Mechanisms and Machine Science

Maneesh Singh · Gunjan Soni · Jyoti Sinha · Andrew D. Ball · Fengshou Gu · Huajiang Ouyang · Carol Featherston *Editors*

Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences

UNIfied 2024–Volume 2

IFToMM

Springer

Mechanisms and Machine Science

Volume 185

Series Editors

IFToMM, Fredericton, Canada

Marco Ceccarelli, Department of Industrial Engineering, University of Rome Tor Vergata, Roma, Italy

Advisory Editors

Burkhard Corves, RWTH Aachen University, Aachen, Germany

Victor Glazunov, Mechanical Engineering Research Institute, Moscow, Russia

Alfonso Hernández, University of the Basque Country, Bilbao, Spain

Tian Huang, Tianjin University, Tianjin, China

Juan Carlos Jauregui Correa, Universidad Autonoma de Queretaro, Queretaro, Mexico

Yukio Takeda, Tokyo Institute of Technology, Tokyo, Japan

Sunil K. Agrawal, Department of Mechanical Engineering, Columbia University, New York, USA

This book series establishes a well-defined forum for monographs, edited Books, and proceedings on mechanical engineering with particular emphasis on MMS (Mechanism and Machine Science). The final goal is the publication of research that shows the development of mechanical engineering and particularly MMS in all technical aspects, even in very recent assessments. Published works share an approach by which technical details and formulation are discussed, and discuss modern formalisms with the aim to circulate research and technical achievements for use in professional, research, academic, and teaching activities.

This technical approach is an essential characteristic of the series. By discussing technical details and formulations in terms of modern formalisms, the possibility is created not only to show technical developments but also to explain achievements for technical teaching and research activity today and for the future.

The book series is intended to collect technical views on developments of the broad field of MMS in a unique frame that can be seen in its totality as an Encyclopaedia of MMS but with the additional purpose of archiving and teaching MMS achievements. Therefore, the book series will be of use not only for researchers and teachers in Mechanical Engineering but also for professionals and students for their formation and future work.

The series is promoted under the auspices of International Federation for the Promotion of Mechanism and Machine Science (IFToMM).

Prospective authors and editors can contact Mr. Pierpaolo Riva (publishing editor, Springer) at: pierpaolo.riva@springer.com

Indexed by SCOPUS and Google Scholar.

Maneesh Singh · Gunjan Soni · Jyoti Sinha ·
Andrew D. Ball · Fengshou Gu · Huajiang Ouyang ·
Carol Featherston
Editors

Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences

UNIfied 2024–Volume 2

Springer

Editors
Maneesh Singh
Department of Mechanical and Marine Engineering
Western Norway University of Applied Sciences
Bergen, Norway

Jyoti Sinha
Department of Aerospace and Mechanical Engineering, School of Engineering
The University of Manchester
Manchester, UK

Fengshou Gu
Department of Engineering and Technology
University of Huddersfield
Huddersfield, UK

Carol Featherston
School of Engineering
Cardiff University
Cardiff, UK

Gunjan Soni
Department of Mechanical Engineering and Department of AI and Data Engineering
Malaviya National Institute of Technology
Jaipur, Rajasthan, India

Andrew D. Ball
Department of Engineering and Technology
University of Huddersfield
Huddersfield, UK

Huajiang Ouyang
School of Engineering
The University of Liverpool
Liverpool, UK

ISSN 2211-0984 ISSN 2211-0992 (electronic)
Mechanisms and Machine Science
ISBN 978-3-031-95962-2 ISBN 978-3-031-95963-9 (eBook)
https://doi.org/10.1007/978-3-031-95963-9

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Contents

Detecting Additive Manufacturing Anomalies with Shallow Convolutional Neural Networks 1
Anshul Kumar Yadav, Dhiraj, Gunjan Soni, Suveg V. Iyer, and Kuldip Singh Sangwan

Vehicle Wheel Monitoring and Diagnosis Based on the Vibration Signals of an On-Rotor Sensor 15
Kunzuo Zhong, Dongqin Li, Yang Zeng, Chun Li, Bing Li, Zhexiang Zou, Fengshou Gu, and Andrew D. Ball

Speed Tracking and Dynamic Analysis of Condition Monitoring of Industrial Collaborative Robot Based on Vibration Signal 35
Huanqing Han, Dongqin Li, Fengshou Gu, and Andrew Ball

Fatigue Damage Assessment of FOWT Mooring Lines Using Sequence-to-Sequence Based Indirect Sensing 47
Rohit Kumar, Subhamoy Sen, and Arvind Keprate

Temporal Feature Extraction Based Real Time Damage Detection of Floating Offshore Wind Turbine Mooring Lines 59
Rohit Kumar, Ananay Thakur, Subhamoy Sen, and Arvind Keprate

An Energy Management Strategy for Piezoelectric Energy Harvester for Low-Power Applications 69
Naveen Tyagi, Anshu Mli Gaur, and Rakesh Mishra

A Framework for Development of Industrial Metaverse in Maintenance 83
Ravdeep Kour, Ramin Karim, Naveen Venkatesh, and Veronica Jagare

Inter-Organizational and Multi-objective Operational Risk Management: A Case Study of Well Control Management for Land Drilling Operation in Iraq 95
Guicang Peng, Knut Erik Bang, and Tore Markeset

An Improved Framework to Transition from TBM to CBM in the Maintenance of Freight Locomotives 113
Mohamed Ali Mohamed Alaswed and Jyoti K. Sinha

Application of Learning Curves in Maintenance 4.0: A Comprehensive Review 127
Prakash Naraian Sharma, Gunjan Soni, and M. L. Meena

A Review on Cable-Driven Soft Robotic Systems Used for Surgical Applications 141
Samaksh Judson, Chetan Jalendra, and B. K. Rout

Potential of Artificial Intelligence in Millets Supply Chain Risk Management 163
Hemant Kumar, Rishi Kumar Jha, and Om Ji Shukla

Sequential Hybrid Solution to Alleviate Congestion Utilizing Distributed Generations and FACTs Devices in the Power Network 175
Mohammad Ibrar, Charu Sharma, and Pawan Sharma

Understanding the Role of Cyber Physical Technology in Supporting Supply Chain Flexibility: A Systematic Literature Review 187
Angel Singh, Bharti R. Soni, and Sachin Ghai

A Conceptual Framework to Evaluate Customers' Response to Use Product Recommendation Systems 213
Aman Mathan, Deepak Verma, and Divesh Kumar

Probabilistic Assessment on Natural Frequencies of Functionally Graded Plates Using Artificial Neural Network Approach 225
R. K. Rout, P. K. Karsh, and R. R. Kumar

Damage Detection of Box Girder Based on Impact Load and the Improved Wavelet Packet-Permutation Entropy 237
Yutao Zhou and Jyoti K. Sinha

Gearbox Anomaly Detection in Wind Turbines Using Classical Machine Learning 253
Duy Tran, Pedro Lind, and Arvind Keprate

Enhanced Gear Fault Diagnosis Using Optimized Spectrogram-Based CNNs for Multiple and Compound Fault Detection 273
Priyom Goswami, Vedant Agrawal, Uday Chaudhary, and Rajiv Nandan Rai

Experimental Investigation of Vibrational Characteristics of Prosthetic Feet 289
Preeti Chauhan, Amit Kumar Singh, and Naresh K. Raghuwanshi

Vibration Trend Monitoring of Indigenous Wankel Rotary Engine 301
Shah Brijeshkumar, Sadanand Kulkarni, Umesh Kumar Sinha, Santhosh Kumar, and A. J. Steve Mithran

Implementation of Multi Criteria Decision Making (MCDM) Techniques in Selecting Influencing Factors for the Affecting of Artificial Intelligence (AI) in the Manufacturing 311
Ashish Mogra, Man Mohan Siddh, and Peeyush Vats

Digital Governance RFP Analysis: BERT Model Adoption for Document Classification 321
Manisha Tiwari, Padmanabhan Aital, and Padmaja Joshi

Application of Condition-Based Monitoring for Improvement of Diagnostic Coverage of Safety Mechanisms 347
Pramit Nag

Predictive Maintenance of Hydraulic Systems Using Multi-task Transfer Learning with Multi-layer Perceptron 359
Tarun Kumar, Mahak Kokate, Pavan Kumar Kankar, and Ankur Miglani

Condition Monitoring of Hydraulic Systems Using Multi-output Classification Convolutional Neural Networks 371
Mahak Kokate, Tarun Kumar, Pavan Kumar Kankar, and Ankur Miglani

Securing Enterprise Data with 4D Security Framework in Advanced Cyber-Physical Systems 385
Meenakshi Malik, Vijay Prakash Sharma, and Surya Prakash

Combined Effect of Charge Rate and Temperature on Battery Health of Electric Vehicles 405
Namrata Mohanty, Neeraj Kumar Goyal, and V. N. Achutha Naikan

Bi-LSTM-Based Model for Software Reliability Prediction in Multi-sprint Agile Based Software 413
Shikha Dwivedi and Neeraj Kumar Goyal

A Temporal Decomposition-Based LSTM Model for Forecasting Reservoir Inflow Using Real-Time Data 423
Kshitij Tandon and Subhamoy Sen

Weak Key Analysis of HQC 435
Ajay S. Rodge and Bhupendra Singh

A Survey on Hamming Quasi Cyclic (HQC) 449
Ajay S. Rodge and Bhupendra Singh

Enhancing Phishing Detection: The Impact of Visual Warning Designs and User Engagement in Email Clients 461
Sandeep Malipeddi

The Integration of Artificial Intelligence in the Context of Industry 4.0 495
Aleksey G. Tashkinov

The Role of Artificial Intelligence in Sustainable Food Retailing: A Pathway to Reducing Waste and Enhancing Efficiency 503
Şirin Gizem Köse, Ipek Kazancoglu, and Yigit Kazancoglu

Pathway to Net-Zero Emissions in Supply Chain: A Comprehensive Analysis of Industry 4.0 Technologies and Emission Reduction Practices 515
Shobha Rathore, Rakesh Jain, and Gunjan Soni

Integrating Automation and Orchestration in Security Incident Handling: A Review of SOAR Frameworks and Platforms 529
Pragya Rathore and Kishor Kolhe

Trust in Social Media Through Deepfake Detection: A Literature Review 553
Debasmita Guha and Sumedha Sirsikar

Framework for a Non-intrusive Monitoring Solution for Independent Elderly 569
Halldor Broddi Thorsteinsson, Anne-Lena Kampen, Kai Goebel, and Maneesh Singh

Assessment of *ProgPy* as an Open-Source Condition Monitoring and Diagnostics Tool 585
Magnus S. Nordeide, Anne-Lena Kampen, Chetan S. Kulkarni, Mayank S. Jha, and Maneesh Singh

Development of Human Cognition Inspired Digital Twins for Optimized Performance of Process Equipment 603
Maneesh Singh, Venkataramana Runkana, Knut Øvsthus, Arvind Keprate, and Anne-Lena Kampen

Application of a Large Language Model (LLM) for Failure Modes and Effects Analysis (FMEA) 619
Kamal V. Charan, Surender Redhu, Gunjan Soni, and Maneesh Singh

Enhancing Circularity in SMEs by Incorporating Industry 4.0 Technologies—An Analysis on Gaps and Barriers 635
Rohit Singh, M. L. Mittal, Rohit Joshi, and Chhavi Solanki

Modern Tools for Understanding Causality: A Comparative Review of Inference Methods 647
Ayushi Rajpoot, Paras Garg, Gunjan Soni, and A. P. S. Rathore

Machine Learning-Driven Clustering Based Environmental, Social, and Governance Performance Prediction Model 661
Gorkem Sariyer, Mert Sarioglu, and Bharti Ramtiyal

Detecting Additive Manufacturing Anomalies with Shallow Convolutional Neural Networks



Anshul Kumar Yadav, Dhiraj, Gunjan Soni, Suveg V. Iyer, and Kuldip Singh Sangwan

Abstract Additive manufacturing often known as 3D printing, has been significant in the manufacturing industry in recent decades. However, the method encounters significant challenges in the form of printing errors, adversely impacting end-user product experience and obstacles to widespread adoption. The current manual and sensor-based continuous monitoring techniques lack a clear distinction between anomalies and healthy data points, making them ineffective for implementation in industrial environments. This research introduces a computer vision-based methodology for detecting anomalies in real-time. Two Convolutional Neural Networks versions are created, Model V1 using residual connection with decreased parameters and computational complexity and Model V2 to facilitate effortless deployment on constraint devices without compromising performance. The proposed CNN networks are evaluated against state-of-the-art classification models, namely ResNet18, ResNet34, and Deep LSTM classifier, to assess their performance. Model V1 and Model V2 achieved comparable performances with 86.7% and 11.86% reduced parameters compared to ResNet18. Afterward, quantization is applied to produce a compact model representation for edge-device deployment. The quantization model proposed has no loss in performance. Lastly, an inference study is conducted on multiple edge devices where the TI AM68A board proved fast, with 0.246 and 0.04 s inference time for models V1 and V2 respectively.

Keywords 3D printing · Anomaly detection · Deep learning · Computer vision

A. K. Yadav · Dhiraj
CSIR-Central Electronics Engineering Research Institute (CSIR-CEERI), Pilani, Rajasthan, India

Dhiraj
Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh, India

G. Soni
Malaviya National Institute of Technology, Jaipur, Rajasthan, India

S. V. Iyer (✉) · K. S. Sangwan
Birla Institute of Technology and Science Pilani, Pilani, Rajasthan, India
e-mail: P20200051@pilani.bits-pilani.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_1

1 Introduction

Additive manufacturing (AM) is a growing manufacturing technology where computer graphical modeling is molded into physical objects with the help of a material extrusion process. The AM method offers the possibility of a unique design with multi-material capability, low-cost manufacturing, and little to no post-manufacturing processing requirement [1]. Considering its usefulness and possibility, AM methods are being utilized in prominent areas like healthcare [2], power electronic devices [3], robotics [4], and aerospace [5]. Despite its advancements in recent years, the AM method is still susceptible to diverse errors, including shrink holes in the structure, edge erosion, abnormal vibration, and total build failure due to mechanical weakness [1]. To identify errors during printing, multiple monitoring systems have been developed through research, and the existing literature on AM anomaly detection can be bifurcated into 3 categories: Manual observation-based, Sensor-based, and Image-based anomaly detection [6]. In manual observation-based anomaly detection, a skilled worker usually monitors the printing process to discover any errors, rectify them by adjusting printing and machinery parameters, and restart the process from scratch. This observation method relies on observer expertise and thus is prone to error and other disadvantages, including waste of material, energy, and human resources. The reliability decreases further for complex multi-material/multi-layer lattice structures [7].

The sensor-based approach employs various current, inertial, and acoustic sensors to capture monitoring parameters and use a detection algorithm to detect anomalies. For example, [8] used an MPU 6050 sensor to acquire vibration data of a fused deposition modeling (FDM) 3D printer with different filament material. The anomaly in this study is manually introduced. The authors reported 97.17% accuracy using the Long Short-Term Memory (LSTM) model. Similarly, [9] employed an additional acoustic sensor with an accelerometer to record vibration patterns to determine printer condition. The analysis of recorded data demonstrates that the vibration peaks at sudden acceleration or deceleration. However, the authors could not differentiate between anomalous and non-anomalous data points.

Although the reported accuracy using sensor-based approaches is within acceptable range, the lack of clear distinction between anomalous and non-anomalous data proved to be the biggest hurdle [10]. Multiple authors [11, 12] have implemented different signal processing/conditioning steps, which have led to mediocre results. Another disadvantage of sensor-based monitoring systems is the high cost of sensors and the requirement for calibration. Therefore, researchers have moved toward image-based error detection systems, also known as camera-based monitoring systems. An image-based error detection system is designed by utilizing one or more cameras positioned on the printer frame, coupled with a classification algorithm to detect diverse errors associated with AM. For example, [13] utilized the thresholding method on Hue, Saturation, and Value (HSV) images captured during the printing process. However, utilizing a static thresholding method can result in misclassification. In [14], the authors designed an image examination method for layer wise

anomaly detection using multiple stage monocular(s). The algorithm also employs a hierarchical clustering algorithm to detect structural defects, while the iterative closest point algorithm helps with contour anomaly detection. The developed framework depends on multiple preprocessing and segmentation algorithms, leading to high computation requirement making them inappropriate for real-time detection. Tunable parameters of each individual algorithm must also be manually perfected as per requirement or use an optimizer, further adding complexity. Therefore, for the image-based error detection systems to reach their full potential, traditional classification and image processing algorithms are rendered useless for real-time error detection in 3d printers [15].

The recent rise in machine learning (ML) and particularly deep learning (DL) methods for image-related tasks has enabled state-of-the-art performance in multiple domains. For AM error detection, [15] introduced a detection model based on UNet architecture and trained using a net-scrapped dataset, reporting 98% as accuracy on the testing dataset. Lyu and Manoochehri [16] developed a CNN classification model using the VGG16 model as the backbone, trained on cloud data obtained with the help of a 3D scanner. Karna et al. [17] demonstrated the use of improved Yolov8 architecture, equipped with hyperparameter tuning and trained on 5000 images for six types of fault detection. The improved architecture achieved an accuracy of 89.7%. Brion et al. [6] introduced a multi-head neural network for diverse error detection and suggested parameter changes to mitigate identified errors. Similarly, the authors of [18] designed a monitoring system for the Fused Filament Fabrication (FFF) method using the camera to identify deviations in target and printed contour. However, all the above-mentioned DL models have extensive network structures and require large datasets to achieve acceptable performance scores. The developed models are also deployed on high-end computer systems, which are not suitable for the industrial floor environment. Using non-constraint computation devices can be costly and is restricted to research applications exclusively [19].

ML-empowered edge computing devices are rising technologies that bring ML computing to edge hardware devices and can be a viable solution for real-time anomaly detection in the AM process. The hardware devices are generally programmable using MATLAB, python, and C++ language and feature X86 or ARM architecture computer processors with possible reusable capability [20]. Therefore, this work focuses on devising a real-time, industry-relevant image-based anomaly detection system coupled with a shallow CNN-based architecture having fewer parameters, lower model size, and accurate classification. The major contributions of the proposed work are listed below:

- Two convolutional neural networks are designed with reduced parameters and computational complexity for faster classification, even on low-end microcontrollers (MCU).
- Quantization is investigated to optimize the generated classifier for deployment on edge devices.
- A quantitative analysis is conducted to determine the ideal device for deployment having a faster inference time.

The rest of the article is organized as: Sect. 2 contains information on the dataset, Model engineering, quantization, and node devices. The detailed benchmarking result of the employed model, quantized model, and inference analysis are provided in Sect. 3. The study's conclusion is presented in Sect. 4.

2 Methodology

This section explains the open-source dataset, the employed classification model, and the inference device used in this study, as shown in Fig. 1.

2.1 Dataset

This study features an open-source dataset [21] hosted on the Kaggle website, collected using a Raspberry Pi and Ender-3 Pro 3D printer. Ender-3 Pro is an industrial printer specially designed to produce great print quality with reduced ripple. Figure 2a presents few instances of the collected dataset, highlighting the distinction between anomalous and non-anomalous images. The images clearly distinguish between the classes, which is impossible in a sensor-based dataset. Figure 2b presents instances of the processed image. For preprocessing, every image is cropped to feed only the information associated with the printing space and printer nozzle to the model. Following the cropping, each image has dimensions of 180 pixels by 500 pixels (H by W).

Figure 3 presents the number of image points for both labels. The label "Anomalous" has 759 images, whereas 798 images belong to "Mon-anomalous" label. Since the number of data points in each label is about equal, the trained classifier will not exhibit bias towards any specific label. Before training, the dataset is divided into a ratio of 80:20, totaling training, and testing images to 1246 and 311, respectively.

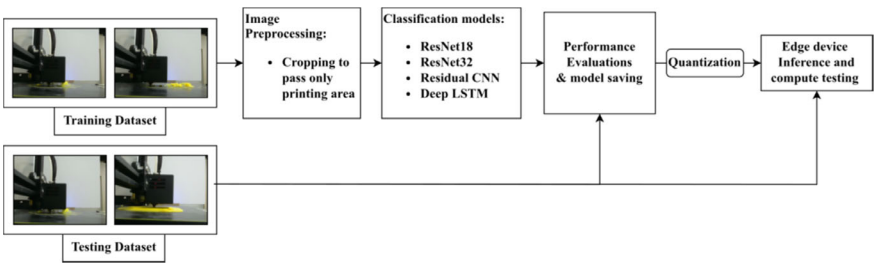
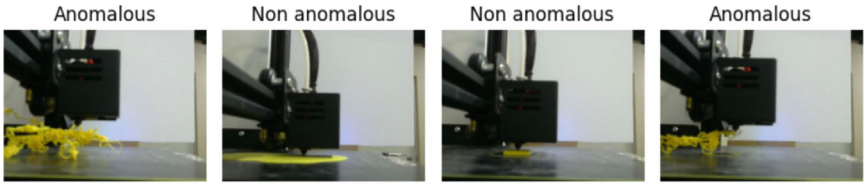


Fig. 1 Pipeline of the study



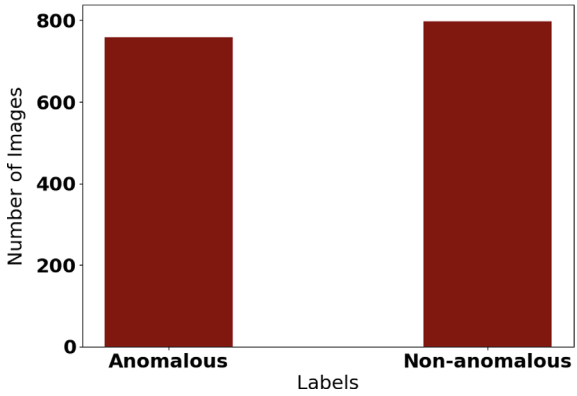
(a) Captured dataset using Raspberry Pi



(b) Processed dataset (after cropping)

Fig. 2 Dataset before and after pre-processing

Fig. 3 Number of image points per label



2.2 Anomaly Detection Models

This subsection highlights the proposed architecture and enumerates ResNet18, ResNet34, and Sequencer: Deep LSTM models that were utilized for performance comparison.

ResNet18 and ResNet34: ResNet18 [22] belongs to the ResNet architecture family and is the most lightweight among other family members. The architecture is based on 18 CNN layers with 8 residual blocks. The architecture involves a novel residual learning framework introducing skip connections (or shortcut connections) that bypass one or more layers. The skip connections allow information merger in the form of identity mapping between two non-sequential layers.

ResNet34 [22] also belongs to the ResNet architecture family and is the second-most lightweight among other family members. As the name suggests, the architecture contains 34 CNN layers with 16 residual blocks. The functionality of skip connections and identity mapping remains the same.

Sequencer: Deep LSTM: The Deep LSTM sequencer [23] is designed using LSTM cells instead of convolution or attention layer. This model consists of patch embedding, sequencer, and patch merging blocks. The patch embedding layer utilizes the convolution layer to separate the input image into patches. The sequencer block consists of a BiLSTM layer and a multi-layer perceptions (MLP) network. The BiLSTM layer helps with better spatial information extraction, while MLP is utilized for channel-mixing. The patch merging block function is opposite to that of the patch embedding block, merging the spatial information containing the patch back to the input image shape. The developed sequencer architecture is also memory-economical due to lack of attention and transformer modules.

Proposed CNN Network: Two CNN-based classifier networks are designed where the residual connection of the ResNet architecture inspires CNN classifier V1. The designed shallow residual-based classifier V1 is a comprehensive model like ResNet 18 and consists of an input convolution layer, four blocks, and an output block. Table 1 shows the layer configuration of the shallow residual CNN network. Each block consists of 3 residual blocks as sub-blocks followed by an average pooling layer of kernel size 2 * 2, where each sub-block has two 2d convolution layers separated by batch normalization and ReLu activation functions. Learning of each concentrated sub-block is carried forward with the help of identity mapping.

The CNN classifier V2 model is a more compact version and is specifically developed for microcontroller deployments. The V2 model consists of two convolution layers followed by 3 fully connected layers, as shown in Table 2.

The increased fully connected layer helps with high-level reasoning by performing complex non-linear input and output mappings, compensating for the reduced residual block while reducing the model computation.

2.3 Dynamic Quantization

Dynamic quantization is performed on each employed model after training. Dynamic quantization involves representing the model's internal parameters using a reduced precision integer representation, such as Int 8 instead of Float 64. The quantization procedure involves multiplying the weight/activations of the model, which are initially in Float 64 precision, by a scaling factor and then rounding the result to a whole number. The scaling factor for model activations is dynamically recognized by analyzing the observed data range during the conversion process using a representative dataset. The dynamic scaling nature of the utilized mode learning retention assures consistent accuracy during deployment. The dynamic scaling is restricted to activations only, whereas the model weights are readily converted to Int 8.

Table 1 Layer details for proposed Model V1

Block	Operation	Output channel	Kernel size
–	Input layer: convolution	16	3 × 3
–	BatchNorm2d and ReLu	16	–
1	Residual block	16	3 × 3
1	Residual block	16	3 × 3
1	Residual block	16	3 × 3
1	Average pooling	16	2 × 2
2	Residual block	32	3 × 3
2	Residual block	32	3 × 3
2	Residual block	32	3 × 3
2	Average pooling	32	2 × 2
3	Residual block	64	3 × 3
3	Residual block	64	3 × 3
3	Residual block	64	3 × 3
3	Average pooling	64	2 × 2
4	Residual block	128	3 × 3
4	Residual block	128	3 × 3
4	Residual block	128	3 × 3
4	Average pooling	128	2 × 2
–	Flatten and fully connected layer	43,648	–
–	Output layer	2	–

Table 2 Layer details for proposed Model V2

Block	Operation	Output channel/neurons	Kernel size
–	Input layer: convolution	6	5 × 5
–	MaxPool2d and ReLu	16	–
1	Convolution	16	3 × 3
1	MaxPool2d and ReLu	16	–
2	Flatten, fully connected layer	120	–
2	Fully connected layer	84	–
–	Final layer: linear	2	–

2.4 Edge Devices

Jetson Orin NX: The Jetson Orin NX is part of Nvidia's high-performance edge devices suitable for deep learning platforms with compact size. The 8 GB variant

offers 100 tera operations per second and is equipped with a 1024-core ampere GPU with 32 tensor cores architecture with a max frequency of 765 MHz.

Raspberry Pi 4: The Raspberry Pi 4 is a compact edge platform with a quadcore ARM Cortex-A72 processor with an operating frequency of 1.5 GHz. Raspberry Pi 4 does not feature any special GPU or accelerators for deep-learning architecture. Therefore, it produces slow inference for DL-based models.

TI AM68A Board: The TI AM68A board is designed for versatile, high performance vision processing applications with power efficiency. It is part of the AM68x family architecture, offering a heterogeneous multicore system-on chip processor. The board has a speed of 8 tera operations per second, which is suitable for DL deployment.

Arduino Nicla Vision: The Arduino Nicla Vision is a powerful MCU chip with a dual-arm cortex processor with 2 MB flash and 1 MB of RAM memory. The MCU also features an M4 core as an AI accelerator along with a 2 MP color camera, Wi-Fi, Omnidirectional microphone, and Bluetooth connectivity.

Espressif ESP-32: The Espressif ESP32 is a low-cost MCU chip with 520 KB SRAM (static). The MCU also features Wi-Fi and Bluetooth connectivity, which is essential for edge deployment in an industrial setting. The SRAM consists of IRAM, where 200 kb is designated for program instructions, and DRAM (dynamic) reserves 320 KB to perform computing.

3 Results and Discussion

The mentioned anomaly classifiers are built using the PyTorch framework and trained using a Linux machine equipped with a single GeForce RTX 3060 GPU system with 12 GB of graphical memory for 20 epochs. Cross-entropy loss is used in all the models with ADAM optimizer and a learning rate set to 0.001. The developed model is saved in ONNX format. The edge device inference analysis uses the Edge impulse platform [24].

3.1 Model Training and Computational Complexity

The accuracy and loss scores achieved by each model on the training and testing sets are given in Table 3. Overall, all the employed model performances are competitive for both training and validation sets.

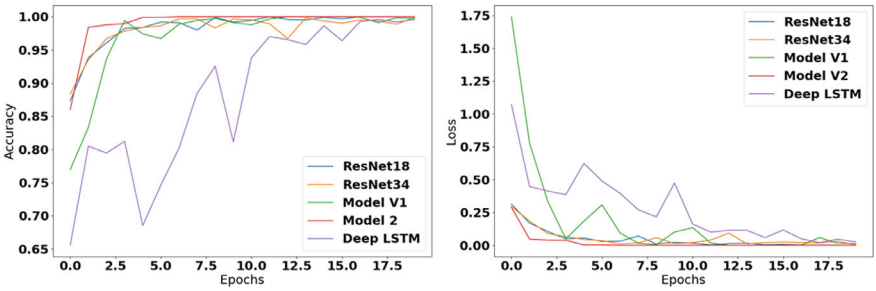
ResNet18 and the proposed Model V2 CNN network are the best-performing classifiers for the training set, achieving accuracy scores of 1 each. In the validation dataset, a perfect accuracy score of 1 is achieved by each model except Deep LSTM. The results suggest that all the models may be employed for anomaly identification in additive manufacturing settings. Figures 4 and 5 presents the loss and accuracy

Table 3 The predictive accuracy and loss of used anomaly classifier

Model	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
ResNet18	0.0003	**1**	0.0007	1
ResNet34	0.0028	0.9992	0.0010	1
Model V1 (Residual)	0.0124	0.9976	0.0019	1
Model V2	**0**	**1**	0.0021	1
Deep LSTM	0.0252	0.9960	0.0249	0.9936

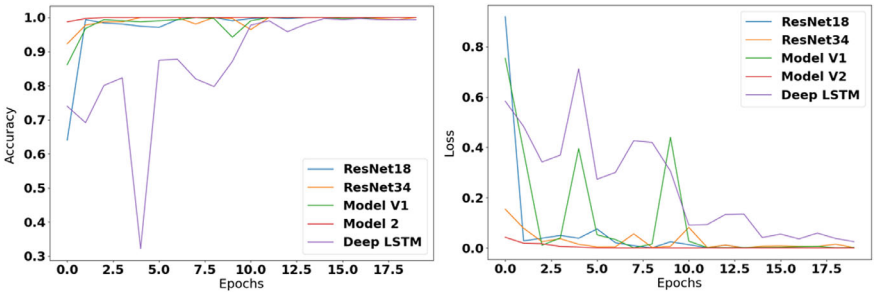
graph achieved by each model during training and validation for 20 epochs. Both the proposed networks are on par with ResNet family architecture, showing quick convergence during training, and are more stable than the Deep LSTM model. A similar trend is observed in Fig. 5, showcasing better convergence and stability than the Deep LSTM model.

To prove the proficiency of the proposed model, computational complexity in the form of floating point operations (FLOPs) to execute a single instance of the classifier



(a) Accuracy curve (b) Loss curve

Fig. 4 Model training performance

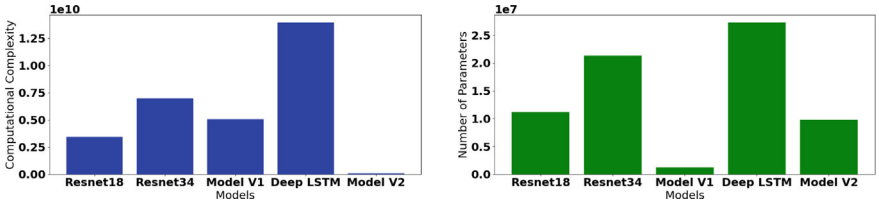


(a) Accuracy curve (b) Loss curve

Fig. 5 Model validation performance

and the number of the parameters is shown in Fig. 6. Regarding flip-flop requirements (Fig. 6a), which is the deciding factor and dictates the inference time, Model V2 has the lowest computational complexity. Specifically, the proposed model has 97.035, 98.54, and 99.967% less computational resources than the ResNet18, ResNet34, and Deep LSTM models. Meanwhile, Model V1 has the second-lowest computational resource requirement. As seen in Fig. 6b, the proposed model V1 uses very few parameters compared to other models. Specifically, the model V1 has 86.7, 93.01, and 94.58% fewer parameters compared to ResNet18, ResNet34, and Deep LSTM classifiers. The proposed Model V2 ranks second, regarding the lowest number of parameters.

Another factor that dictates the disk space required to store the trained model, the bar graph representation of model size, is shown in Fig. 7. Regarding storage space, the saved V1 model file is only 6.0 MB compared to 44.8 MB, 85.3 MB, and 109.2 MB of ResNet18, ResNet34, and Deep LSTM. The V2 model file has the second lowest storage requirement of 39.4 MB. Therefore, it can be concluded that the suggested models demonstrate the ability to attain equivalent outcomes and stability as significantly bigger models while maintaining a low level of computational complexity and model size.



(a) Flip Flop requirements (b) Number of parameters

Fig. 6 Computational complexity analysis of each model

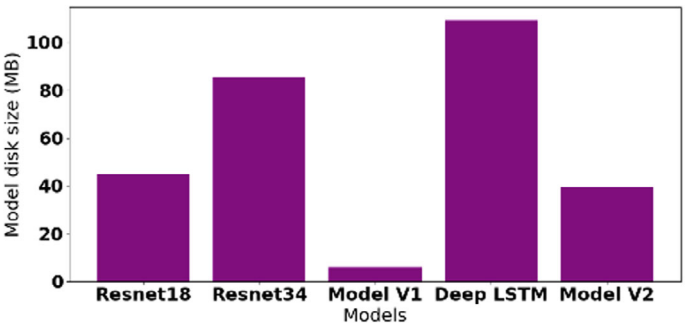


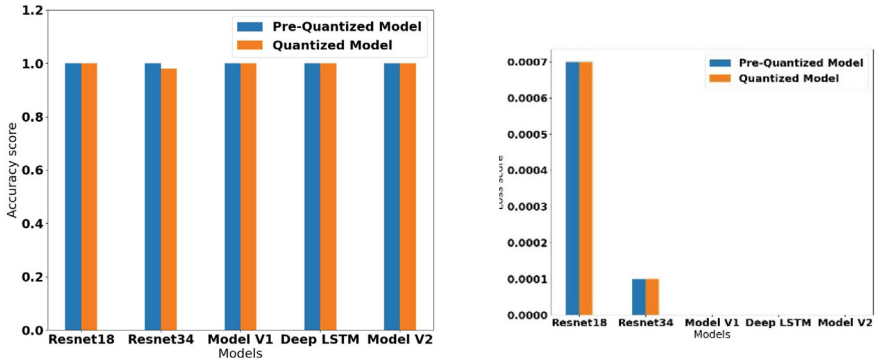
Fig. 7 Disk space analysis

3.2 *Quantization and Deployment Analysis*

Figure 8 presents the performance of the quantized model on the testing dataset. As seen, dynamic quantization can retain the performance of all models with similar accuracy.

For inference analysis, five devices (Espressif ESP-32, Arduino Nicle Vision, TI AM68A, Jetson Orin NX, and Raspberry Pi4) are simulated with the help of the edge impulse platform. Table 4 compares the inference time of all five devices for Model V1 and Model V2. The Model V1 can be deployed to CPU or higher-end devices, whereas the Model V2 is compatible with all the five selected edge devices. The RAM and ROM usage of Model V2 is 3.814 MB and 3.945 MB, whereas Model V1 requires RAM greater than 6 MB, making it not feasible to implement on the chosen MCU and MCU + AI accelerator edge device.

Utilizing TI AM68A, the proposed models V1 and V2 can produce inference in 246 s and 0.04 s respectively. Meanwhile, the Jetson Orin NX is the second fastest, having an inference time of 0.7 s and 0.1 s, respectively.



(a) Accuracy (b) Loss

Fig. 8 Quantize model performance

Table 4 Deployment analysis

Devices	Device features				Inference time (s)	
	MCU	CPU	GPU	AI accelerators	Model V1	Model V2
Espressif ESP-32	✓				–	69.315
Arduino Nicla						
Vision	✓			✓	–	9.674
Raspberry Pi 4		✓			13.081	0.135
Ti AM68A		✓		✓	**0.246**	**0.04**
Nvidia Jetson Orin		✓	✓		0.7	0.1

4 Conclusion

This study aims to solve anomaly detection, a pressing problem of additive manufacturing in Industry 4.0, using a deep learning method. In this research, utilizing an open-source dataset, two defect detection algorithms are designed using convolution blocks. Model V1 uses residual block, whereas model V2 is a simple CNN network with a feed-forward neural network designed for microcontroller deployment. Both the proposed models have competitive performance similar to the ResNet18, ResNet34, and Sequencer Deep LSTM models. The developed models offer an advantage over others by significantly reducing the number of parameters and computational complexity. Additionally, the study also dives into the dynamic quantization of the model to gauge performance reduction by reducing floating point precision to Int 8. Furthermore, model deployment and data inference was performed by using five edge devices, and it was found that the TI AM68A device has the shortest inference time, taking only 0.246 s for model V1 and 0.04 s for model V2. In future research, a control feedback loop that adjusts parameters can be targeted to resolve errors in real-time. Exploring data collection using various printing machines and materials is a promising avenue to further evaluate the efficacy of the developed models.

References

1. Ngo, T.D., Kashani, A., Imbalzano, G., Nguyen, K.T., Hui, D.: Additive manufacturing (3d printing): a review of materials, methods, applications and challenges. Compos. B Eng. **143**, 172–196 (2018)
2. Kumar, A., Kumar, D., Faisal, N., Sharma, A., Kumar Ansu, A., Goyal, A., Saxena, K. K., Prakash, C., Kumar, D.: Application of 3d printing technology for medical implants: a state-of-the-art review. Adv. Mater. Process. Technol. **10**(2), 357–372 (2024)
3. Wang, F.-H., You, C.-Y., Tian, N., Liu, H.-G., Zhang, J., Zhu, X.-P.: 3d printing of soft magnetic materials: from printing to applications. J. Alloy. Compd. **990**, 174486 (2024)
4. Khalid, M.Y., Arif, Z.U., Tariq, A., Hossain, M., Khan, K.A., Umer, R.: 3d printing of magneto-active smart materials for advanced actuators and soft robotics applications. Eur. Polymer J. **205**, 112718 (2024)
5. Adeleke, A.K., Montero, D.J.P., Lottu, O.A., Ninduwezuor-Ehiobu, N., Ani, E.C.: 3d printing in aerospace and defense: a review of technological breakthroughs and applications. World J. Adv. Res. Rev. **21**(2), 1149–1160 (2024)
6. Brion, D.A., Pattinson, S.W.: Generalisable 3d printing error detection and correction via multi-head neural networks. Nat. Commun. **13**(1), 4654 (2022)
7. Mohamed, O.A., Masood, S.H., Bhowmik, J.L.: Optimization of fused deposition modeling process parameters: a review of current research and future prospects. Adv. Manuf. **3**, 42–53 (2015)
8. Kumar, R., Sangwan, K.S., Herrmann, C., Ghosh, R., Sangwan, M.: Development and comparison of machine-learning algorithms for anomaly detection in 3d printing using vibration data. Prog. Addit. Manuf. **9**(2), 529–541 (2024)
9. Waheed, M.F., Bernadin, S.: In-situ analysis of vibration and acoustic data in additive manufacturing. In: SoutheastCon 2024, pp. 812–817. IEEE (2024)

10. Becker, P., Roth, C., Roennau, A., Dillmann, R.: Acoustic anomaly detection in additive manufacturing with long short-term memory neural networks. In: 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), pp. 921–926. IEEE (2020)
11. Dirhamsyah, M., Riza, H., Rizal, M.S.: Utilizing a knowledge-based training algorithm and time-domain extraction for pattern recognition in cylindrical features through vibration and sound signals. J. Meas. Eng. **12**(1), 40–52 (2024)
12. Kammerer, K., Hoppenstedt, B., Pryss, R., Stökler, S., Allgaier, J., Reichert, M.: Anomaly detections for manufacturing systems based on sensor data—insights into two challenging real-world production settings. Sensors **19**(24), 5370 (2019)
13. Langeland, S.A.K.: Automatic error detection in 3d printing using computer vision. Master's thesis, The University of Bergen (2020)
14. Petsiuk, A.L., Pearce, J.M.: Open source computer vision-based layer-wise 3d printing analysis. Addit. Manuf. **36**, 101473 (2020)
15. Lalitha, T., Anushkannan, N., Shreepad, S., Sasireka, S., Anandaram, H., Razia, S.: Deep learning-based automatic 3d printer anomaly detection during the printing process. In: 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), pp. 1343–1348. IEEE (2022)
16. Lyu, J., Manoochehri, S.: Online convolutional neural network-based anomaly detection and quality control for fused filament fabrication process. Virtual Phys. Prototyp. **16**(2), 160–177 (2021)
17. Karna, N.B.A., Putra, M.A.P., Rachmawati, S.M., Abisado, M., Sampedro, G.A.: Toward accurate fused deposition modeling 3d printer fault detection using improved yolov8 with hyperparameter optimization. IEEE Access **11**, 74251–74262 (2023)
18. Werkle, K.T., Trage, C., Wolf, J., Möhring, H.-C.: Generalizable process monitoring for FFF 3d printing with machine vision. Prod. Eng. Res. Devel. **18**(3), 593–601 (2024)
19. Cao, M., Fanrong, F., Ai, et al.: Improved 3d printing extrusion defect detection method based on yolo-v8. In: Improved 3d Printing Extrusion Defect Detection Method Based on Yolo-V8, 2024. Available at SSRN: https://doi.org/10.2139/ssrn.4800007
20. Lu, S., Lu, J., An, K., Wang, X., He, Q.: Edge computing on IoT for machine signal processing and fault diagnosis: a review. IEEE Internet Things J. **10**(13), 11093–11116 (2023)
21. Ruan, J.: 3d-printer defected dataset. Data retrieved from Kaggle website, (2022) https://www.kaggle.com/datasets/justin900429/3d-printer-defected-dataset
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Arxiv e-prints, arXiv preprint arXiv:1512.03385, vol. 10 (2015)
23. Tatsunami, Y., Taki, M.: Sequencer: deep LSTM for image classification. Adv. Neural. Inf. Process. Syst. **35**, 38204–38217 (2022)
24. Hymel, S., Banbury, C., Situnayake, D., Elium, A., Ward, C., Kelcey, M., Baaijens, M., Majchrzycki, M., Plunkett, J., Tischler, D., et al.: Edge impulse: an MLOps platform for tiny machine learning. arXiv preprint arXiv:2212.03332 (2022)

Vehicle Wheel Monitoring and Diagnosis Based on the Vibration Signals of an On-Rotor Sensor



Kunzuo Zhong, Dongqin Li, Yang Zeng, Chun Li, Bing Li, Zhexiang Zou, Fengshou Gu, and Andrew D. Ball

Abstract Wheel bearings and tyres are vital components of various vehicles. Their conditions are one of critical factors influencing fuel efficiency, vehicle handling, ride comfort and safe operations, which are more significantly concerned by modern autonomous vehicles. This study introduces a novel On-Rotor Sensing (ORS) method to collect acceleration responses on wheel hub to obtain more accurate and comprehensive information for on-line monitoring wheel conditions. The method is exemplified by a field experiment when a on road car is equipped with an ORS triaxial MEMS accelerometer and ran with different tyre pressures at different speeds. The high signal to noise data from the ORS allows instantaneous wheel rotation speed to be easily calculated for online wheel lockage monitoring and wheel order tracking analysis. Based on tyre resonance responses in the frequency bands from 70 to 200 Hz, the radial and lateral vibrations give a comprehensive detection of changes in tyre pressures using either changes in spectral amplitudes or spectral centroids. However, the lateral amplitudes along can produce sufficiently good detection in the speed ranges tested. Wheel order tracking analysis can be implemented with the instantaneous speed information, making it possible to observe bearing ball pass frequency outer race (BPFO) to assess bearing health conditions. In addition, the amplitudes of wheel harmonics are more accurate, allowing better monitoring and assessing wheel eccentricity and imbalances.

Keywords Wheel bearings · Tyre · On-rotor sensing · On-line monitoring · Order tracking analysis · Spectral centroid

K. Zhong · D. Li · C. Li · Z. Zou · F. Gu (✉) · A. D. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, UK
e-mail: F.Gu@hud.ac.uk

K. Zhong
e-mail: Kunzuo.Zhong@hud.ac.uk

K. Zhong · D. Li · Y. Zeng · C. Li · B. Li · Z. Zou
School of Industrial Automation, Beijing Institute of Technology, Zhuhai, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_2

1 Introduction

The passenger car wheel bearing unit is one of the vital parts of the car. Its main function is to carry the mass of the whole car, drive the wheel hub to rotate, and provide the rotation function. With the development of the automotive industry, the demands for lightweight, unitization, comfort, and multifunctionality of wheel hub bearings are increasing. Wheel bearings perform as an intermediate medium in the automotive driving system, connecting the tyre to the suspension system. Via the wheel bearing, the tyre is connected to the constant velocity axle (CV axle), to ensure smooth rotation of the tyre. At the same time, the suspension system is connected to the wheel bearing through swing arms and struts, which control the up-and-down movement of the wheels and keep them in constant contact with the ground. This structural design not only provides a smooth steering and driving experience, but also effectively absorbs vibrations caused by uneven road surfaces, thus improving the overall stability and comfort of the vehicle [1].

During vehicle operation, wheel bearings are frequently in a state of variable rotational speed owing to the constant variation of vehicle speed and load, which results in the vibration signals collected from the bearings showing cyclic non-stationary characteristics. Adopting the traditional spectrum analysis technique to process such signals may lead to spectral blurring. The order tracking technique effectively separates the influence of rotational speed by converting the vibration signal from the time domain to the order domain, and enables accurate tracking of rotational speed changes, showing significant advantages in fault diagnosis under changing operating conditions [2].

Selecting the correct inflation pressure is vital to the operational condition of a vehicle's tires, affecting not only the performance of the vehicle but also the longevity of the tires and the safety of the driving experience [3]. Operating a tyre at more than the required inflation pressure may result in reduced driving comfort, reduced grip, uneven tyre wear, increased risk of blow-outs and damage to the suspension. Similarly, operating below the required inflation pressure can lead to reduced grip, reduced fuel efficiency, faster tyre wear, increased risk of blow-outs and reduced handling, as well as overheating of the tyre, damage to the sidewalls and tyres coming off the wheel hubs, which can reduce the vehicle's performance [4]. Hamed et al. [3] developed a seven-degree-of-freedom model and validated it experimentally to analyse the effects of different tyre pressures on suspension vibration, handling and ride comfort. Tang et al. [5] suggested a remaining useful life prediction method based on fault signals, utilizing signal processing and machine learning techniques to effectively identify the health state of wheel bearings and predict their remaining service life. Sorniotti et al. [6] assessed the effect of friction inside wheel bearings through analytical modelling and experimental methods, pointing out the importance of friction on bearing performance, fuel economy and driving experience. Toma et al. [7] study the effect of tyre pressure on the diagnostic results of braking and suspension systems, and find that tyre pressure changes significantly affect the diagnostic

results of the suspension system, but have a more minor effect on braking performance at low speeds. Pardeshi et al. [8] monitored two-wheeler tyre pressure through machine learning methods and data acquisition systems, collected wheel vibration data using accelerometers, and successfully achieved over 90% classification accuracy using decision trees and random forest classifiers. Kim et al. [9] identified the sources of automotive wheel bearing noise through experimental measurements and modal analysis, and determined that deformation of the outer ring due to bolt tightening was the main cause of noise and vibration through finite element analysis. Guan et al. [10] investigated the vibration characteristics of radial tyres by experimental and numerical techniques and analysed the vibration behaviour under different conditions, including the frequency response and dynamic characteristics of the tyres. Xu et al. [11] analysed the association between the vibration characteristics of complex tread tyres and the influencing factors such as inflation pressure and load by means of numerical simulation models. Gonzalez Diaz et al. [12] investigated the vibration characteristics and dynamic response of tyres during rolling by experimental and numerical simulation methods. Feng et al. [13] proposed a frequency compensation algorithm for four-wheel correlated random road surfaces to solve the problem of the effect of road surface unevenness on the vibration response of vehicles through mathematical modelling and algorithmic analysis. Tang et al. [14] developed a low-power wireless monitoring system for wheels, incorporating a micromotor to continuously monitor wheel alignment in real time. They verified the system's effectiveness through field experiments conducted under different vehicle speeds and tyre angles. Du et al. [15] used a high-speed uniformity tester to analyse the effects of varying speeds and pressures on multiple uniformity parameters of tyres. Marton et al. [16] monitored tyre pressure by observing changes in tyre vibration frequency. Through comparative analysis of different signal processing methods, they examined the spectra of tyre acceleration signals to identify characteristic frequencies that vary with tyre pressure changes. Prażnowski et al. [17] used vehicle longitudinal acceleration data to monitor pavement conditions in real time, capturing the effects of different pavement types on the vehicle and constructing an assessment model.

Although researchers have conducted numerous studies on condition monitoring of vehicle transmission systems, however, the relationship between tyre pressure variations and wheel bearings has rarely been explored. This study utilises ORS sensors [18] to monitor the vibration response of wheel bearings under variable inflation pressures. Due to the complex structure of the vehicle's transmission system and the random excitation of the road surface, the vibration signals of the wheel bearings are submerged in the complex background noise, and the frequency components of the bearings are not distinguishable by direct Fourier transform. Thus, based on the characteristics of ORS sensors, the zero-crossing point is extracted from the gravitational acceleration component of the vibration signal and the instantaneous velocity of the vehicle is accurately estimated by differential computation, followed by angular resampling to obtain the order spectra of the system, and finally, the characteristic frequency of the bearing's faults is clearly observed by order analysis.

This paper is organised as follows: Sect. 2 demonstrate the development of a tyre-hub dynamics model to investigate the relationship between wheel vibration

response, hub angular velocity, and coordinate acceleration. Section 3 describes the structure and working principle of the ORS vibration measurement system. In Sect. 4, a field experimental study is carried out under varying tyre pressures and speeds, and the captured measurement data are analysed to validate the ORS vibration signals. Finally, Sect. 5 summarises the study and proposes directions for future work.

2 Theoretical Backgrounds

2.1 Wheel Vibration Model and Characteristic

While the wheel rotates on the road it will be excited by the waviness of road surface and local defects and vibrates and produces vibrations with restriction of by the springs and dampers of tyres and suspension systems, as well as the driving CV axle. The tyre has two types of vibration modes in radial and lateral directions respectively. A tyre-hub dynamics model is developed in this section, and the schematic of the flexible sidewall tyre model utilized is shown in Fig. 1 [19]. The analysis focuses on the correlation between wheel radial and lateral vibrations and hub angular velocity and acceleration in the coordinate direction, which is used to describe the vibration behaviour related to wheel speed, hub assembly and tyre pressure monitoring. The tyre is connected to the hub by springs and dampers in the horizontal, lateral and torsional directions. The rim has only one rotational degree of freedom ω_r, with an associated moment of inertia J_r. The wheel has translational degree of freedom x, vertical degree of freedom z, lateral degree of freedom y, and rotational degree of freedom φ_r ($\varphi_r \doteq \omega_r$).

The displacement of a point on the tyre rim (points O, A and B) is shown in Fig. 2 for a fixed coordinate system (x, z) at different stages of the system's motion. The

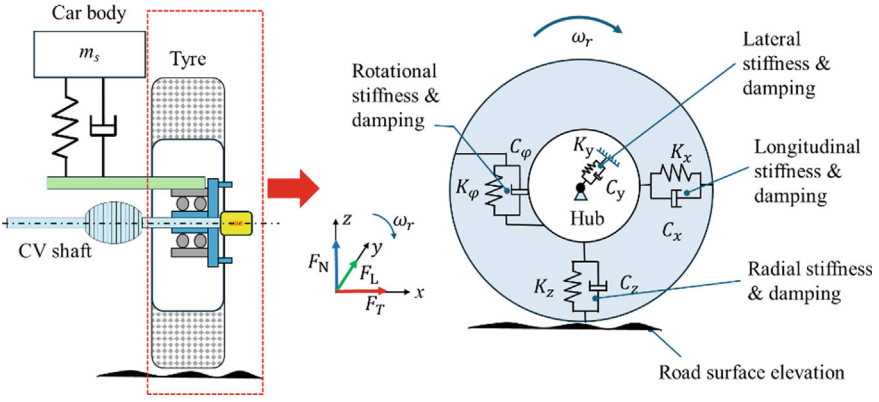


Fig. 1 The tyre-hub vibration model

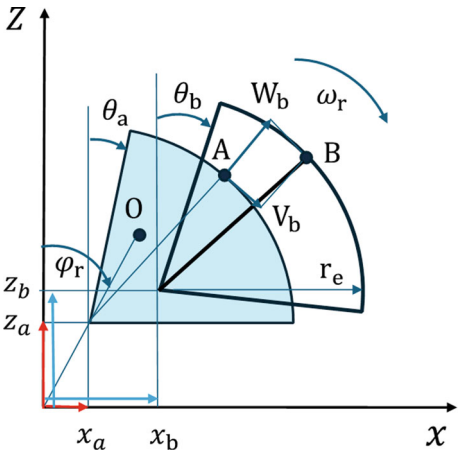


Fig. 2 The displacement of a point on the tyre

tyre-wheel system rotates around the wheel axis with a constant angular velocity ω_r. Point O is defined as a point on the tyre ring, with the wheel axle rotating at a constant speed and a radius of r. φ_r denotes the angle of this undisturbed rotation. Point A represents a point on the tyre ring displaced by the axes x_a and z_a and an additional rotation of the rim θ_a. The position of the tyre ring after the displacement is denoted by B. The horizontal and vertical positions for the case of B are denoted by x_b and z_b, respectively, and the rotation angle is denoted by θ_b.

From Fig. 2, the tangential V_b and radial W_b components of the tyre deformation are equal to the positional difference between point A and point B:

$$\begin{bmatrix} V_b \\ W_b \end{bmatrix} = \begin{bmatrix} \cos(\varphi_r + \theta_a) & -\sin(\varphi_r + \theta_a) & r_e \\ \sin(\varphi_r + \theta_a) & \cos(\varphi_r + \theta_a) & 0 \end{bmatrix} \begin{bmatrix} x_b - x_a \\ z_b - z_a \\ \theta_b - \theta_a \end{bmatrix} \tag{1}$$

Based on the Eq. (1), the tangential and radial deformation of the tyre can be known as the x- and z-direction displacements are known. The corresponding dynamic equations can be solved based on the model in Fig. 1. To simplify the analysis, consider the case where the wheel/hub is supported on a rigid structure (no suspension flexibility) and the tyre rolls on the horizontal road surface. Three types of forces are generated at the contact between the tyre and the road surface: the normal force F_N, which carries the weight of the vehicle, the lateral force F_L of the tyre, and the tangential force F_T, which is used for traction and braking, according to the expression for the dynamics of the flexible carcass, which is given by:

$$\begin{aligned} &m_a\ddot{x}_b + K_x(\dot{x}_b - \dot{x}_a) + C_x(x_b - x_a) - K_z\left(\omega_r + \dot{\theta}_a\right)(z_b - z_a) = F_T + F_N \\ &m_a\ddot{z}_b + K_z(\dot{z}_b - \dot{z}_a) + C_z(z_b - z_a) + K_x\left(\omega_r + \dot{\theta}_a\right)(x_b - x_a) = -F_T + F_N \\ &m_a\ddot{y}_b + K_y(\dot{y}_b - \dot{y}_a) + C_y(y_b - y_a) = -F_L \\ &J_y\ddot{\theta}_b + K_\varphi\left(\dot{\theta}_b - \dot{\theta}_a\right) + C_\varphi(\theta_b - \theta_a) = -r_eF_T + M_y \end{aligned} \tag{2}$$

where r_e is the tyre rolling radius, the tangential force is converted to a driving torque acting on the tyre ring using r_e. The rolling resistance is modelled as the additional torque M_y acting on the tyre ring from the road.

The lateral and radial tyre vibration as a function of time is obtained by combining Eqs. (1) and (2). The forces acting on the wheels are related to dynamic vehicle loads, road unevenness, drive/brake torque fluctuations and tyre inhomogeneities. Dynamic changes in load forces result in transverse and longitudinal vibrations of the system. Therefore, the tyre-hub system's mechanical vibration characteristics can be obtained through hub-mounted ORS sensors.

The tyre-hub system also subjects to excitations from wheel imbalance and eccentricity in the tyre-wheel system. They will lead to the vibration responses mainly at the wheel rotation frequency which becomes more significant at high speeds.

In addition, the wheel bearing can be another effective excitation. Even if there is no clear local defect on the race ways, the contact forces between the races and rolling element can vary periodically as the contact spring rotates around the shaft when it subjects to asymmetric static load. This contact force and its induced vibration are more observable at ball pass frequency outer ring (BPFO), which is about several times of wheel frequency such as $BPFO \approx (5 \sim 7)f_r$. In addition, its amplitude is also proportional to wheel speeds.

2.2 ORS Vibration Measurement System

The ORS is an integrated wireless rotating machinery sensing system that is integrated into a small 3D printed box with external fastening bolts that attach to the rotor of any rotating machinery, and the ORS rotates with the rotor axis for stable vibration signal acquisition. The hardware system includes a sensing unit, a processing unit, a wireless communication unit and a power supply unit. A MEMS accelerometer unit is selected for the sensor module, which can acquire triaxial vibration data at a high sampling rate and high resolution. The signals collected through the sensors are acquired by the processing unit and then wirelessly transmitted to the host software for data processing and other operations through the communication module. The communication features include Bluetooth for connecting to an Android application for real-time data visualisation and storage. The hardware system modules are integrated on a single printed circuit board to improve space utilisation. A 3.7 V Li-ion battery is used as a power module to power the circuit. The relative position of the ORS sensors to the rotor is fixed and the ORS signals are not subject to complex transmission paths.

From Fig. 3, it can be derived that ORS responses in U-V-W rotation frame can be obtained by:

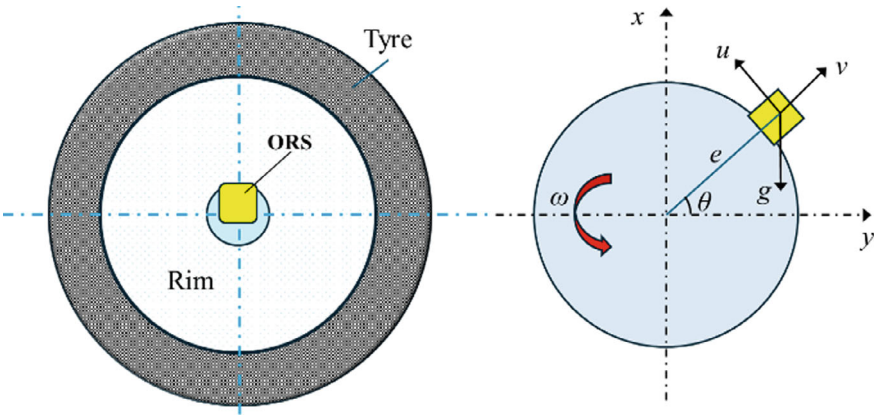


Fig. 3 The outputs of the ORS installed on the wheel hub

$$\begin{aligned}\ddot{u}(t) &= g\cos(\omega t) + \ddot{\theta}(t)e - \ddot{y}(t)\sin(\omega t) + \ddot{x}(t)\cos(\omega t)\\ \ddot{v}(t) &= -g\sin(\omega t) + \omega^2(t)e + \ddot{x}(t)\sin(\omega t) + \ddot{y}(t)\cos(\omega t)\\ \ddot{w}(t) &= \ddot{x}(t)\sin(\omega t) + \ddot{y}(t)\cos(\omega t)\end{aligned} \tag{3}$$

where g represents the acceleration of gravity, ω presents the rotational angular velocity of the rotor wheel. $\ddot{\theta}(t)$, $\ddot{x}(t)$ and $\ddot{y}(t)$ indicate the vibration acceleration in the torsional, vertical and horizontal directions, respectively, in the stationary frame. Additionally, compared with the gravity acceleration g, the projection components of $\ddot{\theta}(t)$, $\ddot{x}(t)$ and $\ddot{y}(t)$ are much smaller. And the $\omega^2(t)e$ can be neglected as the sensor installation deviation $e \ll 1$.

The change process of gravitational acceleration can be found to involve both tangential and radial directions according to Eq. (3). If the decoupling of gravitational acceleration is implemented, the rotational speed can be estimated by calculating the phase of the acquired gravitational acceleration components. Moreover, either of the three outputs can reflect the vibrations in the stationary frame.

3 Experimental Methods

3.1 ORS Setup

The vehicle used in the experiment is a Honda minicar as illustrated in Fig. 4, and with the ORS sensor mounted at the centre of the right front wheel on the passenger side. During the experiment, the car speed was controlled by the driver, and the vibration signal was collected at the passenger position through a mobile phone application and transmitted wirelessly in real time to the acquisition device via a

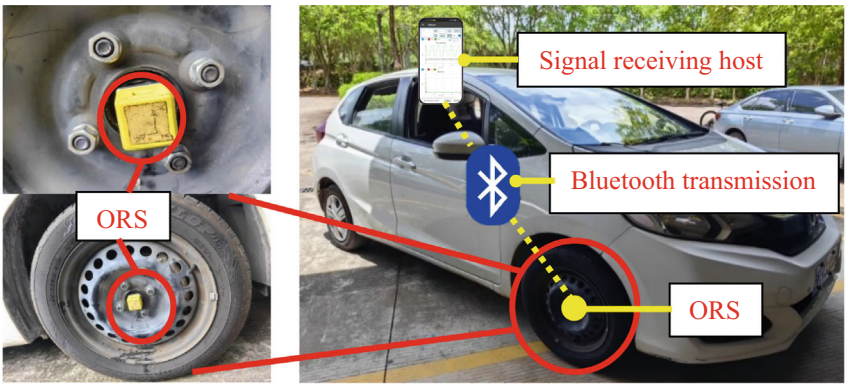


Fig. 4 ORS mounted on the centre end of the right front wheel

Bluetooth connection. The tyre pressure was adjusted by a pressure regulator and the sampling frequency was set to 8000 Hz with a measurement range of 8 g. The wheel hub bearing kit consists of a hub and a bearing, connected through an interference fit.

3.2 Datasets for Different Speeds and Tyre Pressures

The field experiment was conducted on a city express road (Grade C) in the suburb of Zhuhai, as shown in Fig. 5, with a total mileage of approximately 3 km, and point A was recorded as the initial point and point B as the terminal point, the test road was smooth overall, with no significant potholes or undulations. During the tests, car speed was adjusted at the test speed as stable as possible, making it easier to analyse the signals with conventional tools such as FFT.

The car was driven at different speeds from 40 to 80 km/h, which are typical speed ranges in city suburbs. The data was recorded as one test when the car was driven from point A to point B. In addition, the tyre inflation pressure near the passenger side was set to different pressures ranging from 1.6 to 2.5 bar, and the pressures of the remaining three tyres remained as the standard one of 2.5 bar. In total, there are 2 × 20 datasets acquired corresponding to 5 speeds (40, 50, 60,70 and 80 km/h), 4 tyre pressures (1.6, 1.9, 2.2 and 2.5 bar), 2 reciprocating journeys. These tests allow a comprehensive examination of the performance of using ORS vibration to monitor the wheel health conditions such as issues with tyre pressure, wheel bearing, imbalance and suspension.

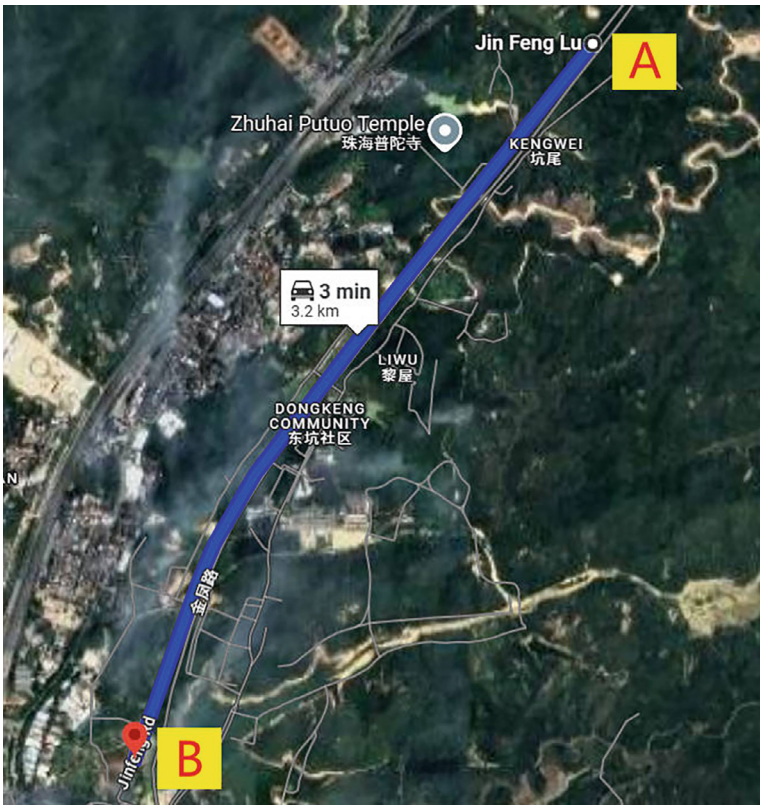


Fig. 5 The route of data acquisition

4 Wheel Condition Monitoring and Discussion

To examine monitoring information from ORS vibration datasets, the data are presented in the time, time–frequency domain and angular domain respectively. These will help to obtain information regarding to wheel speed, tyre inflation pressures, wheel unbalance and wheel bearings of which are common problems with the wheel.

4.1 Waveform Analysis for Wheel Speed Monitoring and Vibration Levels

As the car travels on the uneven road, the wheel system is subjected to excitations due to uneven road surfaces. Thus, the ORS accelerometer can effectively acquire responses to these excitations.

Figure 6 shows typical acceleration waveforms of wheel for 2 s at 50 and 80 km/h for the normal tyre inflation pressure of 2.5 bar. It can be seen that the variation of acceleration with time is significant, exhibiting a strong periodicity of the waveforms in the two radial directions (U and V) and U and V having a phase difference of 90°. However, the vibration responses in W representing in lateral vibrations are much lower and show less periodicity because it is less influenced by the gravitational acceleration, as shown in Sect. 2. The waveforms show higher amplitudes and faster fluctuation at the high speed of 80 km/h, which is agreeable with the that of speed induced excitations. Moreover, due to the unevenness of the road surface, significant impulsive signals can be noticed, which also shows that high speed causes more impulses.

To show more detailed vibration responses, RMS values, representing vibration intensity are calculated for the datasets. Figure 7 demonstrates the trend of the RMS values of wheel in different directions with the increase of car speed under different tyre inflation pressures. It shows that with the increase of car speed, the RMS values in lateral and radial directions show an increasing trend overall, and the radial direction RMS can be about 1 g, being significantly higher than the lateral. The magnitude of RMS value varies with the increase of car speed is different for different tyre pressures. In general, with higher tyre pressures and faster vehicle speeds, the RMS values

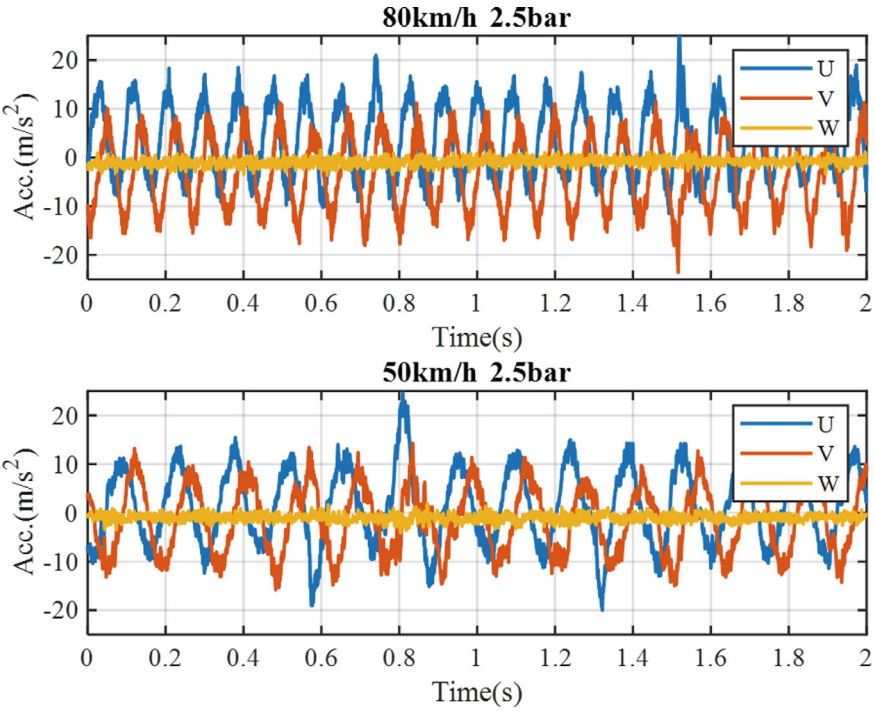


Fig. 6 Acceleration waveforms at different speeds

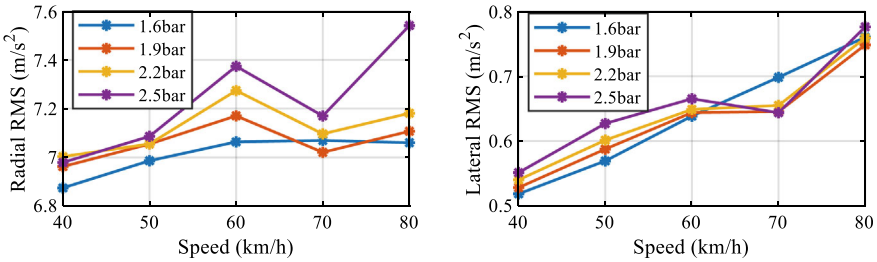


Fig. 7 RMS values at different speeds and tyre pressures

increase, which indicates that the vibration intensity of the wheel is also increased. At lower tyre pressures, the vibration is relatively small, exhibiting its superiority in absorbing vibration, which may lead to better driving comfort. Nevertheless, excessively low tyre pressures may adversely affect other aspects of vehicle performance, such as fuel efficiency and less comfortable riding.

With the strong periodicity characteristics, the variation of wheel instantaneous rotation speed under different car speeds can be calculated online by a simple zero-crossing detection algorithm. The algorithm obtains a time incremental pulse tarin tacho signal by a multiplication between the original signal and its one sample shifted version, which can be easily implemented by a micro-CPU for online applications. Figure 8 shows the instantaneous speed for all datasets. It can be seen that when the car speed increases, the wheel instantaneous rotation speed gradually increases, and the fluctuation amplitude of rotation also increases, especially at high speeds the fluctuation is more significant, which shows that the car speed is proportional to the wheel rotation speed. At high and low tyre pressures, the pressure changes have little effect on the instantaneous wheel speed. Which may provide little information on detecting the pressure changes. Nevertheless, wheel instantaneous speed can be useful to find any slippages of the wheels that caused by ABS, worn tyres and blocking brakes. Moreover, it will help to check the stationarity of signals for choosing adequate data analysis tools.

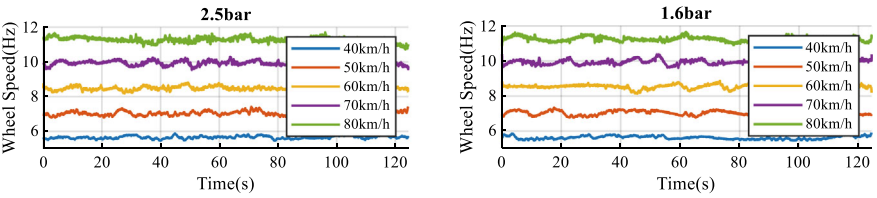


Fig. 8 Instantaneous frequency of wheel rotation for high and low pressures

4.2 *Time–Frequency Analysis for Pressure Monitoring*

As the speed shows clear fluctuations with time and intermediate road impacts, the vibration signals fluctuate largely in instantaneous amplitudes and instantaneous frequencies, showing strong non-stationarity. So common time–frequency analysis tool—short time Fourier transform (STFT) needs to be applied to the data to real more detailed and accurate diagnostic information.

Typical STFT results of the wheel hub vibrations in radial and lateral directions are presented in Fig. 9 for tyre pressure of 2.5 bar at vehicle speed of 40 km/h. The vibrations in the UV radial directions are mainly concentrated in the high frequency range, especially in the high frequency vibration energy near 400 Hz, and the signal amplitude is significantly higher than that in the W direction. The vibration in the W lateral direction is mainly concentrated in the low-frequency range, and the vibration amplitude is smaller and more stable.

As the speed is increased to 80 km/h, the results of the STFT in Fig. 10 show that. The vibration energy is more concentrated in the higher frequency bands, while

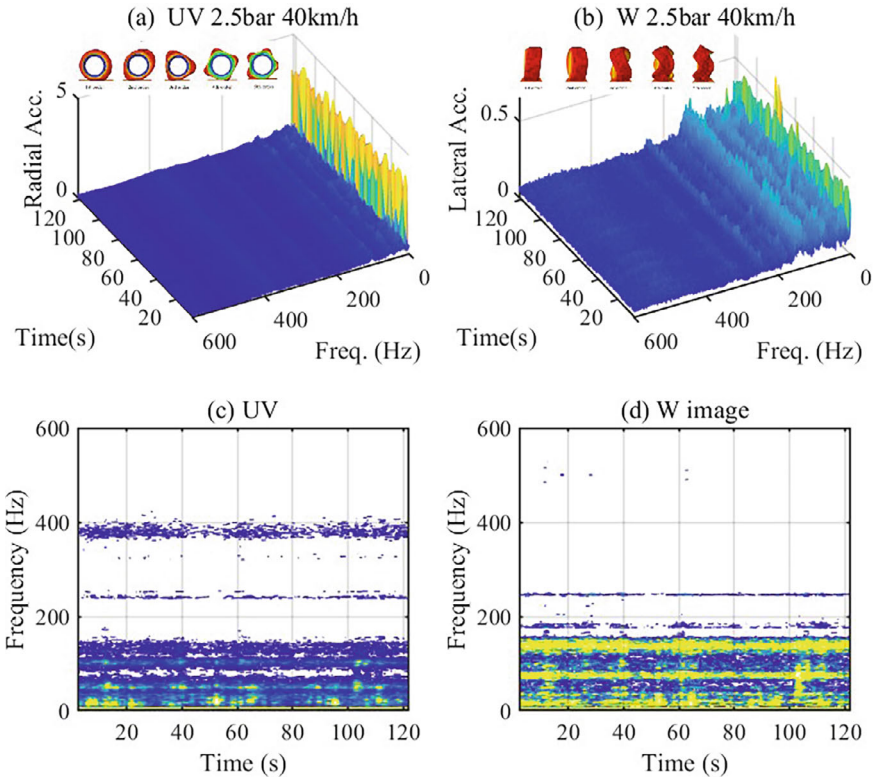


Fig. 9 STFT results of radial and lateral vibrations at the low speed of 40 km/h

in the W direction it is concentrated in the lower frequency bands, which is similar to the analysed results at low speeds. Nevertheless, some periodic features of the frequency distribution can be observed in all directions at high speeds. The rotation frequency in the lateral vibration is the principal vibration response, reflecting the effect of the gravitational acceleration of the ORS sensing methods. Lateral vibration is not affected by the gravitation. In addition, both radial and lateral vibrations show high responses below 200 Hz, which are mainly allocated in several distinctive peaks such as at frequencies of 47, 98, 248, 257, 390 Hz for the radial vibrations and at: 33, 70, 135, 245 and 500 Hz for the lateral vibrations. Such vibrations reflect tyre vibration modes in the lateral and radial directions respectively. In particular, these characteristics are well consistent with that of FEM predictions [10, 11]and the 1st several modes spread over the frequency range from 70 to 200 Hz.

To examine if these vibrations can be used for wheel diagnosis, two common vibration features: spectral amplitude and spectral centroid (SC) are extracted from STFT results for different tyre pressures at different driving speeds. In calculating these features, the impulses are excluded from STFT results in performing the average

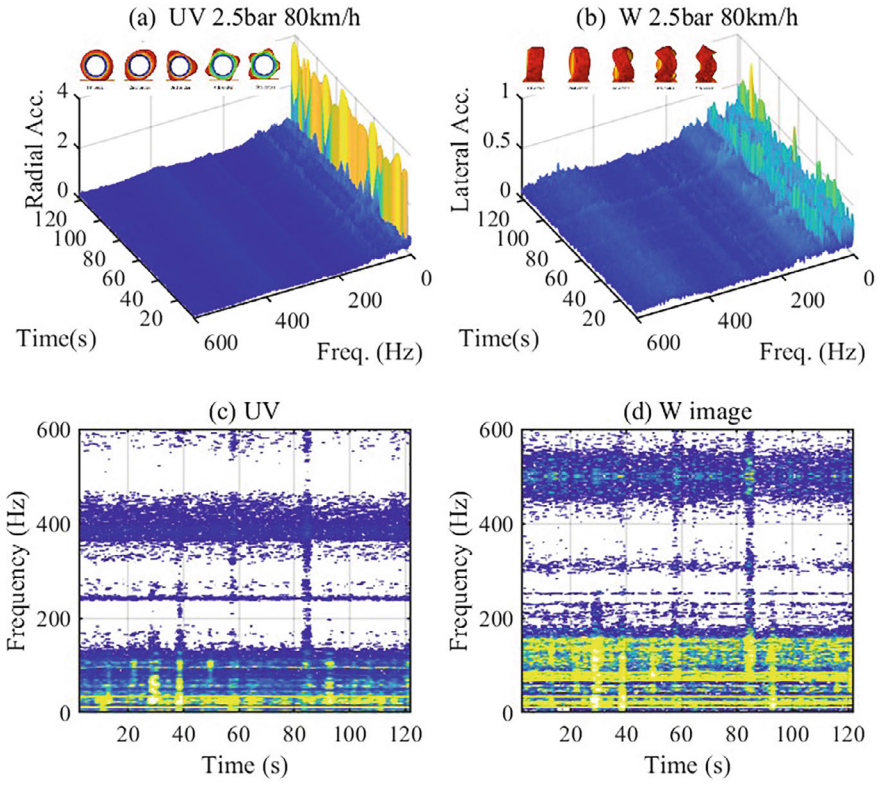


Fig. 10 STFT results of radial and lateral vibrations at the high speed of 80 km/h

over time. It will remove the influences of the large contributions from such impulses, making the result more stable over time.

In addition, although these frequency values exhibit little changes with time, there are high amplitudes at several time instants due to road surface defects such as large dents and cracks. Such impulses are more significant at high speeds as shown in Fig. 6.

Figure 11 shows the feature values in lateral and radial directions for different tyre pressures at different speeds. It can be seen in Fig. 11a that spectrum amplitude of radial vibration increases with speeds and produces fairly good results to separate the pressure different over different speeds. It means the high damping effect of the tyre in low pressure ranges. However, at the high speed of 80 km/h, the low pressure of 1.6 bar shows very high amplitude, making it impossible to separate it at the high speed. This probably caused by the high vibrations in the lateral direction. As shown in Fig. 11b, the spectral amplitude in lateral direction also exhibits increasing with speeds as that of radial one but the amplitude varies reversely with pressure increases. It indicates that as the tyre pressure becomes lower the tyre can oscillate more. Combining these two vibration behaviours it is feasible to comprehensively diagnose the changes in tyre pressures, showing effective information inside ORS signals.

Based on FEM predictions in [10, 11] that tyre pressure decrease will lead to vibration modal frequency decreases in the frequency range from 60 to 200 Hz. The spectral centroid parameter is employed to show these changes. It allows a comprehensive measure of both frequency values and corresponding spectral changes, rather than using the frequency values determined by spectral peaks which are very unstable due to the random excitation of road surfaces. Figure 11c, d show the SCs changes with speed and pressure. SC in radial directions agrees the model prediction of frequency decrease only when the speed lower than 60 km/h. At the high speeds, SC values become lower, being consistent with the model predictions, probably because the flat spectrum caused by high excitations of road surface. However, the SC values in the lateral direction exhibit good agreement with the prediction and can be based to differentiate the pressure differences in the high speed range from 60 to 80 km/h.

4.3 Angular Spectrum for Bearing and Eccentricity Monitoring

Based on the effect of tyre vibration modes, STFT analysis leads to good results for tyre pressure diagnosis. However, because of the time-varying speeds, it is not sufficiently accurate to characterise the components relating to wheel rotations such as wheel imbalance and bearing characteristic vibrations. To overcome the shortage, a wheel order tracking method is implemented based on the instantaneous speeds calculated in Sect. 4.2. It divides the continuous time-based data into short consecutive segments, each corresponding one revolution of the wheel, according to the

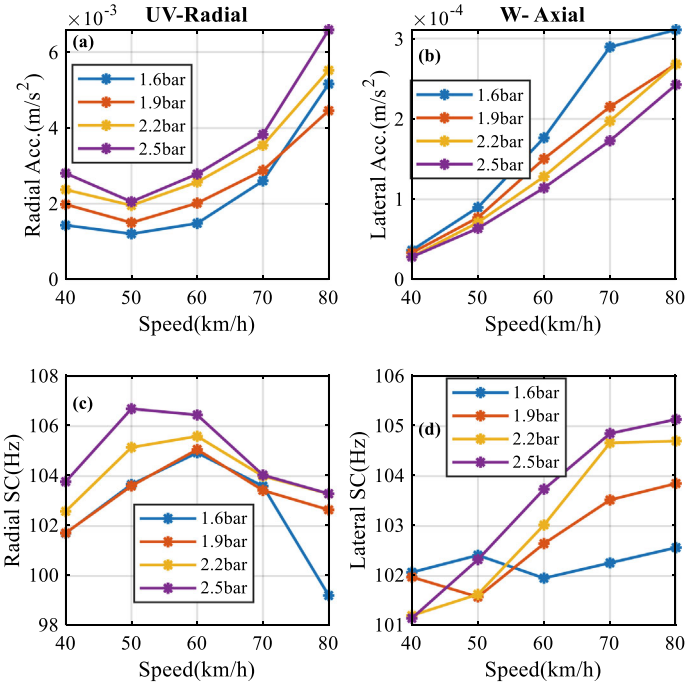


Fig. 11 Spectral amplitude (SA) and spectral centroid (SC) with speed and pressure: **a** SA of lateral vibration in 70–200 Hz, **b** SA of lateral vibration in 90–110 Hz, **c** SC of radial vibration in 90–125 Hz, **d** SC of lateral vibration in 90–125 Hz

tacho pulse signal created during the calculation of instantaneous speed. These short segments are then interpolated into same length so that the long signal cascaded with these segments will become closely cyclic according to the wheel rotation period, resulting data in the angle domain. Then conventional FFT can be applied to the angle domain data to result in the angular spectrum for more accurately representing the cyclic components associated with wheel rotations.

Figure 12 demonstrates typical angular spectrums for three speeds and different tyre pressures. The X-axis ticks show both the wheel rotation orders and their approximate frequency values estimated by the average rotation frequency. Because of the effect of suppression of nonlinearity and random noise by the wheel order tracking, the spectrums highlight the significant amplitudes at the wheel orders and also those at the factional orders relating to wheel bearing characteristic frequency of BPFO. These spectral amplitudes increase largely with speeds, which agree well with vibration mechanisms induced by the excitations due to road waviness and local defects. Moreover, the components relating to BPFO can be observable at the speed higher than 60 km/h at which the vibration becomes more significant than the noise level.

To use these spectra for online monitoring, the mean amplitude of wheel harmonics at different orders is calculated and presented in Fig. 13. It can be seen in Fig. 13a that

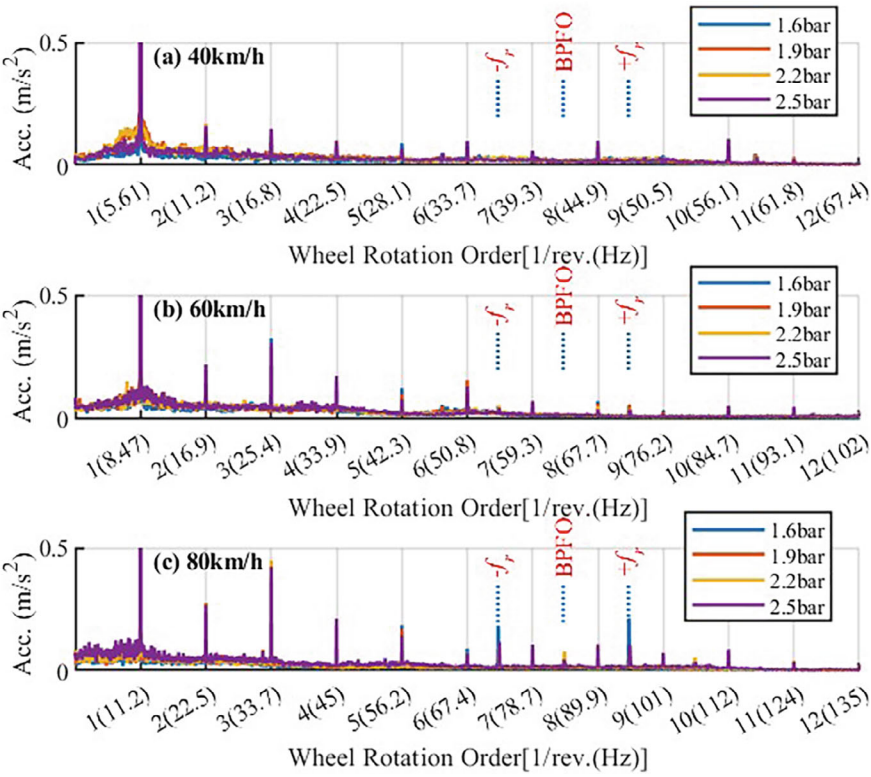


Fig. 12 Spectrum of angle domain signals at 40 km, 60 and 80 km/h

the radial vibrations exhibit a linear increase trend with speed but cannot differentiate pressure differences. This linear change probably can indicate more the eccentricity characteristics of the driving shaft, which is usually not aligned perfectly with wheel hub. In Fig. 13b the amplitude in lateral direction exhibits typical nonlinear increases with speeds. This increase is probably more associated with wheel imbalance characteristics. Especially, the imbalance can be more significant as the tyre pressure becomes low. So, the vibration in lateral direction can also be more effective to separate the pressure changes.

The mean amplitude associated with BPFO including its sidebands at BPFO $\mp f_r$, where f_r is wheel rotation frequency, is also calculated as the indicator for bearing detection. As shown in Fig. 14, vibration responses in both radial and lateral directions increase nonlinearly with speeds, which is consistent with the nonlinear behaviour of Hertz contact dynamic effects when the contact dynamic forces become larger at higher speeds. In addition, both the two types of vibration can be based to detect the bearing conditions when the speeds are higher than 60 km/h but the lateral detection shows better performance as the amplitudes are more stable with pressure variations,

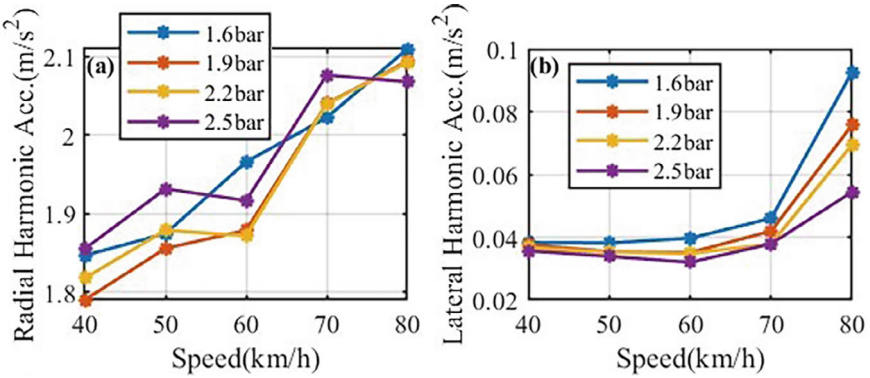


Fig. 13 Spectral amplitudes of wheel harmonics

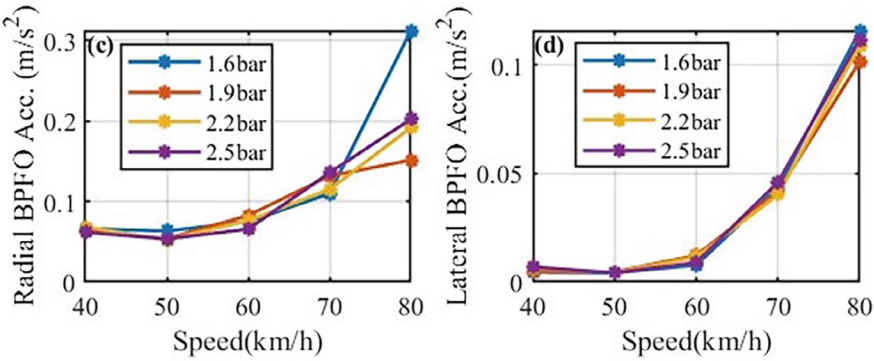


Fig. 14 Spectral amplitudes at bearing characteristic frequency (BPFO)

meaning that bearing conditions can be assessed adequately even if tyre pressures are different.

5 Conclusions

With moderate signal processing effort, the ORS vibration signal on vehicle wheels has been shown to provide valuable information for the online monitoring and diagnosis of wheel health conditions. Due to the high SNR of the radial signal, it is possible to easily calculate instantaneous wheel rotation speed for online wheel lockage monitoring and wheel order tracking analysis. The vibration responses in the frequency range of 70–200 Hz are primarily attributable to tyre resonance modes. By combining the vibrations in the radial and lateral directions, a comprehensive detection of tyre

pressure changes can be achieved, either by monitoring changes in spectral amplitudes or through the spectral centroid. However, lateral amplitude alone has proven sufficient for reliable detection within the tested speed ranges. Using instantaneous speed information, wheel order tracking analysis can also be implemented, allowing for the observation of BPFO to assess bearing health conditions. Moreover, the amplitudes of wheel harmonics can be measured more accurately, facilitating improved monitoring and assessment of wheel eccentricity and imbalance.

This study demonstrates that the ORS test system provides a novel way to achieve wireless signal transmission and online monitoring of the dynamic conditions of axles, wheel bearings, and tyres. This facilitates timely observation of developing faults, and the ability to replace or repair components such as bearings promptly ensures operational continuity and cost efficiency by preventing potential breakdowns. However, this research did not incorporate the effects of variables such as suspension dynamics, vehicle load, road type, complex working conditions, or climate influences, which might significantly impact the performance of the ORS system. Moreover, the absence of simulation experiments means that the results are applicable mainly to the specific conditions studied. Future research will address these factors to extend the applicability of the result.

Acknowledgements This research was supported in part by the 2024 Guangdong Province Science and Technology Innovation Strategy (Climbing Plan Project) Special Fund (No. pdjh2024a437), in part by the Special Projects in Key Areas in Fundamental and Foundational Applied Research of Guangdong Provincial Education Department (No. 2023ZDZX3047), and in part by the National Undergraduate Training Program for Innovation and Entrepreneurship (No. 2024008DCXM).

References

1. Naunheimer, W. N., Harald, Bertsche, B., Ryborz, J.: Automotive Transmissions: Fundamentals, Selection, Design and Application (2011)
2. Lu, S., Yan, R., Member, S., Liu, Y., Wang, Q.: Tacholess speed estimation in order tracking : a review with application to rotating machine fault diagnosis. IEEE Trans. Instrum. Meas. **68**(7), 2315–2332 (2019). https://doi.org/10.1109/TIM.2019.2902806
3. Hamed, M., Tesfa, B., Gu, F., Ball, A.D.: Effects of tyre pressure on vehicle suspension performance. Int. Lett. Chem. Phys. Astron. **55**, 102–111 (2015). https://doi.org/10.56431/p-l6t1lc
4. Reiter, M., Wagner, J., Ph, D.: Automated automotive tire inflation system—effect of tire pressure on vehicle handling 10–15 (2013). https://doi.org/10.3182/20100712-3-DE-2013.00013
5. Tang, S., Gu, J., Tang, K., Zou, R., Sun, X., Uddin, S.: A fault-signal-based generalizing remaining useful life prognostics method for wheel hub bearings. Appl. Sci. **9**(6) (2019). https://doi.org/10.3390/app9061080
6. Sorniotti, A., Sampo, E., Velardocchia, M., Bonisoli, E., Galvagno, E.: Friction inside wheel hub bearings: evaluation through analytical models and experimental methodologies. SAE Tech. Pap. **2007** (2007). https://doi.org/10.4271/2007-24-0138
7. Toma, M., Andreescu, C., Stan, C.: Influence of tire inflation pressure on the results of diagnosing brakes and suspension. Procedia Manuf. **22**, 121–128 (2018). https://doi.org/10.1016/j.promfg.2018.03.019

8. Pardeshi, S.S., Patange, A.D., Jegadeeshwaran, R., Bhosale, M.R.: Tyre pressure supervision of two wheeler using machine learning. SDHM Struct. Durab. Heal. Monit. **16**(3), 271–290 (2022). https://doi.org/10.32604/sdhm.2022.010622
9. Kim, J., Kwon, S., Ryu, S., Lee, S., Jeong, J., Chung, J.: Noise identification for an automotive wheel bearing. Appl. Sci. **12**(11) (2022). https://doi.org/10.3390/app12115515
10. Guan, Y., Cheng, G., Zhao, G., Zhang, H.: Investigation of the vibration characteristics of radial tires using experimental and numerical techniques. J. Reinf. Plast. Compos. **30**(24), 2035–2050 (2011). https://doi.org/10.1177/0731684411431764
11. Xu, M., Ge, Y., Du, X., Meng, Z.: Analysis of vibration characteristics and influencing factors of complex tread pattern tires based on finite element method. Machines **12**(6), 1–14 (2024). https://doi.org/10.3390/machines12060386
12. Gonzalez Diaz, C., et al.: Dynamic behaviour of a rolling tyre: Experimental and numerical analyses. J. Sound Vib. **364**, 147–164 (2016). https://doi.org/10.1016/j.jsv.2015.11.025
13. Feng, J., Zhang, X., Guo, K., Ma, F., Karimi, H.R.: A frequency compensation algorithm of four-wheel coherence random road. Math. Probl. Eng. **2013** (2013). https://doi.org/10.1155/2013/986584
14. Tang, X., et al.: A miniature and intelligent low-power in situ wireless monitoring system for automotive wheel alignment. Meas. J. Int. Meas. Confed. **211**(February), 112578 (2023). https://doi.org/10.1016/j.measurement.2023.112578
15. Du, M., Sun, P., Zhou, S., Huang, H., Zhu, J.: A study on the influence of tire speed and pressure on measurement parameters obtained from high-speed tire uniformity testing. Vehicles **2**(3), 559–573 (2020). https://doi.org/10.3390/vehicles2030031
16. Marton, Z., Fodor, D., Enisz, K., Nagy, K.: Frequency analysis based tire pressure monitoring. 2014 IEEE Int. Electr. Veh. Conf. IEVC. 1–5 (2014). https://doi.org/10.1109/IEVC.2014.7056187
17. Prażnowski, K., Mamala, J., Śmieja, M., Kupina, M.: Assessment of the road surface condition with longitudinal acceleration signal of the car body. Sensors (Switzerland) **20**(21), 1–19 (2020). https://doi.org/10.3390/s20215987
18. Li, C., Zou, Z., Lu, K., Wang, H., Cattley, R., Ball, A.D.: Assessment of a three-axis on-rotor sensing performance for machining process monitoring: a case study. Sci. Rep. **12**(1), 1–12 (2022). https://doi.org/10.1038/s41598-022-21415-w
19. Lechner, G., Naunheimer, H.: The dynamic response of tyres to brake torque variations and road unevennesses (1998)

Speed Tracking and Dynamic Analysis of Condition Monitoring of Industrial Collaborative Robot Based on Vibration Signal



Huanqing Han, Dongqin Li, Fengshou Gu, and Andrew Ball

Abstract Within the realm of intelligent manufacturing, industrial robots occupy a paramount position as electromechanical keystones. Consequently, the scrutiny of advanced state monitoring methodologies for these robotic systems is of utmost significance. Currently, dynamic analysis based on vibration signal has been used to evaluate robot macroscopic behavior. Nonetheless, these tools fall short for the rigorous scrutiny demanded by such endeavours, which necessitate a meticulous dissection of each subcomponent's functionality. This paper aims to fill this gap by combine traditional dynamic analysis methods with robot joint state monitoring, containing an accurate description of its full dynamic characteristics. Harmonic reducer was used to transfer torque, so that arm can move. To conclude, the real robot on a test rig were tested with a set of different condition and speed. The experimental results highlight the ability of the dynamic analysis method to accurately replicate joint angular rotation and speed in a wide range of operational scenarios. The amplitude of characteristic frequency can be used to monitor the running state of robot joint. This research endeavours to construct a comprehensive maintenance CM-oriented framework, tailored for conducting in-depth and sophisticated analyses that unravel the intricate impacts of manipulator degradation.

Keywords Industrial collaborative robots · Condition monitoring · Dynamic analysis · Characteristic frequency · Speed tracking

H. Han (✉) · D. Li
School of Industrial Automation, Beijing Institute of Technology, Zhuhai, China
e-mail: Huanqing.Han@hud.ac.uk

H. Han · D. Li · F. Gu · A. Ball
Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, UK

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_3

1 Introduction

Today, industrial robots have become the backbone of smart production systems. By leveraging various modern technologies, including the Internet of Things, machine learning, and neural networks, these robots are becoming increasingly "intelligent" [1, 2]. However, because robots consist of complex combinations of physical structures and electrical systems, a complete failure necessitates systematic inspection and maintenance, which can lead to costly downtime for the entire production line [3, 4]. To mitigate this issue, several state monitoring technologies have been proposed to assess the condition of industrial robots, identify potential risks, and predict future performance.

Currently, across industries such as aerospace, automotive, and energy, three primary state monitoring approaches are employed [5, 6]: model-based (white-box) techniques, data-driven (black-box) strategies, and hybrid methodologies [7]. Model-based methods entail contrasting the outcomes from dynamic simulations with those from physical models to identify potential system anomalies. Data-driven monitoring constructs models based on historical data to reflect the performance trends of a system [8]. Unlike model-based methods, data-driven approaches depend entirely on the quality and accuracy of available data, which may not fully capture all characteristics of a physical system. Model-based methods are generally regarded as superior for achieving precise and continuous fault detection and state monitoring [9]. Various model-based fault detection techniques have been developed, all aiming to pinpoint critical component, sensor, or operational failures by evaluating discrepancies in measurable signals [10].

For industrial robots, data-driven state monitoring methods are the most widely used. But this lacks in-depth exploration of the causes of internal joint faults, only knowing that there is a fault here, without truly locating which component has caused the fault and what kind of fault has occurred. Of course, this is also because the joint structure of robots is too complex, making traditional dynamic analysis methods difficult to use in this scenario. So, the dynamic monitoring method of industrial robot joints is an urgent problem that needs to be solved. This article aims to make some attempts in this direction, based on On Rotor Sensing (ORS) sensors to track joint speed, and further analyze the dynamic characteristics of joints to obtain fault features based on the characteristic frequencies of internal transmission components.

2 Key Components of Industrial Robot Joint

The harmonic drive reducer has become ubiquitous in industries ranging from advanced robotics and collaborative machines to the aerospace sector, owing to its exceptional attributes such as a high gear reduction ratio, minimal backlash, and a design that is both compact and lightweight. This sophisticated mechanism revolves around three pivotal components: the Wave Generator (WG), the Flexspline (FS),

and the Circular Spline (CS). The WG, a cornerstone of this design, features an elliptical cam, akin to a wave generator plug, nestled within a flexible ball bearing, imparting a unique deformation capability. This assembly interacts with the FS, a delicately thin-walled gear adorned with external teeth, which in turn engages with the CS, a robust gear characterized by internal teeth. The operational dynamics of this system are intricately linked to how these elements are assembled; in applications like the UR5 robotic arm, the input shaft seamlessly integrates with the WG, while the output shaft is directly connected to the CS, with the FS securely affixed to the manipulator's joint housing, ensuring precise motion control. This leads to a reduction ratio (t), defined as Eq. 1:

$$R = \frac{Z_C}{(Z_C - Z_F)} = 101 \tag{1}$$

where R is the gear reduction ratio. $z_c = 202$ is the number of teeth on the circular spline gear. $z_F = 200$ is the number of teeth on the flexible gear. The wave generator's distinctive elliptical form imparts a precise elastic deformation to the flexspline along its principal axis upon rotation. This deformation facilitates an intricate engagement with the circular spline. Notably, the design incorporates a deliberate tooth discrepancy, where the flexspline possesses two fewer teeth than the circular spline. Consequently, when the input shaft completes a full revolution, this meticulous mechanism drives the output shaft to rotate by an angle equivalent to the movement across two teeth, embodying the harmonic drive's efficient gear ratio translation (Fig. 1).

In order to monitor the operating status of harmonic reducers in joints using vibration signals, it is necessary to calculate the characteristic frequencies of each component. As the Table 1 shown. Base on the characteristic frequency, the common faults such as circular spline gear crack fault, flex-spline gear wear fault, flex-spline gear crack fault, et al. can been monitor by the vibration signal.

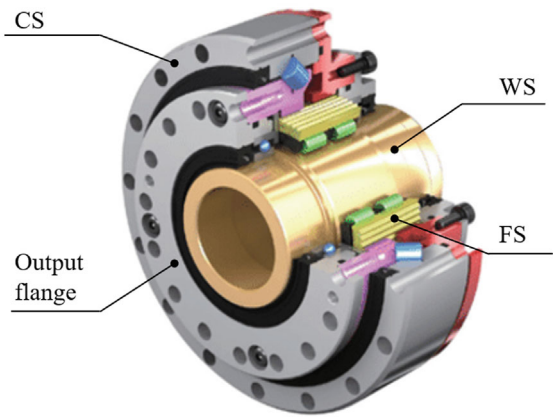


Fig. 1 Schematic diagram of harmonic reducer

Table 1 Characteristic frequencies of each component of harmonic reducer

Contents	Characteristic frequency formula
Wave generator	$n/60$
Circular spline	$2n/60$
Flexspline	$\frac{2n}{60} \times \frac{z_C}{z_F}$
Meshing frequency	$\frac{2n}{60} \times 160$
WG bearing outer ring	$n_r = \frac{z_F n_i + z_C n_o}{z_F + z_C}$
WG bearing ball rotation	$n_x = \frac{n_r - n_o}{d_r} + n_r$

3 Abnormal State Simulation and Procedure of Experiments

The UR5 collaborative robot, a flagship product from Universal Robots, served as the centerpiece for this study. The objective was to capture vibration data during the robot's operational cycle, necessitating the development of an innovative wireless acquisition system, designated as the ORS (Optimized Remote Sensing) system. This sophisticated system was meticulously designed to quantify acceleration across three dimensions—linear and rotational—providing a comprehensive analysis of vibratory forces. Its structure is shown in Fig. 2. The acceleration sensor module serves a critical role in machinery health monitoring. It is specifically designed to capture vibration data that occurs during the operational phase of machines, including industrial robots and rotating machinery. These sensors meticulously record the minute vibrations, which are subsequently analyzed to assess the machine's condition. The vibration signals act as indicators of the machine's health status, enabling early detection of potential issues. By scrutinizing these data, maintenance teams can diagnose any anomalies, such as imbalance, misalignment, or impending component failure, allowing for proactive maintenance and preventing unscheduled downtime. This application is fundamental to predictive maintenance strategies, where the continuous monitoring of vibration data helps in maintaining the optimal performance and longevity of industrial equipment.

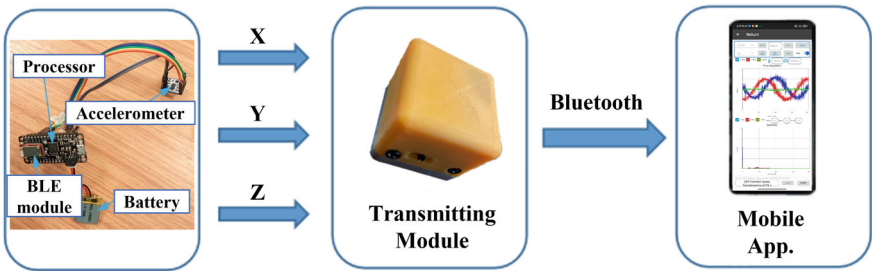


Fig. 2 Three-axis accelerometer based on ADXL345

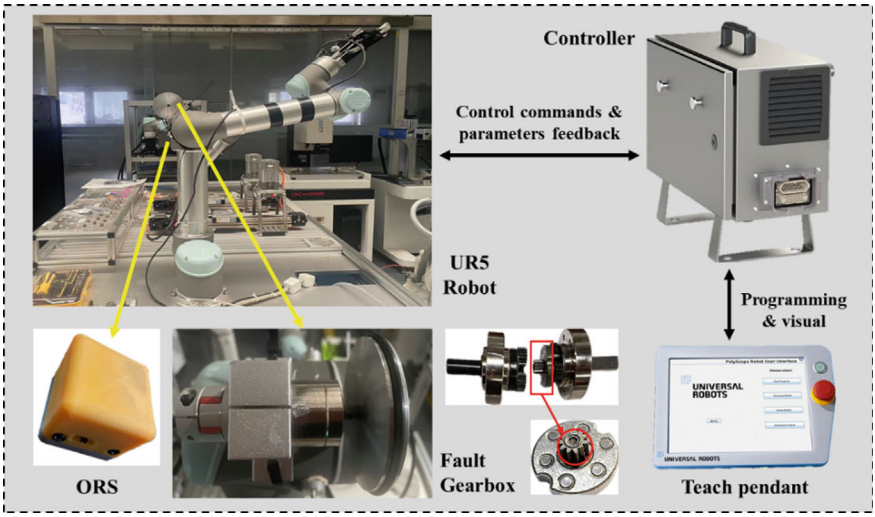


Fig. 3 Schematic diagram of experiments

In the course of our experimental endeavors, the data acquisition frequency was meticulously maintained around 8000 Hz, a rate that underscores our commitment to capturing data with utmost precision. Although this frequency slightly diverges from the initially intended benchmark, the resultant dataset is of remarkable richness, yielding invaluable insights into the nuanced performance and behavioral characteristics of the robotic system. The high-resolution data, despite the minor discrepancy, serves as a cornerstone for in-depth analysis, enriching our understanding of the robot's operational dynamics (Fig. 3).

On the one hand, in order to verify that ORS vibration signals can monitor the motion status of robot joints, four different rotating speed were used, namely 02, 03, 04 and 05 which means 36 deg/s, 54 deg/s, 72 deg/s and 90 deg/s. On the other hand, to verify the ability of ORS vibration signals to distinguish abnormal joint states, two different operating states were set. In each speed, normal condition is the baseline. In this case, no external force was added to the joint of the robot. The other is an abnormal state, which was simulated using an external planetary gearbox with a fault. The planetary gearbox is fixedly connected to the joint 3 of the robot and will rotate with the rotation of the joint 3. The ORS sensor is installed outside the joint to obtain status signals at different speeds. Figure 4 shows the time-domain signal under normal conditions.

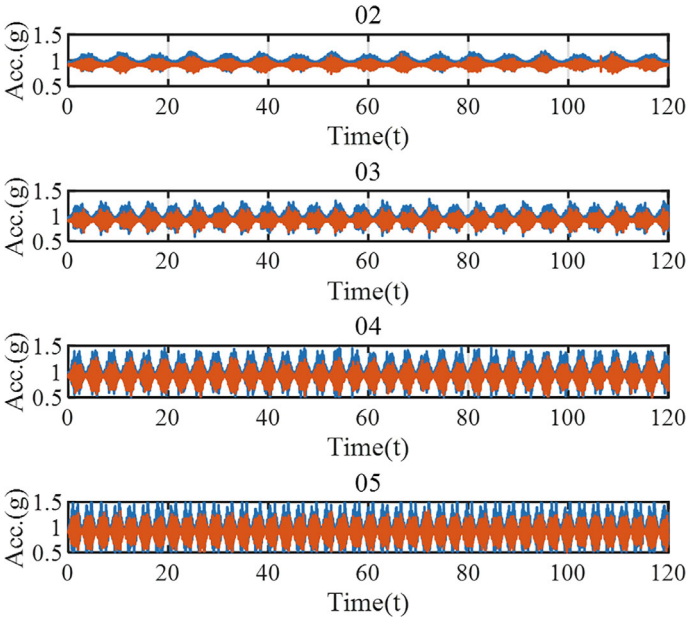


Fig. 4 Time stamp of ORS vibration signals of normal state

4 Data Analysis and Discussion

ORS vibration signals were used to track joint rotating speed and diagnose the abnormal running state. Based on the set speed of robot joint 3, the speed of the motor output shaft can be calculated according to the reduction ratio of the harmonic reducer, and the theoretical values of the characteristic frequencies of the reducer input shaft at four different set speeds can be calculated as 20.20 Hz, 30.30 Hz, 40.40 Hz, and 50.50 Hz, respectively. Based on time-domain signals at different speeds, corresponding FFT results can be obtained, as shown in Fig. 5, which are calculated based on vibration signals in the U direction. By querying the characteristic frequencies in Fig. 5, the actual signal characteristic frequencies at four different set speeds can be found to be 19.62 Hz, 29.64 Hz, 39.72 Hz, and 49.50 Hz, respectively. It can be seen that the characteristic frequency of the joint can be accurately obtained based on the ORS sensor signal, and dynamic analysis can be achieved based on the characteristic frequency.

From Fig. 5, it can be seen that at the frequency conversion position, the signal patterns of the two different operating states are consistent, with the amplitude in the abnormal operating state being smaller than that in the normal state. To further verify the regularity of this dynamic characteristic, Envelop and STFT were used for analysis. The calculation results of Envelop are shown in Fig. 6. It can be seen that the reflection of joint speed is also accurate. In terms of the amplitude characteristics

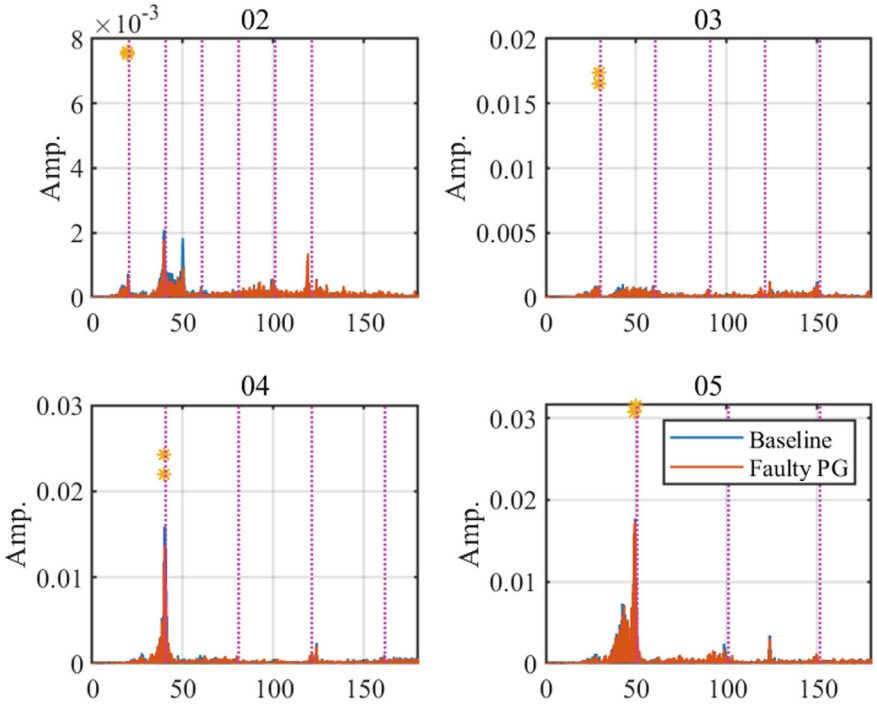


Fig. 5 FFT result of vibration signal of U axis

of the first-order rotation frequency and its sidebands, the amplitude under abnormal working conditions is smaller than that under normal conditions.

The calculation results of STFT are shown in Fig. 7. From Fig. 7a, it can be seen that different main frequency ranges can be observed at different speeds, and at the same speed, the characteristics of normal working state and abnormal working state are different. The amplitude characteristics can be obtained from Fig. 7b, and it can be seen that the results obtained by STFT are opposite to the methods discussed earlier, both showing that the amplitude under abnormal working conditions is greater than that under normal conditions.

The results indicate that the ORS three-axis accelerometer can perform dynamic analysis on industrial collaborative robots, accurately track joint speeds, and reflect certain fault patterns in the amplitude of characteristic frequencies based on dynamic analysis methods.

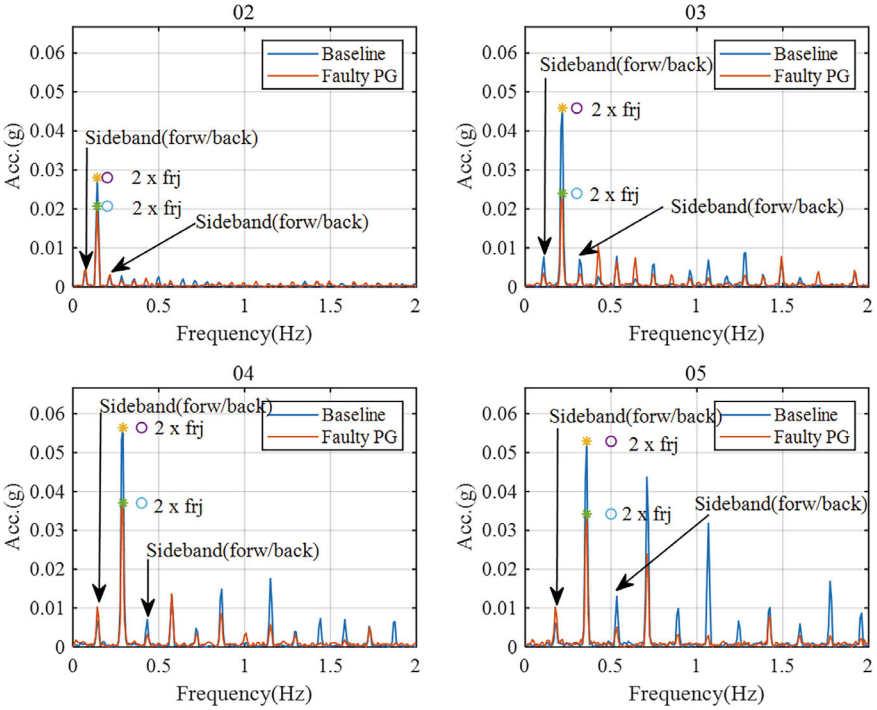


Fig. 6 Envelop of vibration signal of U axis

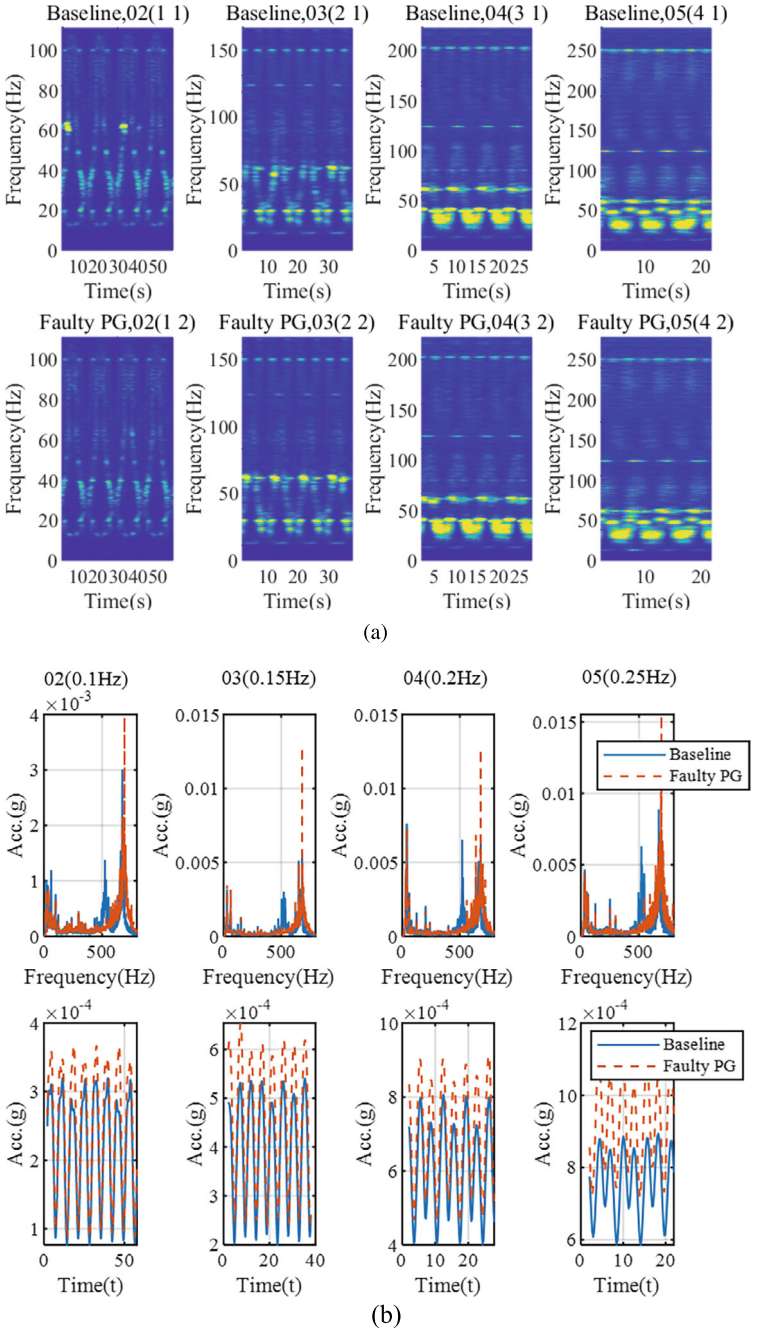


Fig. 7 STFT result of vibration signal of U axis

5 Conclusion

Structural failures in industrial robots frequently stem from exacerbated friction and wear within the critical joints, highlighting the pivotal role of proactive detection. This underscores the necessity for employing model-based fault detection methodologies, a strategy that enables the identification of nascent failure signs, thereby forestalling catastrophic structural breakdowns. The core objective of this research is to delve into a diagnostic methodology tailored to address abnormal operational states in industrial robots, leveraging the precision of model-based approaches. To facilitate this investigation, an experimental setup featuring an UR5 robot was meticulously constructed, designed to emulate a spectrum of abnormal running conditions, thereby providing a controlled environment for studying these critical issues. A planetary gearbox with sun gear broken tooth was used to simulate inner harmonic reducer fault of the joint. The results indicate that the use of ORS sensors can track the dynamic speed of robots in real time, and based on this, analyze the dynamic response to distinguish different operating states. The rotation frequency of input shaft of harmonic reducer with different speed can be seen in FFT spectrum. The amplitude of characteristic frequency can be used to monitor the running state of robot joint. This approach not only identifies impending equipment failures before they occur but also categorizes the severity of issues, enabling targeted maintenance actions. By doing so, it significantly reduces unplanned downtime—those unexpected halts in operations that can lead to substantial financial losses due to idle workforce and delayed production.

Acknowledgements This work was supported by Youth Innovative Talents Projects of Guangdong Province's University (2024KQNCX159) and the Quality Engineering Project of Beijing Institute of Technology, Zhuhai (2023053ZLGC).

References

1. Bevilacqua, M., et al.: Digital twin reference model development to prevent operators' risk in process plants. Sustainability (Basel, Switzerland) **12**(3), 1088 (2020)
2. Huang, J., Li, C., Chen, B.: Optimization design of RV reducer crankshaft bearing. Appl. Sci. **10**(18), 6520 (2020)
3. Nentwich, C., Reinhart, G.: A method for health indicator evaluation for condition monitoring of industrial robot gears. Robotics (Basel) **10**(2), 80 (2021)
4. Han, H., Shi, D., Gu, L., Wei, N., Gu, F.: Modelling and vibration signal analysis for condition monitoring of industrial robots. In: Proceedings of IncoME-VI and TEPEN 2021. Mechanisms and Machine Science, vol. 117. Springer, Cham (2023)
5. Xu, Y., Tang, X., Feng, G., Wang, D., Ashworth, C., Gu, F., Ball, A.: Orthogonal On-Rotor Sensing Vibrations for condition monitoring of rotating machines. J. Dyn. Monit. Diagn. **1**(1), 29–36 (2021)
6. Editorial Office, J.D.M.D., Fan, Z., Gao, R.X., He, Q., Huang, Y., Jiang, T., Peng, Z., Thévenaz, L., Xiong, Y., Zhong, S.: New sensing technologies for monitoring machinery, structures, and manufacturing processes. J. Dyn. Monit. Diagn. **2**(2), 69–88 (2023)

7. Han, S., Feng, Z.: Deep residual joint transfer strategy for cross-condition fault diagnosis of rolling bearings. J. Dyn. Monit. Diagn. **2**(1), 51–60 (2023)
8. Jaber, A.: Design of an Intelligent Embedded System for Condition Monitoring of an Industrial Robot. Newcastle University, United Kingdom (2016)
9. Lomakin, A., Deutscher, J.: Reliable algebraic fault detection and identification of robots. IEEE Trans. Autom. Sci. Eng. 1–17 (2022)
10. Isermann, R.: Model-based fault-detection and diagnosis—status and applications. Annu. Rev. Control. **29**(1), 71–85 (2005)

Fatigue Damage Assessment of FOWT Mooring Lines Using Sequence-to-Sequence Based Indirect Sensing



Rohit Kumar, Subhamoy Sen, and Arvind Keprate

Abstract In floating offshore wind turbines (FOWTs), mooring lines face fatigue degradation under wind and hydrodynamic loads, demanding real-time monitoring for safety. Conventional methods, like the TN or SN curve, mandate stress fluctuations under tensile loads to be measured from each mooring line via integrating costly underwater sensors, potentially inflating the cost of energy. This study introduces a sequence-to-sequence (Seq2Seq) surrogate model for real-time fatigue monitoring of a FOWT mooring system, offering a novel approach to estimating mooring line fatigue by connecting platform displacement to tension variations. A 5MW OC4 semi-submersible wind turbine model is used to simulate the dataset, train the Seq2Seq model, and establish the intricate connection between platform motion and tension fluctuations across diverse sea conditions. This surrogate model effectively captures the nonlinear relationship, allowing real-time predictions of fatigue damage in mooring lines.

Keywords Mooring lines · Fatigue · Seq2Seq · Deep learning · FOWT

1 Introduction

Floating offshore wind turbines (FOWTs) have revolutionized offshore wind energy by unlocking vast resources in deep waters, offering advantages like increased energy potential, resource stability, rapid deployment, and minimal environmental impact compared to fixed structures. The transition to FOWTs addresses issues associated with land usage, wildlife impact, and visual pollution, overcoming limitations of fixed structures confined to shallow depths. Despite these advantages, FOWTs encounter

R. Kumar · S. Sen (✉)
i4S Laboratory, Indian Institute of Technology Mandi, Mandi, Himachal Pradesh, India
e-mail: subhamoy@iitmandi.ac.in

A. Keprate
GrønnMet—Green Energy Lab, Department of Mechanical, Electrical and Chemical Engineering, Oslo Metropolitan University, Oslo, Norway

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_4

high operation and maintenance (O&M) costs, which significantly impact the overall cost of energy [10].

Recent efforts have focused on mitigating these costs through structural health monitoring (SHM) techniques, including damage detection and prediction in blade, tower, and catenary mooring predominantly with machine learning (ML)-based approaches [4]. Additionally, acoustic emission techniques are employed for gearbox, bearing, shaft, and blade fault detection [1]. Despite advancements, O&M costs owing to harsh operational environments for FOWTs have remained the primary concern necessitating continuous monitoring to ensure the operational status and condition of FOWT components.

Towards monitoring the fatigue in the catenary mooring, sampling cable tension directly has been undertaken by employing load cells or strain gauges [7]. This is important since a substantial portion of mooring chain failures is attributed to fatigue, occurring especially near fairleads [2]. Primarily due to installation complexities and accessibility issues related to maintenance, the sensors typically face issues leading to failure in harsh environmental conditions.

Contrary to the safety concern of the sensors patched on the mooring, their placement on floating platforms is perceived to be robust [7]. Accordingly, the recent attention in offshore wind energy has shifted towards integrating ML-based approaches for indirect sensing. These approaches are highly effective in capturing complex non-linear relationships where conventional methods often fall short. Moreover, once these relationships are established, the models can generate results quickly, facilitating real-time output. Considering these features, studies have explored the indirect prediction of tension time series from FOWT platform motion, employing advanced deep learning (DL) techniques like recurrent neural networks and deep neural networks (DNN) [6]. DNNs have already been employed extensively for damage detection utilizing insights derived from platform motions [6]. Nevertheless, unlike instant detection of other damage types, fatigue damage requires continuous monitoring of tension or strain time series [5]. Yet, there is a notable research gap in connecting platform motion responses with tension ranges, a linkage that could potentially streamline computational processes and enable real-time fatigue damage assessment.

A sequence-to-sequence (Seq2Seq) DL technique, known for handling long sequential input and output mapping [8, 13, 14], thus can establish a correlation between platform displacement time history (surge) and output tension ranges, represented as long arrays for a single mooring line section. Renowned for predicting data sequences, Seq2Seq models have been applied in various fields [9]. However, a primary limitation is the substantial data requirement for effective model training, posing challenges due to limited real-field data for FOWTs. To address this, numerical simulations recreate authentic FOWT behavior, aiming to capture environmental and structural influences. Despite these efforts, achieving optimal model performance under real conditions remains a significant challenge. The research objective is therefore to develop a Seq2Seq DL regression model. Trained using platform responses,

the model predicts tension ranges of the mooring line within designated time intervals. This innovative approach replaces the need for expensive sensors, contributing to a reduction in the overall cost.

The following section discusses the proposed seq2seq model, accompanied by a multitude of simulations over the FOWT model. The dataset generation process offers comprehensive details for constructing input and output datasets, followed by the subsequent presentation and discussion of results.

2 Methodology

2.1 Sequence-to-Sequence Model

Seq2Seq models belong to a class of neural networks designed for sequence prediction tasks. These models incorporate both an encoder and a decoder, enabling them to handle input and output sequences of varying lengths [8, 13]. In this architecture, both the encoder and decoder include layers of Long Short-Term Memory (LSTM) networks, making them well-suited for tasks where the relationship between input and output is not fixed.

Training. The Seq2Seq model during training is trained over an input dataset containing fixed-length input sequence X of time steps T given as, $\mathbf{X} = \{x_1, x_2, \ldots, x_t, \ldots x_T\}$.The sequence X is fed to the encoder which compresses the sequential information into a fixed-size context vector W representing the hidden state h_T and cell state C_T of the encoder's LSTM unit. The hidden state and cell state are updated at each time step as shown in Eq. (1).

$$h_t, C_t = LSTM_{encoder}(x_t, h_{t-1}, C_{t-1}) \tag{1}$$

After generating the context vector, the decoder starts with the context vector $W = (h_T, C_T)$, which initializes its LSTM states. The decoder is then trained over the original output y to produce a hidden state d_t and cell state e_t at each t^{th} step using the previous time step's original output $y_{(t-1)}$ as shown in Eq. (2).

$$d_t, e_t = LSTM_{decoder}(y_{t-1}, d_{t-1}, e_{t-1}) \tag{2}$$

The produced hidden output d_t then goes through a linear transformation to produce the final desired output $\hat{y}$, while updating their respective weights W_o and biases b_o over each iteration as shown in Eq. (3).

$$\hat{y}_t = Linear(d_t) = W_o d_t + b_o \tag{3}$$

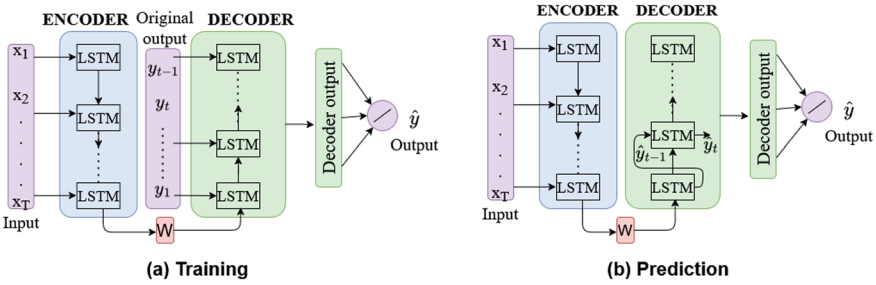


Fig. 1 Sequence-to-sequence model, (W = context vector)

Predictions. Once fully trained, the Seq2Seq model can predict any input sequence of time steps T by processing it through the encoder LSTM layers, compressing it into a fixed-size context vector W. Subsequently, the decoder LSTM layers generate the hidden output d_t at each t^{th} time step based on the context vector and the previously generated output $\hat{y}_{t-1}$ from the decoder as shown in Eq. (4).

$$d_t, e_t = LSTM_{decoder}(\hat{y}_{t-1}, d_{t-1}, e_{t-1}) \tag{4}$$

This final decoder output sequence is then fed to an output layer with a linear activation function to get the final prediction output $\hat{y}$. The overall training and prediction process is illustrated in Fig. 1a, b respectively.

The aforementioned model requires a large amount of data for training. However, in the case of FOWTs, such data is scarce thus synthetic or simulated data can serve as a viable alternative for training these DL models.

2.2 FOWT Model Description and Simulations

To fulfill the data requirements, NREL's OpenFAST software [3], a powerful simulation tool for wind turbines (WTs), was employed. In the current study, the OC4-DeepCWind semisubmersible platform model, equipped with a 5MW NREL wind turbine [11], was selected. This model features three catenary mooring lines spread 120°, as detailed in Table 1 and shown in Fig. 2. Equation (5) presents the dynamics of the FOWT platform:

$$(\mathrm{M_m} + \boldsymbol{AM_m})\ddot{\boldsymbol{y}} + \int_{-\infty}^{t} \boldsymbol{R_m}(\boldsymbol{t}-\boldsymbol{\tau})\dot{\boldsymbol{y}}(\boldsymbol{\tau})\boldsymbol{d\tau} + \boldsymbol{K_m y} = \boldsymbol{F_{wa}}(\boldsymbol{t}) + \boldsymbol{F_w}(\boldsymbol{t}) + \boldsymbol{F_v}(\boldsymbol{t}) + \boldsymbol{F_m}(\boldsymbol{t}) \tag{5}$$

Here, M_m is mass matrix, AM_m is added mass, and ÿ is acceleration, ẏ is velocity and y, is displacement, respectively, for six independent motion axes. R_m is retardation

Table 1 Catenary mooring system details

Specifications				
Mooring lines		3		
Diameter		0.0766 m		
Mass density		113.35 kg/m		
Axial stiffness		759.03 MN		
Unstretched length		835.35 m		
Line No.	Connection	X (m)	Y (m)	Z (m)
1	Anchor	418.80	725.38	− 200.00
	Fairlead	20.43	35.39	− 14.00
2	Anchor	− 837.60	0.00	− 200.00
	Fairlead	− 40.87	0.00	− 14.00
3	Anchor	418.80	− 725.38	− 200.00
	Fairlead	20.43	− 35.39	− 14.00

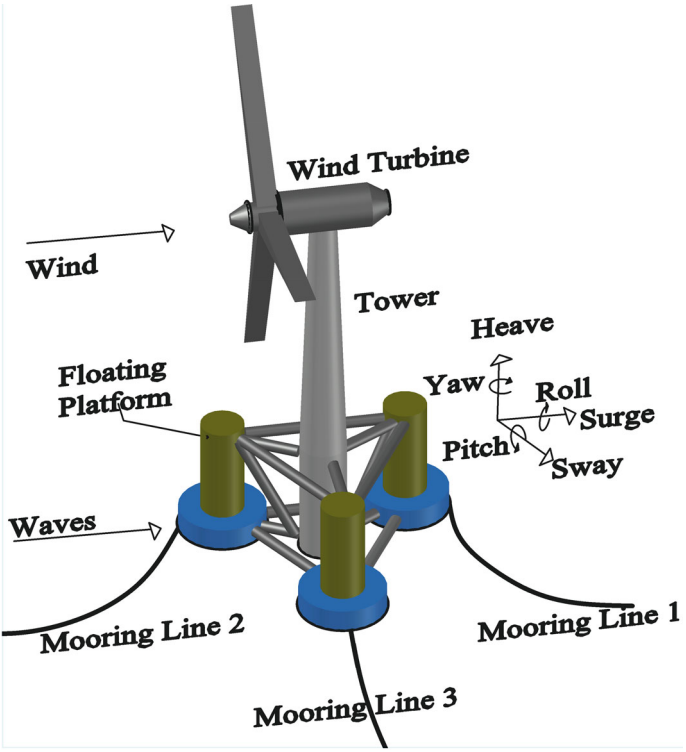


Fig. 2 OC4-DeepCWind semisubmersible platform model

function, F_{wa} represent wave excitation forces, F_w is the wind force, F_v is the drag force, and F_m represent the mooring force [10].

Time histories from OpenFAST simulations, captured at a fixed sampling frequency of 40 Hz, encompass platform motion (surge) and tension responses at the fairlead location of each mooring line. These data pertain to a single wind and wave direction. A total of 197 cases were simulated for a constant wind speed of 8 m/s along with different sea states as listed in Table 2, each signal extending over 4100 s. The initial 500 s, containing noise and transient portion, were excluded from the analysis.

2.3 Dataset Generation

Preprocessing is crucial in simulating raw data for the Seq2Seq-based model input and output datasets. The chosen model takes the platform surge time series sequence as input, as surge shows larger variations for any loads than any other platform displacements.

The DL model is LSTM-based, hence its limitations in recalling long sequences conflict with the need for accurate tension range estimation. Despite the initial signal creating a long sequence at a 40 Hz sampling rate, which could pose a risk of the model forgetting certain sequence patterns, experimentation with different time steps showed that a 1-s time step yielded satisfactory results during model training. Therefore, a sequence length of 20 s downsampled to 1 Hz was deemed a balanced choice, demonstrating effectiveness in providing nearly identical tension ranges for the same time interval (Fig. 3). The tension ranges in Fig. 3 are clubbed together in 10 *kN* intervals known as bins. The bin size is adopted from the study by Li et al. (2018), where a bin size resolution of 10 *kN* is assumed [5].

Output Dataset. Evaluating fatigue damage using the DL model in the mooring line requires correlating the platform's surge time history with the required output dataset (tension range and cycle). Using the rainflow counting method, tension history is analyzed to extract the output dataset. Hence, the output dataset was computed for each 20-s time interval of the tension history with the original 40 Hz sampling frequency. To ensure comprehensive representation, tension ranges with larger cycles are duplicated, attributing each tension range in the output dataset to 0.5 cycles. Normalization of all datasets is undertaken to confine values within the range of 0 and 1, ensuring precise representation during the training process. Considering a 20 s time interval for every 3600 s long 195 cases, resulted in a total of 35,100 samples. These samples were divided into an 80% training dataset and a 20% validation dataset. Additionally, 2 cases were considered as the test dataset for the loading conditions provided in Table 2.

DL model is trained over raw tension ranges (incorporating cycles) as output extracted from different 20 s intervals of tension signals for a single mooring line at

Table 2 Wave scatter diagram for the North Sea (Song and Wang [12])

$\frac{Tp(s)}{H_s(m)}$	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13..5	14.5	15.5	16.5	17.5	18.5	19.5
0.5	*	*	*	*	*	*	*	*	*	*	*						
1.5		*	*	*	*	*	*	*	*	*	*	*	*	*			
2.5		*	*	*	O	*	*	*	*	*	*	*	*	*	*		
3.5		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
4.5			*	*	*	*	*	*	*	*	*	*	*	*	*	*	
5.5			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
6.5			*	*	*	O	*	*	*	*	*	*	*	*	*	*	*
7.5				*	*	*	*	*	*	*	*	*	*	*	*	*	*
8.5				*	*	*	*	*	*	*	*	*	*	*	*	*	*
9.5					*	*	*	*	*	*	*	*	*	*	*	*	*
10.5						*	*	*	*	*	*	*	*	*	*	*	*
11.5						*	*	*	*	*	*	*	*	*	*	*	
12.5							*	*	*	*	*	*	*	*	*		
13.5								*	*	*	*	*	*	*	*		
14.5									*	*	*	*	*	*	*		
15.5									*	*	*	*	*	*	*		
16.5										*	*	*	*	*			

*: Training load cases, O: Testing load cases

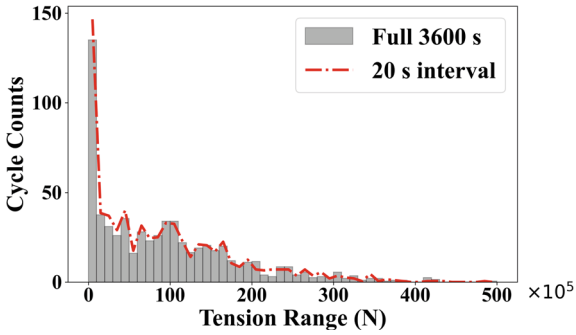


Fig. 3 Comparison plot of tension ranges obtained from 20 s time interval signals with 1 h signal

Table 3 Sequence-to-sequence model characteristics

Model architecture		
Layers	Activation function	Units
Encoder layer	–	128
Decoder layer	–	128
Output layer	Linear	1
Input sequence	–	20
Max. output length	–	23
Hyperparametres		
Optimizer		Adam
Epochs		200
Learning rate		0.0005

fairlead section, and surge as input. Based on the evaluation of training and validation loss, the model's architecture is outlined in Table 3.

3 Results and Discussion

The Seq2Seq model as discussed earlier is trained using the teacher forcing technique, wherein the ground truth from a prior time step is provided as decoder input along with input sequence. Once fully trained, predictions are made for any input sequence by connecting the predicted prior time step output as decoder input. Although the model was validated using 20 percent of the validation data, an additional 2 cases were used as testing load cases. The load cases outlined in Table 2 differed from the training datasets. The DL model demonstrated satisfactory performance with the testing datasets, as evident from the accuracy indices presented in Table 4.

The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics showcase the model's ability to make predictions with minimal error, as reflected in their low values when compared with the varying data range from 1 *kN* to 1089.3 *kN*.

Table 4 Accuracy indices of the trained Seq2Seq model

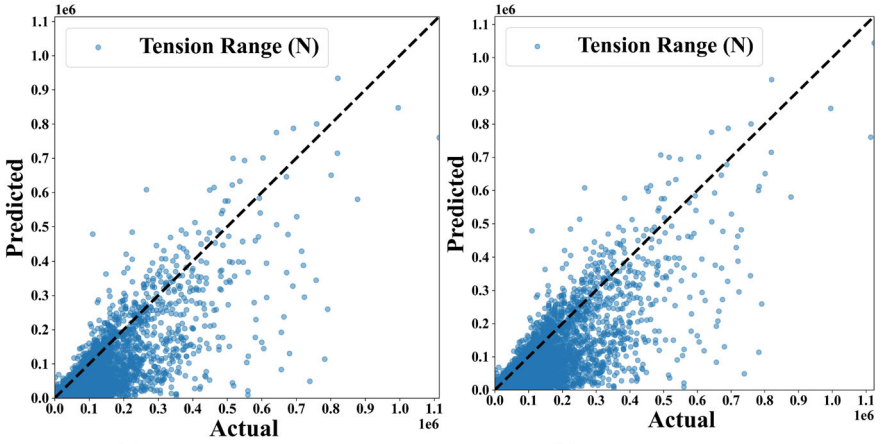
RMSE	MAE	R^2	CC
37.733 KN	17.7834 KN	0.70262	0.8183

Additionally, the Coefficient of Determination r^2 with a value of 0.70262 suggests a substantial portion of the variance in the dependent variable is captured by the model. Additionally, the correlation coefficient (CC) of 0.8183 highlights a strong positive correlation between the predicted and actual values.

The scatter plot presented in Fig. 4 illustrates the deviation of predicted raw tension ranges from the actual raw tension ranges for the testing load cases. The predicted raw tension ranges were grouped into bins of 10 *kN*, providing the respective number of cycles, as shown in Fig. 5. Overall, the DL model was able to predict the output using surge time history. Apart from capturing the output, the model also demonstrates high accuracy in predicting the output length.

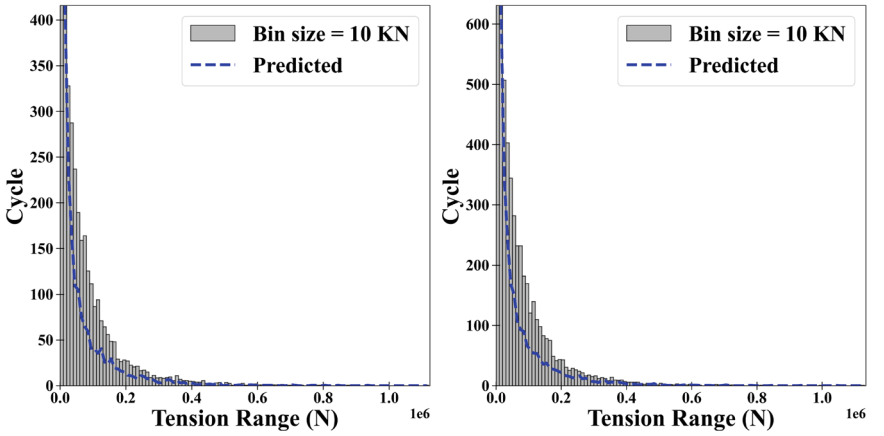
The observed dispersion in the scatter plot, particularly in the low-tension ranges, indicates reduced accuracy within this specific range. This can be attributed to the consideration of only one input sequence; incorporating combinations of other platform responses has the potential to yield improved results. Nevertheless, given that low tension ranges under the endurance limit of material are less likely to cause significant fatigue damage, this deviation is not expected to substantially impact the overall assessment.

In the current study, the DL model predicts damage for the fairlead, but its applicability can be expanded to assess damage at different positions of the mooring line. This enhancement can be achieved by incorporating labels for different locations for each mooring line.



(a) Scatter plot for Load case 1 **(b)** Scatter plot for Load case 2

Fig. 4 Scatter plot for test load cases



(a) Tension ranges and cycles for Load case 1 **(b)** Tension ranges and cycles for Load case 2

Fig. 5 True and predicted tension ranges and cycles for test load cases (Bin size = 10 *kN*)

4 Conclusion

In summary, employing Seq2Seq modeling to indirectly predict tension ranges of mooring line sections in the deep-sea using platform displacement sensor signal (surge) as inputs yielded effective results in this study. The indirect sensing DL model, trained using the teacher forcing technique, displayed satisfactory performance with the provided datasets, as evidenced by the accuracy indices. While the study has presented promising outcomes, ongoing endeavors to fine-tune the model and investigate additional factors could contribute to its further enhancement. Moreover, the application of Seq2Seq models for predicting tension ranges in diverse offshore wind scenarios holds the potential to advance the comprehension and optimization of floating offshore wind turbine systems.

References

1. Ciuriuc, A., Rapha, J.I., Guanche, R., Domìnguez-Garcìa, J.L.: Digital tools for floating offshore wind turbines (fowt): a state of the art. Energy Rep. **8**, 1207–1228 (2022)
2. Fontaine, E., Kilner, A., Carra, C., Washington, D., Ma, K., Phadke, A., Kusinski, G.: Industry survey of past failures, pre-emptive replacements and reported degradations for mooring systems of floating production units. In: Offshore Technology Conference, p. D041S047R002 (2014)
3. Jonkman, B., Mudafort, R.M., Platt, A., Branlard, E., Sprague, M., Jonkman, H., Vijayakumar, G., Buhl, M., Ross, H., Bortolotti, P., Masciola, M., Ananthan, S., Schmidt, M. J., Rood, J., Rdamiani, N., Sinolonghai, M. Hall, ashesh2512, kshaler, Bendl, K., pschuenemann, psakievich, ewquon, mattr-phillips, KUSUNO, N., alvarogonzalezsalcedo, Martinez, T., and rcorniglion, "Openfast/openfast: Openfast v3.1.0 (2022)

4. Kumar, R., Sen, S., Keprate, A.: Real-time fatigue assessment of Floating Offshore Wind Turbine Mooring employing sequence-to-sequence-based deep learning on indirect fatigue response. Ocean Eng. **315**, 119741 (2025)
5. Li, C.B., Choung, J., Noh, M.-H.: Wide-banded fatigue damage evaluation of catenary mooring lines using various artificial neural networks models. Mar. Struct. **60**, 186–200 (2018)
6. Mao, Y., Zheng, M., Wang, T., Duan, M.: A new mooring failure detection approach based on hybrid lstm-svm model for semi-submersible platform. Ocean Eng. **275**, 114161 (2023)
7. Minnebo, J., Aalberts, P., Duggal, A.: Mooring system monitoring using dgps. In International Conference on Offshore Mechanics and Arctic Engineering, vol. 45387, p. V01BT01A035 (2014)
8. Ounoughi, C., Yahia, S.B.: Sequence to sequence hybrid bi-lstm model for traffic speed prediction. Expert Syst. Appl. **236**, 121325 (2024)
9. Pham, T.Q.M., Im, S., Choung, J.: Prospects and economics of offshore wind turbine systems. J. Ocean. Eng. Technol. **35**(5), 382–392 (2021)
10. Piscopo, V., Scamardella, A., Rossi, G.B., Crenna, F., Berardengo, M.: Fatigue assessment of moorings for floating offshore wind turbines by advanced spectral analysis methods. J. Mar. Sci. Eng. **10**(1), 37 (2021)
11. Robertson, A., Jonkman, J., Masciola, M., Song, H., Goupee, A., Coulling, A., Luan, C.: Definition of the semisubmersible floating system for phase ii of oc4 (Tech. Rep.). National Renewable Energy Lab. (NREL), Golden, CO (United States) (2014)
12. Song, X., Wang, S.: A novel spectral moments equivalence based lumping block method for efficient estimation of offshore structural fatigue damage. Int. J. Fatigue **118**, 162–175 (2019)
13. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. **27** (2014)
14. Wang, Z., Qiao, D., Tang, G., Wang, B., Yan, J., Ou, J.: An identification method of floating wind turbine tower responses using deep learning technology in the monitoring system. Ocean Eng. **261**, 112105 (2022)

Temporal Feature Extraction Based Real Time Damage Detection of Floating Offshore Wind Turbine Mooring Lines

Rohit Kumar, Ananay Thakur, Subhamoy Sen, and Arvind Keprate

Abstract Floating Offshore Wind Turbines (FOWT) are currently facing the challenge of high cost of energy. Mooring lines holding the FOWTs in the deep sea are usually exposed to harsh and dynamic wind and wave loadings, which are responsible for damage among mooring lines. The physical sensors deployed for online monitoring ultimately increase the cost of energy. This study proposes a data-driven LSTM-ANN method to classify the amount of damage among mooring lines. The Long Short-Term Memory (LSTM) network identifies patterns in platform time history responses, which the Artificial Neural Network (ANN) then uses to classify the damage levels. Due to the lack of real data, the model is trained on simulated data from the 5MW-OC4 semisubmersible model available in NREL's OpenFast. The model achieved a classification accuracy of 90%, demonstrating its ability to learn the behaviour of platforms with mooring damage. This approach provides a cost-effective online monitoring solution and enhances the reliability and safety of FOWTs by enabling timely maintenance and reducing downtime.

Keywords FOWT · LSTM-ANN · Mooring lines · Real-time damage detection · Online monitoring

1 Introduction

Floating offshore wind turbines (FOWTs) are promising clean energy sources in the renewable energy sector. However, their large-scale deployment is largely hindered by high cost of energy, with maintenance and operation (M&O) expenses comprising approximately 32% of the total cost [1]. These massive structures float over the

R. Kumar · A. Thakur · S. Sen (✉)
i4S Laboratory, Indian Institute of Technology Mandi, Mandi, Himachal Pradesh, India
e-mail: subhamoy@iitmandi.ac.in

A. Keprate
Department of Mechanical, Electronics and Chemical Engineering, Oslo Metropolitan University, Oslo, Norway

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_5

sea, anchored in place by mooring lines that are continuously subjected to various dynamic loads.

A case study of 40 offshore structures concluded that failure among mooring lines is largely due to corrosion [2]. Corroded moorings, if not properly monitored, can fail before their operational life is over, leading to substantial damage to the superstructure and significant economic losses. Therefore, it is essential to continuously monitor these lines in real-time to detect damage early and mitigate the associated risks.

Current methods for continuously monitoring mooring line damage involve costly sensors that are challenging to install and maintain in the deep sea. As a result, researchers are focusing on reducing dependence on these sensors, which are susceptible to failure in harsh marine environments [3]. One promising approach is to use the motion response of the platform to predict other responses, thereby eliminating the need for expensive deep-sea sensors [4]. Platform sensors are generally more robust and less prone to failure. This indirect sensing is primarily achieved through machine learning (ML) and deep learning (DL) techniques, which utilize more reliable sensor signals of the structure to estimate required responses. Techniques include particle filters [5], artificial neural networks (ANNs) [6] and convolutional neural networks (CNNs) [7] for identifying types of damage in mooring lines. These methods use platform response data as input for their models, providing more resilient and cost-effective virtual sensing. However, these DL approaches classify the type of damage but do not quantify the extent of damage, which is crucial for scheduling maintenance and reducing downtime. This creates a research gap in mapping platform motion to different damage levels in mooring lines. To address this gap, a more advanced DL approach can be employed, focusing on detecting unique patterns from raw platform responses. These patterns can then be used to accurately identify damage in mooring lines, ensuring reliable monitoring across various sea conditions.

This study introduces an *LSTM-ANN* framework implemented within a Python-based tool for classifying mooring line damage. In this framework, the Long Short-Term Memory (LSTM) model extracts unique patterns from raw platform responses, while the ANN classifies the intensity of damage among mooring lines. These data-driven models require a lot of data to understand and accurately predict the patterns associated with different damage scenarios. Due to the limited deployment of FOWTs, real-world data is unavailable [8]. Therefore, the DL model is trained and tested using simulated data. Simulations are performed on a 5MW-OC4 semisubmersible model available in NREL's OpenFAST [9]. By utilizing the capabilities of LSTM for pattern recognition and ANN for classification, this study provides a cost-effective and reliable solution for real-time monitoring of mooring line damage. This proactive approach offers the advantage of online continuous monitoring while reducing the costs associated with expensive sensors.

The following section provides a fundamental overview of the DL method used in this study, along with a detailed explanation of dataset generation and model training using simulated data based on the FOWT model. This is succeeded by a presentation of the findings along with a discussion of their significance, ultimately leading to the real-world applications of the approach.

2 Methodology

The proposed method in this study is a data-driven approach that integrates LSTM networks with ANN. LSTM is renowned for its ability to remember sequences and extract important features. LSTM cells extract the sequential information based on three gates: the input, forget, and output gates. Mathematically, the working of these gates for any input x_t at any t^{th} time step can be given as follows:

The forget gate selects cell state information to reject:

$$f_t = \sigma\left(A_f \cdot \left[h_{t-1}, x_t\right] + b_{\text{fgt}}\right) \tag{1}$$

The input gate decides which values from the input to update the cell state with:

$$i_t = \sigma\left(A_i \cdot \left[h_{t-1}, x_t\right] + b_{\text{igt}}\right) \tag{2}$$

$$\tilde{S}_t = \tanh\left(A_c \cdot \left[h_{t-1}, x_t\right] + b_{\text{ct}}\right) \tag{3}$$

The cell state is updated as follows:

$$S_t = f_t \cdot S_{t-1} + i_t \cdot \tilde{S}_t \tag{4}$$

The output gate governs the selection of information that will shape the next hidden state:

$$o_t = \sigma\left(A_o \cdot \left[h_{t-1}, x_t\right] + b_{\text{ogt}}\right) \tag{5}$$

$$h_t = o_t \cdot \tanh(S_t) \tag{6}$$

Here, σ represents the sigmoid function, A_f, A_i, A_c and A_o represent the weight matrices for the forget gate, input gate, cell state update, and output gate, respectively. $b_{\text{fgt}}, b_{\text{igt}}, b_{\text{ct}}, b_{\text{ogt}}$ are the corresponding bias vectors. x_t represents the input at time step t, h_{t-1} represents the hidden state of the previous time step, and S_t represents the updated cell state at time step t.

Inside the LSTM network, for any given input sequence $\mathbf{X} = \{x_1, x_2, ..., x_t, ..., x_T\}$, the LSTM layer extracts the features at each time step t and creates a final hidden state output $\mathbf{h} = \{h_1, h_2, ..., h_d\}$, where d is the number of LSTM cells in the last layer. This final hidden output from the LSTM layer is fed to the ANN hidden layer. The ANN processes these features and performs the final classification using the softmax activation function in the output layer. The softmax function assigns probabilities to each class label, facilitating the classification process as shown:

$$\hat{p}_i = \frac{\exp(z_i)}{\sum_{i=1}^{\mathcal{K}} \exp(z_i)} \tag{7}$$

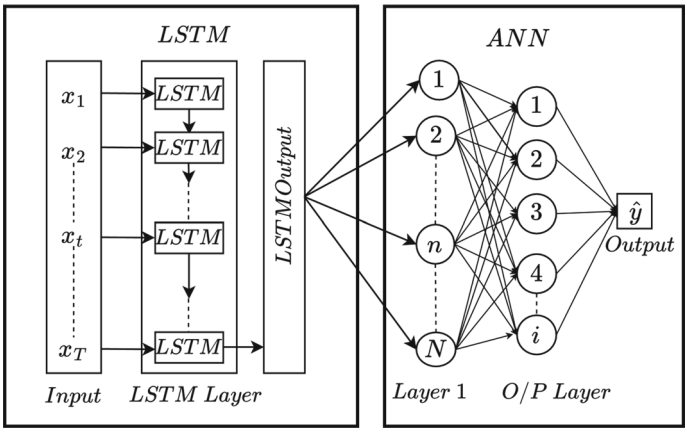


Fig. 1 LSTM ANN framework

where $\hat{p}_i$ is the predicted probability for class i, z_i is the input to the softmax function for class i, and $\mathcal{K}$ is the total number of classes.

The model is refined by adjusting the associated weights and biases to reduce the loss function, which is represented by the cross-entropy loss as shown in Eq. (8):

$$\mathcal{L} = -\frac{1}{M}\sum_{j=1}^{M}\sum_{c=1}^{\mathcal{K}} p_{j,c}\log\left(\hat{p}_{j,c}\right) \tag{8}$$

where M is the total number of samples, $\mathcal{K}$ is the total number of classes, $p_{j,c}$ is the true label, and $\hat{p}_{j,c}$ is the predicted probability for class c.

By using the sequence pattern-capturing capabilities of LSTM, the model extracts crucial information, which the ANN then uses to classify the output based on the provided labels. Figure 1 provides an overview of the proposed framework.

2.1 FOWT Model

The DL model discussed above requires huge datasets for training. The data requirements are fulfilled by simulating the OC4-DeepCWind semisubmersible FOWT model shown in Fig. 2. The FOWT model is equipped with a 5MW NREL wind turbine [10] anchored in a 200 m deep-sea region using three catenary mooring lines attached 120° apart. The detailed specifications of the mooring lines are given in Table 1. The simulations were implemented employing the use of the OpenFAST software [9], which includes various modules for generating forces and estimating responses. The *InflowWind* and *HydroDyn* modules generate steady wind and wave conditions, respectively. *HydroDyn* also handles the hydrodynamic loading on the

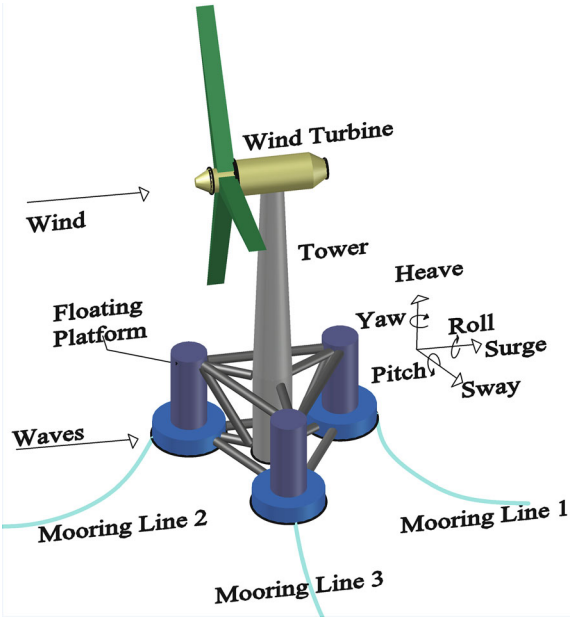


Fig. 2 OC4 Semisubmersible 5MW FOWT

Table 1 Specifications of mooring lines in the undamaged state

Property	Value	Unit
Diameter	0.0766	m
Mass density	113.35	kgm^{-1}
Axial stiffness	7.5903×10^8	N
Unstretched length	835.35	m

platform, while *MoorDyn* addresses the loading related to moorings. *AeroDyn* and *ElastoDyn* deal with aerodynamic and structural loads, respectively.

2.2 Dataset Generation

To ensure the real loading scenario for the simulated dataset, the loading condition assumes a constant 8 m/s wind speed achieved using *InflowWind* module and a sea state representing sea behaviour having a wave height (H_s) of 6 m and a wave period (T_p) of 10 s. Different waveforms for a single sea state are generated using the JONSWAP energy spectrum by providing different wave seed values in *HydroDyn* module. To ensure the DL model is trained over a variety of damaged mooring line scenarios, 22 different cases shown in Table 2 were simulated. The required damage

Table 2 Damaged mooring lines (ML) cases

Case	ML1(%)	ML2(%)	ML3(%)	Case	ML1(%)	ML2(%)	ML3(%)
K_1	0	0	0	K_{12}	15	15	0
K_2	10	0	0	K_{13}	20	20	0
K_3	15	0	0	K_{14}	0	10	10
K_4	20	0	0	K_{15}	0	15	15
K_5	0	10	0	K_{16}	0	20	20
K_6	0	15	0	K_{17}	10	0	10
K_7	0	20	0	K_{18}	15	0	15
K_8	0	0	10	K_{19}	20	0	20
K_9	0	0	15	K_{20}	10	10	10
K_{10}	0	0	20	K_{21}	15	15	15
K_{11}	10	10	0	K_{22}	20	20	20
0: Undamaged				10: 10% Damage			

in the mooring lines is induced by reducing their axial stiffness in the material properties of the mooring lines available in *MoorDyn* module.

Each simulated response was simulated over 3600 s, containing floating platform responses in all 6 *dofs*, i.e. surge, sway, heave, pitch, roll, and yaw sampled at a sampling frequency of 40 Hz. For each damage case, 100 different simulations were recorded for 100 different waveforms. For all 22 scenarios as outlined in Table 2, a total of 2200 samples each having a 3600-s duration were generated. To ensure LSTM extracts the information from an input sequence efficiently a 60 s input dataset window is assumed. This leads to the total dataset of 132,000 samples. This dataset is split into three segments: 70% designated for training, 15% reserved for validation, and the remaining 15% utilized for testing the DL model.

2.3 DL Model Training

The LSTM-ANN model is trained over a training dataset containing all 6 *dofs*. The input sequence of length 60 s down sampled to 2 Hz was fed to the LSTM layer. It was done to get an identifiable pattern while also not making the sequence too long. The LSTM layer extracts the information in the form of an output vector, which is then used by the ANN hidden layer to classify the output accurately. The final architecture of LSTM-ANN model detailed in Table 3 is decided with continuous monitoring of training and validation loss. The input layer has 6 features, representing 6 *dofs*, and the output layer has 22 output nodes, which provide the predicted probability distribution from the network, classifying the most probable class.

Table 3 The architecture of the LSTM-ANN model

LSTM-ANN architecture		
Layer	Activation function	Units
Input layer	–	120
LSTM	–	256
ANN	tanh	512
Output layer	softmax	22
Hyperparameters		
Optimizer		Adam
Epochs		300
Time taken		600 min
Learning rate		10^{-5}

3 Results and Discussion

The proposed model exhibited robust performance on the testing dataset, attaining an overall accuracy of 90% indicating the effective categorisation, which can be seen in the confusion matrix detailed in Fig. 3. The confusion matrix provides a detailed performance analysis of the classifier when evaluated on a test dataset with known true labels.

The detailed performance metrics for each class which shed light on the model's performance, encompass key measures like precision, recall, and F1-score, are summarized in Table 4.

It is observed that the precision is high, with most values greater than 0.8 indicating the DL model is effective in correctly identifying these damage types, with minimal false positives for the above classes. High recall values indicate that the actual positives are correctly identified by the model for the same classes, resulting in a low false negative rate. The F1-score also being above for many cases also indicates that the model maintains a good balance between precision and recall and signifies its robust performance. This implies that the model accurately identifies a high proportion of positive cases (high recall) while minimising false positives (high precision).

For the cases K_{19} and K_{20} denoting 15% and 20% damage scenarios respectively in all mooring lines, the precision is 0.49 and 0.46, recall is 0.43 and 0.52, while the F1-score is 0.45 and 0.49 respectively. These lower values imply less robustness of the model's sensitivity to these specific damage types. It is found that most of the K_{19} samples were misclassified as K_{20} and K_{21}; similarly, for K_{20} misclassified samples were in K_{19} and K_{21} as shown in Fig. 3. This could be due to the resemblance in sequence pattern for the given input sequence to LSTM, the problem can be fixed if the sequence length for LSTM is further increased along with an adequate sampling frequency. Of course, this will also create an attention-based algorithm to remember longer sequence lengths. Except for these two cases, the performance metrics imply

True Label \ Predicted Label	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}	K_{11}	K_{12}	K_{13}	K_{14}	K_{15}	K_{16}	K_{17}	K_{18}	K_{19}	K_{20}	K_{21}	K_{22}
K_1	365	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K_2	0	370	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K_3	0	0	381	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K_4	0	0	0	386	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	7	0
K_5	0	0	0	27	269	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K_6	0	0	0	0	0	359	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K_7	0	0	0	0	0	0	357	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
K_8	0	0	0	0	0	0	1	379	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K_9	0	0	0	0	0	0	0	7	379	0	0	0	0	0	0	0	0	0	0	0	0	0
K_{10}	0	0	0	0	0	0	0	0	0	389	0	0	0	0	0	0	0	0	0	0	0	0
K_{11}	0	0	0	0	0	0	0	0	0	0	408	4	0	0	0	0	0	0	0	0	0	0
K_{12}	0	0	0	0	0	0	0	0	0	0	7	396	0	0	0	0	0	0	0	0	0	0
K_{13}	0	0	0	0	0	0	0	0	0	0	0	0	353	48	0	0	0	0	0	0	0	5
K_{14}	0	0	0	0	0	0	0	0	0	0	0	0	51	274	1	0	0	0	0	0	0	0
K_{15}	0	0	0	0	0	0	0	0	0	0	0	0	0	18	365	0	0	0	0	0	0	0
K_{16}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	354	2	0	0	0	0	15
K_{17}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	413	7	0	0	0	0
K_{18}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	374	0	0	0	0
K_{19}	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	157	125	13	73
K_{20}	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	81	189	85	7
K_{21}	1	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	62	292	0
K_{22}	1	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	80	31	0	1106

Fig. 3 Confusion matrix

Table 4 Performance evaluation for classification

Case	Precision	Recall	F1-score	Samples	Case	Precision	Recall	F1-score	Samples
K_1	0.99	0.95	0.97	384	K_{12}	0.99	0.98	0.99	403
K_2	0.95	0.96	0.96	384	K_{13}	0.86	0.98	0.99	406
K_3	0.96	1.00	0.98	381	K_{14}	0.81	0.87	0.86	326
K_4	0.89	0.94	0.91	410	K_{15}	0.99	0.84	0.82	383
K_5	0.96	0.91	0.93	296	K_{16}	0.94	0.95	0.97	371
K_6	0.99	1.00	1.00	359	K_{17}	0.97	0.95	0.95	441
K_7	1.00	0.99	1.00	359	K_{18}	0.98	0.94	0.95	386
K_8	0.98	1.00	0.99	380	K_{19}	0.49	0.97	0.98	369
K_9	1.00	0.98	0.99	386	K_{20}	0.46	0.52	0.49	364
K_{10}	1.00	1.00	1.00	389	K_{21}	0.74	0.77	0.75	381
K_{11}	0.98	0.99	0.99	412	K_{22}	0.92	0.90	0.91	1225
Overall accuracy: 90%					Total test samples: 9195				

that the model is very effective at correctly identifying these classes with minimal false positives and false negatives leading to consistent and accurate classification.

In comparison to recent studies, such as those by Gorostidi et al. [11] and Chung et al. [12], using ANNs, reported accuracies around 90%. Additionally, [13] employed a CNN, achieving an accuracy greater than 90%. However, these studies primarily classify different types of damage, such as biofouling, mooring failure, and corrosion in mooring lines. In comparison, the current study also provides an accuracy of 90%, but it goes further by detecting not only the presence of damage but also the extent of damage in mooring lines. This advancement aids in optimizing scheduled maintenance, thereby reducing downtime.

4 Conclusion

The proposed model demonstrated exceptional performance in classifying damage under various loading conditions. It effectively captures sequential information, significantly aiding the ANN in accurately classifying damage. This model shows great potential for online monitoring using physical sensors installed on a floating platform, ultimately reducing downtime and enhancing maintenance efficiency. Additionally, this approach can be extended to accommodate different loading conditions and more precise damage intervals, offering a robust solution for the continuous monitoring of offshore structures. A key limitation is the computational expense of training the LSTM-ANN model and its difficulty with long sequences, as LSTM tends to forget over time. Future work could incorporate attention mechanisms to handle longer inputs while retaining crucial information. Additionally, this method can be scaled to larger wind turbines, such as the 15MW models expected to dominate in the future.

References

1. Nava, V., Ruiz-Minguela, P., Perez-Moran, G., Rodriguez Arias, R., Lopez Mendia, J., Villate-Martinez, J.L.: Installation, operation and maintenance of offshore renewables, pp. 397–424, (Institution of Engineering and Technology) (2019)
2. Fontaine, E., Kilner, A., Carra, C., Washington, D., Ma, K., Phadke, A., Laskowski, D., Kusinski, G.: Industry survey of past failures, pre-emptive replacements and reported degradations for mooring systems of floating production units. In: Offshore Technology Conference, p. D041S047R002, OTC, (2014)
3. Salic, T., Charpentier, J.F., Benbouzid, M., Le Boulluec, M.: Control strategies for floating offshore wind turbine: challenges and trends. Electronics **8**(10), 1185 (2019)
4. Wang, K., Gaidai, O., Wang, F., Xu, X., Zhang, T., Deng, H.: Artificial neural network-based prediction of the extreme response of floating offshore wind turbines under operating conditions. J. Mar. Sci. Eng. **11**(9), 1807 (2023)

5. Kumar, R., Thakur, A., Shereena, O., Keprate, A., Sen, S.: Characterizing damage in wind turbine mooring using a data-driven predictor model within a particle filtering estimation framework. PHM Soc. Eur. Conf. **8**, 8–8 (2024)
6. Mao, Y., Wang, T., Duan, M.: A DNN-based approach to predict dynamic mooring tensions for semi-submersible platform under a mooring line failure condition. Ocean Eng. **266**, 112767 (2022)
7. Janas, K., Milne, I., Whelan, J.: Application of a convolutional neural network for mooring failure identification. Ocean Eng. **232**, 109119 (2021)
8. Otter, A., Murphy, J., Pakrashi, V., Robertson, A., Desmond, C.: A review of modelling techniques for floating offshore wind turbines. Wind Energy **25**(5), 831–857 (2022)
9. Jonkman, B., Mudafort, R.M., Platt, A., Branlard, E., Sprague, M., et al.: OpenFAST/OpenFAST: OpenFAST v3.1.0 (2022)
10. Robertson, A., Jonkman, J., Masciola, M., Song, H., Goupee, A., Coulling, A., Luan, C.: Definition of the semisubmersible floating system for phase II of OC4. Tech. Rep. National Renewable Energy Lab. (NREL), Golden, CO (United States) (2014)
11. Gorostidi, N., Nava, V., Aristondo, A., Pardo, D.: Predictive maintenance of floating offshore wind turbine mooring lines using deep neural networks. J. Phys. Conf. Ser., **2257**, 012008, IOP Publishing (2022)
12. Chung, M., Kim, S., Lee, K., et al.: Detection of damaged mooring line based on deep neural networks. Ocean Eng. **209**, 107522 (2020)
13. Sharma, S., Nava, V.: Condition monitoring of mooring systems for floating offshore wind turbines using convolutional neural network framework coupled with autoregressive coefficients. Ocean Eng. **302**, 117650 (2024)

An Energy Management Strategy for Piezoelectric Energy Harvester for Low-Power Applications



Naveen Tyagi, Anshu Mli Gaur, and Rakesh Mishra

Abstract This paper presents a novel design of a piezoelectric cantilever beam for energy harvesting applications. Finite element simulations have been carried out to evaluate the effect of variation of geometrical parameters on the generated voltage output. The effect of thickness and area of the designed piezoelectric structures on the output voltage has been investigated for different piezoelectric materials to select the optimal material for low-power applications. The simulation results have revealed the Aluminum nitride and PZT-7A as potential materials for PEH devices due to a higher output voltage generation (30–35 V) capability. The results have showcased the appropriate selection of piezoelectric materials. It has been demonstrated that the geometrical parameters of the beam have a substantial effect on the output voltage generation of PEH. In addition to the Finite element simulation-based design, an energy management strategy has been implemented to optimize the battery charging/discharging process. An IoT system with sensors has been proposed to monitor the real-time battery conditions and manage the output voltage generation of PEH. The designed system can optimally maintain the healthy charging voltage of the battery in the range of 3.6–3.7 V without overheating. The results from the simulation of the energy management system have showcased the capability to deliver a load of up to 0.05 watts suitable for low-power sensing applications.

Keywords Energy harvesting · Finite element method · Piezoelectric · Energy management · Piezoelectric materials

N. Tyagi
Department of Electronics and Communication, Maharaja Agrasen University, Baddi, India

A. M. Gaur (✉) · R. Mishra
Department of Engineering, School of Computing and Engineering, University of Huddersfield, Huddersfield, UK
e-mail: a.m.gaur@hud.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_6

1 Introduction

Vibration-based energy harvesting has the potential to revolutionize the way continuous machines and structural health monitoring devices are powered in a variety of applications. To maximize the amount of energy captured, the resonance frequencies of an energy harvesting device should generally match the ambient vibration frequencies. Due to their ideal power densities and lack of need for a supplementary power source, piezoelectric energy harvesting (PEH) devices have attracted much interest in the last decade. Adhikari et al. [1] describe the work on the PEHT (piezoelectric energy harvester tile) spring-damper model and demonstrate the effect of parameters on the piezoelectric tile performance. This research investigates the methods to improve piezoelectric performance in tiles by choosing the right piezoelectric ceramics and relaxors. In another work, Li et al. [2] describe a piezoelectric harvester with multi-mass multiple cantilever beams that have a lower resonant frequency, bandwidth, and higher output voltage and power than a piezoelectric harvester with multiple-mass single cantilever beam configuration. Elachi et al. [3] have discussed the effect of several piezoelectric materials on the performance [such as Lead Zirconate Titanate (PZT-5A)] and Barium Titanate attached to different nonpiezoelectric materials [such as Aluminium (Al) and fibre glass]. In this investigation, the maximum output power has been attained when a PZT-5A patch of shorter length has been bonded with an Al patch of longer length. Similarly, Raju et al. [4] discovered that an energy harvesting structure with a trapezoidal cavity with a 1.5-degree taper angle provides higher voltage than structures with rectangular cavities of 2, 2.5, 3, and 3.5 mm cavities based on analytical and experimental analysis. Mangaiyarkarasi et al. [5] have provided a framework for piezoelectric structures with eight various proof mass shapes (Square, Rectangular, L, H, E, Cylinder, p, and T forms) placed at different locations (centre, left, and right), as well as multiple proof masses (Two proof mass and Three proof mass). The power generated by the piezoelectric sensor at a lower frequency has been used to power mobile phone gadgets. Hanif et al. [6] have presented the development of a specially built rotational piezoelectric energy harvester prototype consisting of a 40 mm piezoelectric bimorph beam. The PEH rotates with the help of a rotor and aluminium proof-mass. They have reported that the maximum voltage output could be achieved with an input frequency of 18 Hz.

Some researchers have developed various PEH structures to utilize wind energy for energy harvesting. A PZT-based transducer for multi-direction wind energy harvesting has been designed and realized by Zhou et al. [7]. The developed PEH has a high wind energy utilisation rate in all directions. He et al. [8] have reported a size-dependent micro-beam-based hybrid energy harvester using LASMP (Light activated shape memory polymer). Similarly in another research, Yuan et al. have demonstrated an adjustable hybrid energy harvester made up of two smart materials and a thin plate. The use of LASMP in energy harvesting is effective, superior, and promising due to its unique features, such as dynamic stiffness and noncontact approach [9]. According to Raju et al. [10], a harvester with a non-uniform cross-section near the root produces 76.9% more voltage than a harvester with a uniform

cross-section. The generated voltage has been increased by 91.3% with a rectangular cavity. The benefits of perforation in the cantilever structure have been highlighted by Anand et al. [11]. They have discovered that a multi-perforated cantilever structure produces more output voltage and power than a standard cantilever structure without increasing the size of the cantilever structure. Chen et al. [12] have developed a prototype harvester with the generated average power of 43.64 μW gathered at 4 km/h. This experiment exhibits that the PEH can power a wireless transmitter.

The piezoelectric energy harvester system consists of three main components: a harvester, a buck converter, and a battery/load. So, there is always mismanagement of the power generated and power delivered due to load current variation. For such challenges, an energy management scheme is quite essential to mitigate this problem. In the literature review, few energy management schemes have been worked out. Li et al. have designed an energy flow model to understand the complex flow of power from source to load [13]. Similarly, Omar et al. have proposed a battery-less power management strategy for PEH that has been used for wearable systems. Kassan et al. have proposed a self-powered management system model for piezoelectric harvesters for wireless applications [14]. Amar et al. have designed a self-powered Energy Management Circuit for Piezoelectric Energy Harvesting based on Synchronized Switching Technology for low-frequency applications [15]. However, no one successfully provides an energy management strategy for wide frequency applications. So, in this research work, a simulation-based energy management system has been designed and proposed, which has the capability of working from 10 to 150 Hz.

From all the above discussions, the generated output voltage could be enhanced by varying the geometrical configuration of PEH. Also, the thickness of the piezoelectric cantilever beam is a vital parameter to consider when developing and designing a PEH. The generated power could be increased by choosing an appropriate material for PEH design. For this purpose, a novel PEH has been presented, designed, and simulated to select the material and geometry configuration to maximise output. The article has been organised as follows. Section I discusses the literature review for the piezoelectric harvester systems. Section II focused on the methodology and design consideration for the PEH. This section involves an energy management model for optimizing the charging/discharging of the battery. The results are provided in Section III, followed by the conclusion and future scope in the final section. This investigation will assist in designing a piezoelectric cantilever beam energy management system for optimal throughput power generation for low-power applications.

2 Methodology and Design Considerations

The proposed energy harvesting system comprises multiple components such as PEH, vibration source, full wave rectifier, buck converter, and battery charging circuit (Fig. 1). The selection of rectifier and buck converter depends upon the load capacity as well the amount of energy harvested with the PEH, and the other important parameter for consideration is low threshold voltage. The rectifier (35 A, 600 V)

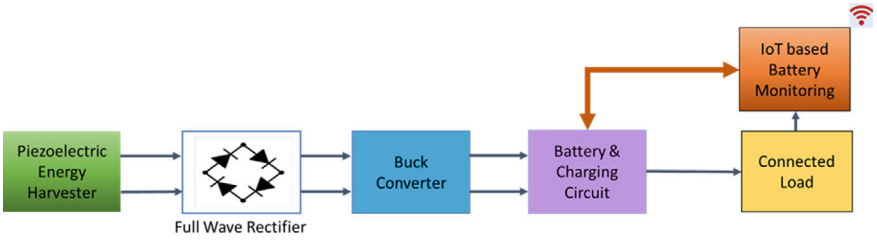


Fig. 1 Block diagram of proposed energy management system for piezoelectric energy harvester

has been used as a high-performance, followed by a filter network to remove the noise components. The buck converter essentially has a DC-DC step-down converter with increased current flow to the battery by adjusting the switching frequency. The buck converter (LM2596) has been utilized which has a Conversion efficiency of 92% (max) with an output voltage of 1.25–35 V, having load output current of 3A (max), having an Output ripple < 30 mV with a switching frequency of 65 kHz. For IoT-based management systems, current, voltage and temperature sensors could be utilized for real-time data acquisition. An embedded system device such as Arduino Mega 2560/ ESP8266 could be added for intelligent system management. This will enable the monitoring of the health of the whole piezoelectric energy harvester system. The piezoelectric energy harvester is the prime component in this whole design strategy to generate the voltage from vibrations. The direct piezoelectric effect has been utilized to scavenge the energy from the vibrations. The other prime objective is to identify the source of the vibrations that could be easily tuned with the designed PEH structure.

The direct piezoelectric effect occurs when a piezoelectric cantilever beam is subjected to mechanical stress. Thus, generating an electrical charge proportionate to the stress and forming an electric field. When an electric field is applied to a piezoelectric material, the piezoelectric material deforms mechanically. The reverse piezoelectric effect is the mechanism by which electrical energy is transformed into mechanical energy. The effect of piezoelectricity in the cantilever beams is described by the Eqs. (1)–(2)

$$X_3 = d_{3i} E_i \tag{1}$$

$$V_{3i} = \sigma_{XX}\, g_{3i} L_i \tag{2}$$

where X_3 is the strain, L_i is the electrode distance, V_{3i} is the open circuit voltage, and E_i is the electric field, and d_{3i} and g_{3i} are piezoelectric constants. The resonant frequency of a cantilever beam with proof mass is defined by the Eq. (3)

$$f_i = \frac{v_{n'}^2}{2\pi L^2} \sqrt{\frac{K}{m_E + \Delta m}} \tag{3}$$

where L is the length, K is the effective spring constant of the cantilever beam, $v_n = 1.875$ is the eigenvalue for the fundamental vibration, $v_{n'}^2 = {v_n}^2 \sqrt{(0.236/3)}$, Δm is the proof mass, $m_E = 0.236\ mwL$ is the effective mass of the cantilever while m is mass per unit length of the beam and w is the width of the cantilever. As the resonant frequency is inversely proportional to the squared length of the beam.

The piezoelectric energy harvester has been designed and simulated in the COMSOL. The structure comprises a beam with a cylinder having a triangular hole array. The idea is to attach the vibration source to the beam, and the cylindrical structure is made free to vibrate. The designed structure is excited with the vibration source and voltage variation has been analysed by fixing the appropriate boundary conditions. Figure 2a shows the design of PEH by applying the mesh analysis to the structure. After the application of mesh analysis, it is transformed into a combination of small elements, as illustrated in Fig. 2b. Table 1 lists the dimensions of the designed structure used for the finite element simulation. Table 2 lists the material properties that have been used for the simulation purposes. For the finite element simulation of PEH, thickness has been varied from 0.09 mm to 0.15 mm and the maximum boundary load is preferred around 10 N. In other analyses, boundary load varies from 1 to 10 Newton, keeping thickness constant. Stress and electric potential distribution due to pressure applied on the PEH structure are shown in Fig. 3a, b. Stress and potential are mainly distributed in the cantilever structure as compared to the cylindrical structure.

Modelling has been carried out to harvest the energy from the piezoelectric harvester. The modelling assists in estimating the power to be utilized for sensor devices. The modelling assists in predicting the value of load current and battery charging voltage for energy management. The Piezoelectric energy harvester-based energy management model designed in Simulink is shown in Fig. 4. To mimic the behavior of the energy harvester from a vibration source, a sine wave/chirp has been selected as a vibration source. The vibration energy has been converted to an AC

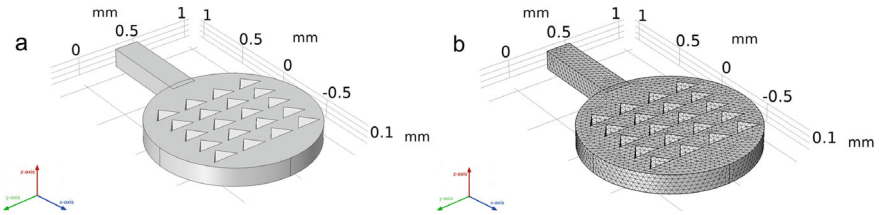


Fig. 2 **a** Geometrical design of PEH, **b** fine mesh structure of PEH

Table 1 Geometrical dimensions of designed PEH

Structure	Dimension
Handle size	0.25 × 0.75 × 0.15
Circular beam size	Radius = 0.65, thickness = 0.15
Triangular hole	0.2 × 0.2 × 0.2

Table 2 Properties of piezoelectric materials employed in simulation

Material	Density (Kg/m^3)	Poisson's ratio	Young's modulus (GPa)
Barium titanate	5700	0.28	79
PZT-4	7500	0.30	63
PZT-5H	7500	0.33	64
Lithium niobate	4647	0.25	170
Zinc oxide	5680	0.34	50
Aluminum nitride	3300	0.27	270
PZT-7A	7700	0.31	83

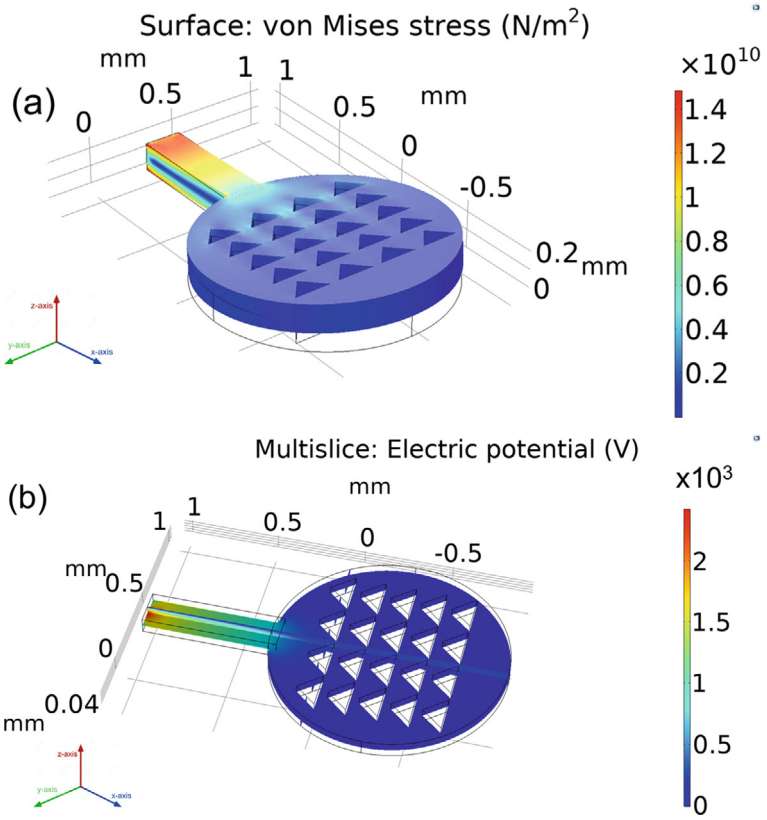


Fig. 3 **a** Stress distribution of PEH, **b** electric potential distribution in PEH

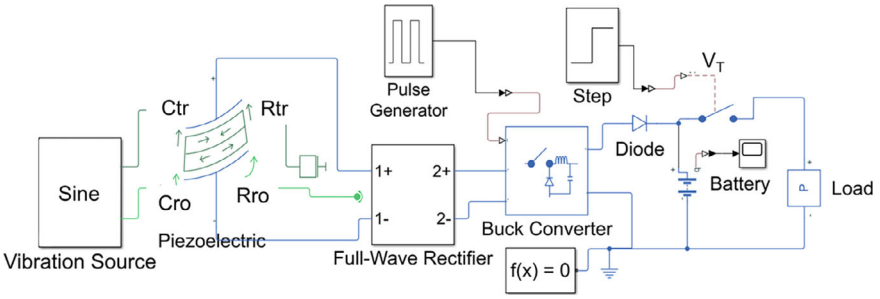


Fig. 4 PEH energy management model using Simulink

voltage, which is to be converted to DC. A Full wave rectifier has been utilized for converting AC into DC voltage with a network of filters to remove the noise, and at the output terminals, a pure DC has been generated. As the voltage generated by the designed PEH is 30–45 V, and the current capacity is low, a buck converter has been utilized to decrease the voltage to 3.5–4 V but with increased current output. The Buck converter is supplied with a pulse signal from the pulse generator. A freewheel diode is connected between the buck converter and the battery to block the reverse flow of a current. A scope has been utilized to observe the battery voltage, current, and charging state. A switch has been utilized to transfer the power to the load for the wireless sensors and microcontroller. An IoT-based monitoring system.

3 Results

The piezoelectric cantilever beam structure has been loaded with impact load. As a result, the vibration energy to electrical energy conversion occurs, and PEH starts to vibrate at its natural resonance frequency. Several vibration modes in PEH structure depend upon the vibration source frequency. As the PEH is loaded, a voltage due to vibration is generated. A plot between the voltage and boundary load is shown in Fig. 5a. The Simulated results have been obtained for different materials, including materials like PZT-4, Barium Titanate, Lithium Niobate, Aluminum nitride, PZT-5H, and PZT-7A. PZT 7A and aluminium nitride outperform the other materials as predicted by the simulation results. The presented results have been obtained for the optimum thickness of the beam, around 0.15 mm. Figure 5b shows the variation of output voltage with the thickness of PEH. The maximum voltage obtained with Aluminum Nitride is 33.57 V, and the minimum Voltage obtained for Barium Titanate is 7.66 V. Lithium Niobate, and PZT-5H show a similar behaviour against the given boundary load as shown in Fig. 5. The insertion of holes in a piezoelectric structure enhances the voltage output which could be observed from the Plots [16]. This analysis provides a framework for selecting the optimal dimensions and material

for novel energy harvesting structures that could power wireless handheld devices, sensors, weather monitoring systems, etc.

Figure 6a, b shows the variation of the battery capacity and battery charging voltage. The capacitor voltage V_c and inductor current i_L variation are shown in Fig. 7a–b. The capacitor charges to maximum voltage V_c which is used to supply power to the battery as well as the load. The value of the capacitor in the buck converter could be selected according to the design requirements and tuned. The internal resistance of the battery is kept at 2 ohms for the simulation. The capacitor charges up to a maximum of 4.2 V with inductor current varying from 0.1 to 0.15 amps. The battery current variation is shown in Fig. 8, which predicts the settling current of 0.05 amps. The total power that could be delivered to the load varies from 0.05 to 0.1 watts, with load current varying from 0.01 to 0.03 amps (Fig. 9a–c). In comparison to other investigations, the designed PEH is particularly suitable for low-frequency applications, i.e. 30–130 Hz. The designed PEH is appropriate for low-current applications such as powering remote sensors, weather monitoring systems

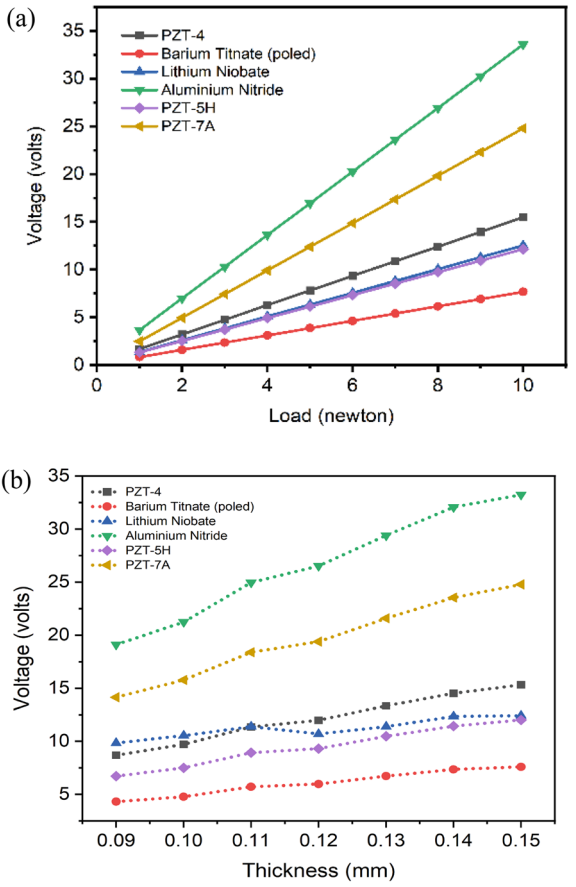


Fig. 5 **a** Output voltage versus applied boundary load for PEH. **b** Output voltage versus thickness of PEH

or powering of LED lights. The limitation of this design is that it cannot supplied large load current to high power devices. For efficiency enhancement, the switching of the buck converter must be carried with other strategies such as Slide mode control, model reference control etc. The theoretical efficiency of PZT based PEH is around 50%–90% depending upon configuration, but the efficiency is practically 15–30% [17]. The efficiency of PEH depends upon multiple parameters such as the design of PEH, source frequency, load resistance, the impedance of the circuit, proof mass, etc. So, design optimization, and different materials selection for PEH are vital in improving the efficiency. The energy management scheme with the embedded buck converter allows the optimal charging of the Li-ion battery, thus assisting in delivering power even when the energy is not generated. The design presented in this paper achieves the optimal control for charging the battery, which delivers the power to the low-power embedded devices.

4 Conclusion and Future Opportunities

A simulation investigation of a cantilever beam, along with an energy management strategy, has been presented in this paper. To explore the cantilever beam's mechanical and electrical characteristics, a piezoelectric material-based energy harvester has been designed and modelled using COMSOL Multiphysics simulation software. A comparison of the harvested voltages of different materials (as PZT-4, Barium Titanate, Lithium Niobate, Aluminium Nitride, PZT-5H, and PZT-7A) for the design has been carried out. The results show the maximum output voltage achieved with the Aluminium Nitride with present geometry at a thickness of 0.15 mm. The present design could be utilized for generating enhanced voltage due to its novel structural variations. An energy management system has been proposed and modelled to optimize the electrical energy generated. The designed system can charge the lithium-ion battery of 900 mAh capacity with an optimal voltage of 3.6 V. The novelty of this approach is that this could be integrated with multiple energy harvesting devices, including wind energy harvesting, Solar energy etc., and this system can deliver the power even in the low-frequency vibrations. This PEH-based energy system could deliver power to low-power devices such as wireless sensors, health monitoring systems, embedded devices, and non-destructive testing systems.

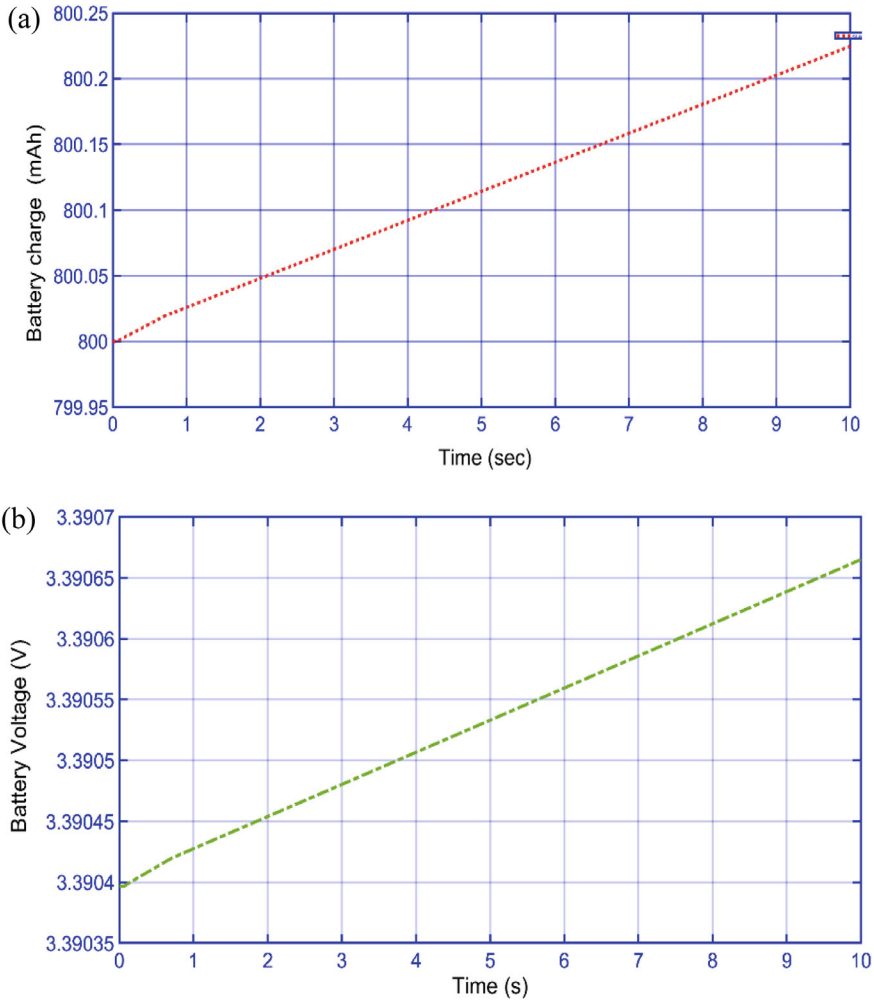


Fig. 6 **a** Battery charging capacity, **b** battery charging voltage variation

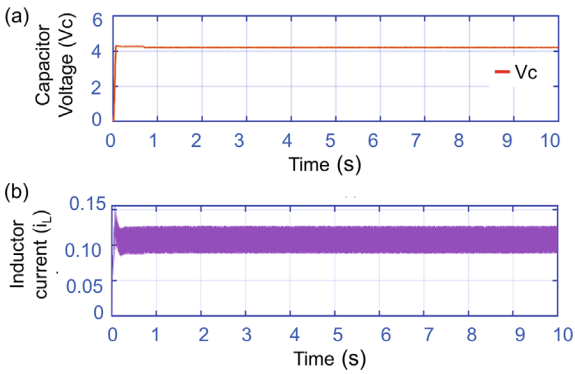


Fig. 7 **a** Variation of capacitor voltage, **b** variation of inductor current for buck converter

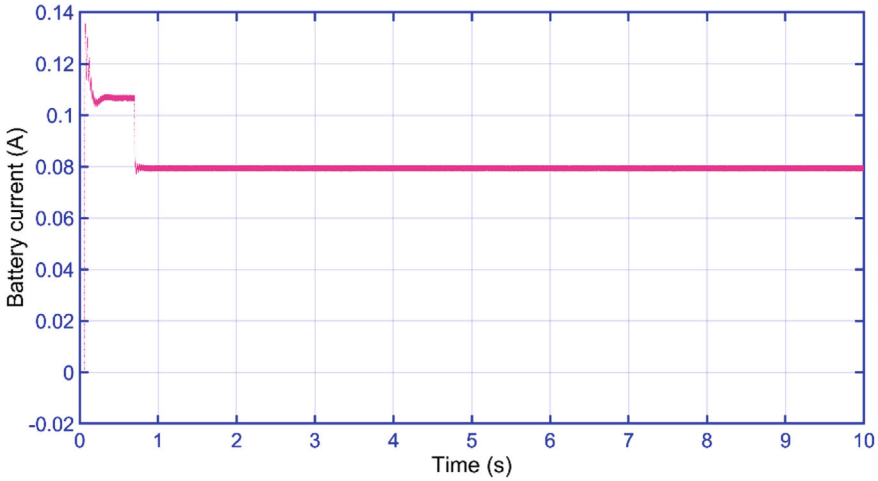


Fig. 8 Li ion battery current variation

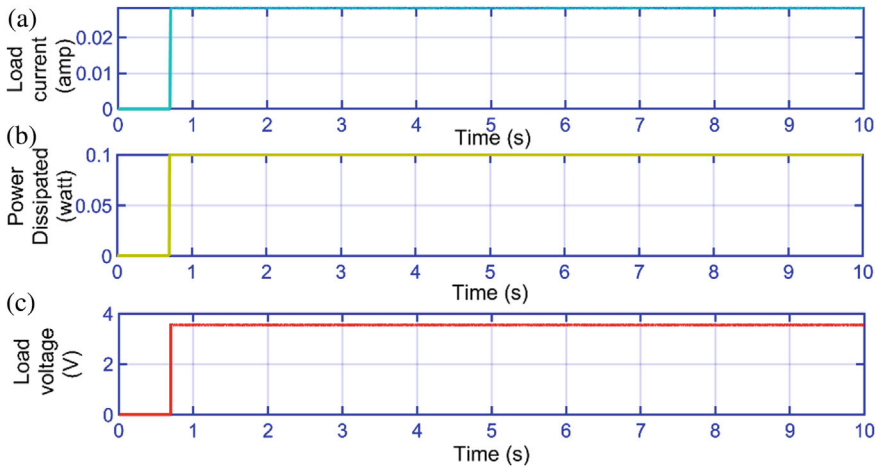


Fig. 9 **a** Load current, **b** power dissipated, **c** load voltage variation

References

1. Adhikari, J., Kumar, R., Jain, S.C.: Modeling and parametric analysis for performance improvement in piezoelectric energy harvesting tile. Ferroelectrics **573**(1), 201–213 (2021)
2. Li, X., Ming, Z., Xiao, M., Yang, K., Ma, Y., Luo, Y.: A wide-band piezoelectric harvester based on cantilever beam. Ferroelectrics **540**(1), 72–87 (2019)
3. Elahi, H., Rizwan Mughal, M., Eugeni, M., Qayyum, F., Israr, A., Ali, A., Munir, K., Praks, J., Gaudenzi, P.: Characterization and implementation of a piezoelectric energy harvester configuration: analytical. Numer. Exp. Approach. Integr. Ferroelectr. **212**(1), 39–60 (2020)
4. Raju, S.S., Umapathy, M., Uma, G.: Performance improvement of piezoelectric energy harvester using non-uniform beam with cavity. Ferroelectrics **540**(1), 193–211 (2019)
5. Mangaiyarkarasi, P., Lakshmi, P.: Effect of various shapes of single proof mass and multiple proof masses on piezoelectric energy harvester for powering mobile phone devices. Ferroelectrics **577**(1), 105–124 (2021)
6. Mohamad Hanif, N.H.H., Jazlan Mohaideen, A., Azam, H., Rohaimi, M.E.: Rotational piezoelectric energy harvester for wearable devices. Cogent Eng. **5**(1), 1430497 (2018)
7. Zhou, X., Zou, J., Gan, D., Gou, X., Chen, Y., Tian, R., Guo, T.: Design of the transducer for multi-direction wind energy harvesting based on PZT. Ferroelectrics **550**(1), 196–209 (2019)
8. He, T., Yuan, J.: A microbeam based piezoelectric energy harvester with LASMP. Ferroelectrics **570**(1), 1–14 (2021)
9. Yuan, J.H., Yang, Y.Z., Zhang, Y.H.: A piezoelectric energy harvester with light-activated shape memory polymers. Ferroelectrics **550**(1), 98–111 (2019)
10. Raju, S.S., Umapathy, M., Uma, G.: Design and analysis of high output piezoelectric energy harvester using non uniform beam. Mech. Adv. Mater. Struct. **27**(3), 218–227 (2020)
11. Anand, A., Pal, S., Kundu, S.: Multi-perforated energy-efficient piezoelectric energy harvester using improved stress distribution. IETE J. Res. **69**(6), 3723–3738 (2021)
12. Chen, F., He, M., Wang, S., Zhong, X., Guan, M.: Study of an inertial piezoelectric energy harvester from a backpack. Ferroelectrics **550**(1), 233–243 (2019)
13. Li, Z., Roscow, J., Khanbareh, H., Taylor, J., Haswell, G., Bowen, C.: A comprehensive energy flow model for piezoelectric energy harvesters: understanding the relationships between material properties and power output. Mater. Today Energy **37**, 101396 (2023)

14. Kassan, S., Gaber, J., Lorenz, P.: Autonomous energy management system achieving piezoelectric energy harvesting in wireless sensors. Mobile Netw. Appl. **25**(2), 794–805 (2020)
15. Mendez-Lira, O., Spinelli, E.M., Gonzalez-Landaeta, R.: Battery-less power management circuit powered by a wearable piezoelectric energy harvester. Energ. Technol. **9**(10), 2100520 (2021)
16. Tyagi, N., Gaur, A.M., Mahajan, A.N.: Performance evaluation of novel piezoelectric cantilever beam structure for energy harvesting. Trends Sci. **19**(22), 836–836 (2022)
17. Yang, Z., Erturk, A., Zu, J.: On the efficiency of piezoelectric energy harvesters. Extrem. Mech. Lett. **15**, 26–37 (2017)

A Framework for Development of Industrial Metaverse in Maintenance



Ravdeep Kour, Ramin Karim, Naveen Venkatesh, and Veronica Jagare

Abstract The industrial evolution has emerged the concept of Industry 5.0 paradigm. The concept of Industry 5.0 emphasizes three main aspects of industrial development, i.e. sustainability, resilience, and human centric. The emerging technologies related to digitalisation and artificial intelligence are expected to augment human-system-interaction in various industrial processes. Integrating AI and digital technologies to enable human-centric solutions in industry can manifest in a concept called "Industrial Metaverse". The Industrial Metaverse is considered a concept that emerged from integrating the Metaverse and Digital Twin, but adapted to specific characteristics of industrial contexts. Industrial Metaverse can enhance Asset Management (AM) within the Industry 5.0 framework. By integrating AI and advanced digital technologies, the Metaverse offers a promising roadmap for optimising human-system interaction in industrial contexts. Metaverse is expected to enhance human centricity of Industry 5.0 and developing metaverse in industrial contexts is challenging and requires a framework. Hence, this paper provides a framework that helps to develop Industrial Metaverse in the context of AM specifically in maintenance. This paper employs literature review in popular databases and discussions with industrial stakeholders to present the current landscape of Metaverse application in Industry 5.0 with a specific focus on maintenance. A case study from the railway sector demonstrates the practical benefits of the Metaverse in improving operational efficiency and maintenance processes. The findings contribute to understanding the Metaverse's role in shaping the future of industrial asset management and maintenance within Industry 5.0.

Keywords Industry 5.0 · Railways · Industrial metaverse · Maintenance

R. Kour (✉) · R. Karim · N. Venkatesh · V. Jagare
Division of Operation and Maintenance Engineering, Lulea University of Technology, Lulea, Sweden
e-mail: ravdeep.kour@ltu.se

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_7

1 Introduction

Use of advanced digital tools and technologies like VR, AR, and AI are redesigning industrial processes for achieving improved performance and efficiency. These tools and technologies create immersive experiences through collaborative spaces where the digital and physical worlds coexist and co-creates, forming the concept of 'Industrial Metaverse' (IM). The word 'Metaverse' was first described by Neal Stephenson in his science fiction novel 'Snow crash' which describes a virtual world that blends with reality and creates social interactions [1]. A historical view of emergence of Metaverse is presented in Fig. 1.

Researchers have provided many definitions for Metaverse in the literature like, "A crossword of "meta" (meaning transcendency) and "universe", describes a (decentralized) three-dimensional online environment that is persistent and immersive, in which users represented by avatars can participate socially and economically with each other in a creative and collaborative manner in virtual spaces decoupled from the real physical world." [2]. When we implement this metaverse in specific industrial contexts it formulates 'Industrial Metaverse'. According to Heath et al., IM is a complementary concept to Industry 5.0 [3]. They have defined IM from the perspective of human-centricity and social manufacturing paradigm as "a human-centric social manufacturing runtime that is compatible with real-time immersive virtual-real symbiosis and supports value co-creation and sustainable industrial ecology through digital incentive mechanisms among prosumers" [3]. This definition of IM is focused on more sustainable digital economy which is represented by Blockchain and Non-Fungible Tokens (NFT) to ensure the security of digital assets [3]. Till date, there is no standard definition of IM. Different researchers have provided their own definition for IM highlighting human-centricity [3, 4], robotized automation [5], sustainability [6, 7], Digital Twins [8–10], remote maintenance [10–12] and other concepts.

In the context of maintenance and Industry 5.0, *Industrial Metaverse is a shared online 3D space consisting of virtual assets, which are visualised and inspected by the avatars of the real operators and maintainers in a collaborative environment, prioritising sustainability, resilience, and human-centric design for more efficient and effective maintenance processes.* Unlike isolated VR or AR, Metaverse is a

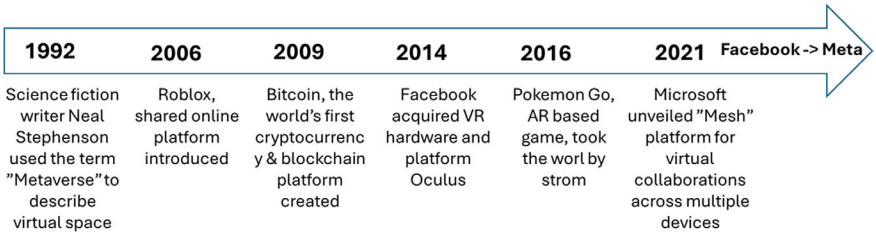


Fig. 1 History of metaverse

shared digital space where human users/operators or in future human and robots (Cobots) interact as avatars, irrespective of time and space.

The primary objective of this paper is to study and explore the development and implementation of a Metaverse within Industry 5.0 which specific focus on maintenance and provide a framework. By achieving this objective, this research aims to contribute valuable insights into the transformative potential of the Metaverse, offering an overview of Metaverse landscape for its effective development and implementation in Industry 5.0 along with a framework for its implementation in future.

The paper is structured as follows: Sect. 2, 'Research Methodology,' outlines the methods used to conduct this review process. Section 3, 'Results and Discussion,' present the overall trends in research, including publication years, literature, technologies/tools, and methods, with a specific focus on industry 5.0 within the context of maintenance along with proposed framework. Finally, Sect. 4 provides the conclusions.

2 Research Methodology

The paper uses literature review in popular databases and discussions with industrial stakeholder for data collection. The approach ensured a transparent and unbiased review by outlining specific criteria for article inclusion, information sources, search strategies, study selection, and data analysis. The focus was on recent literature published in peer-reviewed literature, available in full text and written in English. To identify relevant articles, three major databases (Google Scholar, Scopus, Web of Science) were searched using specific search terms related to the Metaverse and its industrial applications. The two search queries used in the literature title were:

Search query 1: 'Metaverse AND Industry 5.0'.

Search query 2: 'Industrial Metaverse' AND Maintenance.

A total of 13 papers were selected based on search query 1, and a total of 3 papers were selected based on search query 2 in the popular databases as described above.

3 Results and Discussions

This paper provides an overview of the current state of metaverse research in industry 5.0 with special focus on maintenance. A significant number of documents on 'Metaverse' in the title of literature (4241 on Scopus, 2195 on Web of Science, and 16,200 on Google Scholar) have been published, with a strong focus on computer science, engineering, and social sciences. To narrow down the research to 'Industry 5.0' and 'maintenance', the authors of this paper have identified some key trends. A visualisation tool developed using Microsoft PowerBI provides insights into the temporal distribution, publication types, and keywords associated with metaverse literature

within industry 5.0 (See, Fig. 2). The bar chart reveals an increase in the number of papers published over time, from 2022 to 2024. The pie chart indicates that most literature are articles, followed by conference papers and reviews. Additionally, a word cloud is included to visualise the most frequently used keywords by the authors. This word cloud highlights the predominant themes and keywords associated with the research, such as sustainability, digitalization, twin concepts, industrial applications, and human-centric approaches.

Figure 3 presents an analysis of the literature, focusing on the distribution of publications and their citations across different countries and the prevalence of specific authors. The bar charts illustrate the number of papers and the sum of citations for each country. China leads in both categories, followed by Italy and India. The word cloud highlights the most prominent authors, with names like Silvestri, Wang, and Huang appearing frequently. This analysis provides valuable insights into the geographic distribution of research contributions and the prominent authors in the field.

Additionally, Tables 1 and 2 present metaverse literature focused on Industry 5.0 and maintenance along with the technology and methodology used and its outcome.

We found three papers related to "Industrial Metaverse" linked with maintenance of granulator, lathing, turning, and milling machines. These papers have provided demonstrators for supporting remote maintenance of granulator machine with avatars and digital twins in collaborative XR environments, enabling efficient and effective operations in an industrial metaverse [10, 11]. Liu et al. [12] have proposed a framework for the AR-driven industrial metaverse for the auxiliary maintenance [12]. In these environments, remote experts and collaborators are connected through AR devices, and assist in finding out the root causes of the faults and provide solutions. Table 2 shows their details.

Based on analysis and used technologies in Industrial Metaverse we have proposed an 'Industrial Metaverse Framework' for Maintenance. The proposed framework for Industrial Metaverse that supports maintenance leverages enabling technologies to create a human-centric, resilient, and sustainable collaborative environment that bridges the physical and virtual worlds. These enabling technologies facilitating digital transitions, including VR, AR, XR, AI, Industrial Internet of Things (IIoT), big data, cloud/edge/Fog computing, blockchains, digital twins, BCI, and so on. Security is a critical concern in the framework due to the risks like data breaches. Blockchain and smart contracts offer trust and transparency, while AI helps to develop algorithms for data analytics like constructing 3D models from the collected point cloud data using devices like, Lidar scanner. Gamification and User Experience (UX) focus on user engagement and satisfaction within the metaverse environment. Web 4.0 aims to integrate various technologies for seamless human–machine interaction, and 6G is the next generation of wireless communication to support the metaverse's demands. Cloud, edge, and fog computing optimise data management and network performance for metaverse applications. 3D modeling creates virtual environments, while IoT and IIoT connect physical objects to the internet. Brain-Computer Interface (BCI) enables direct brain-computer interaction, and spatial computing integrates digital information into the physical world.

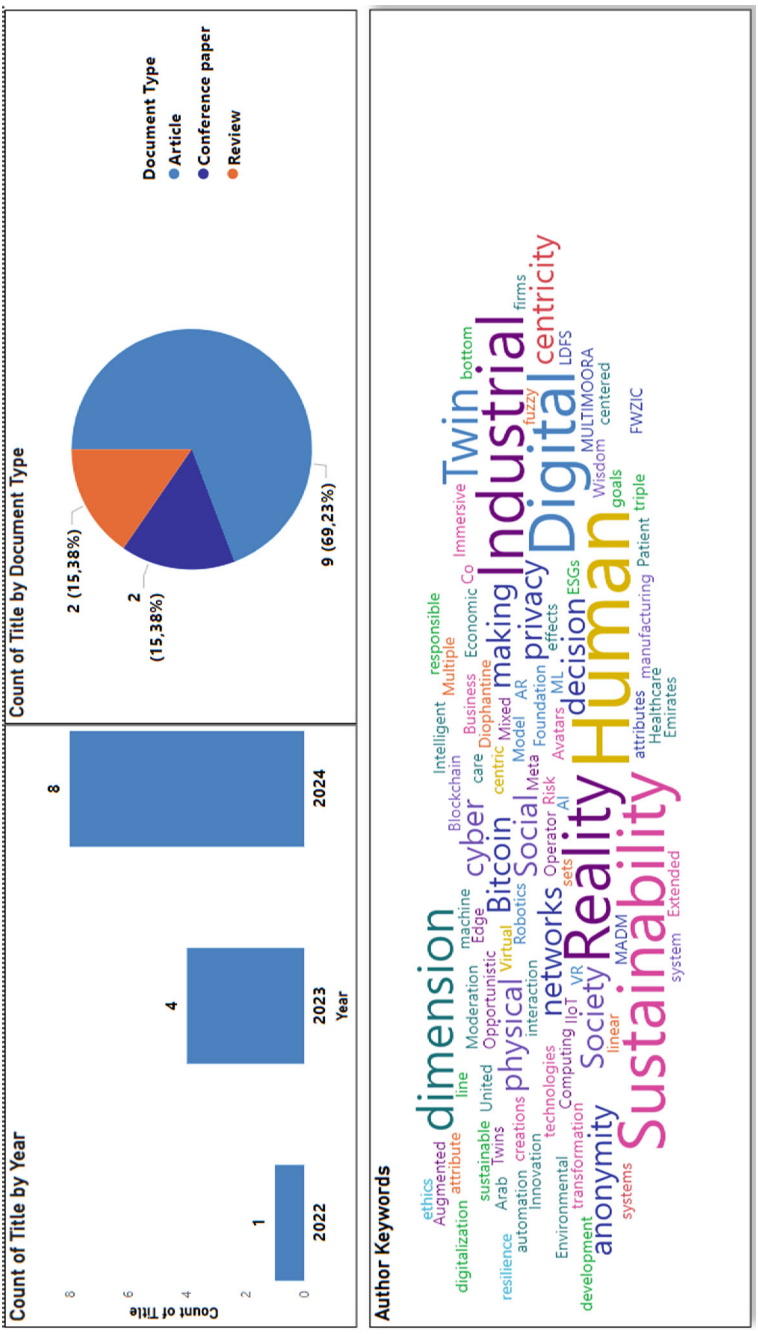


Fig. 2 Statistics of 'metaverse' literature in the context of Industry 5.0 according to occurrence, types, and author keywords

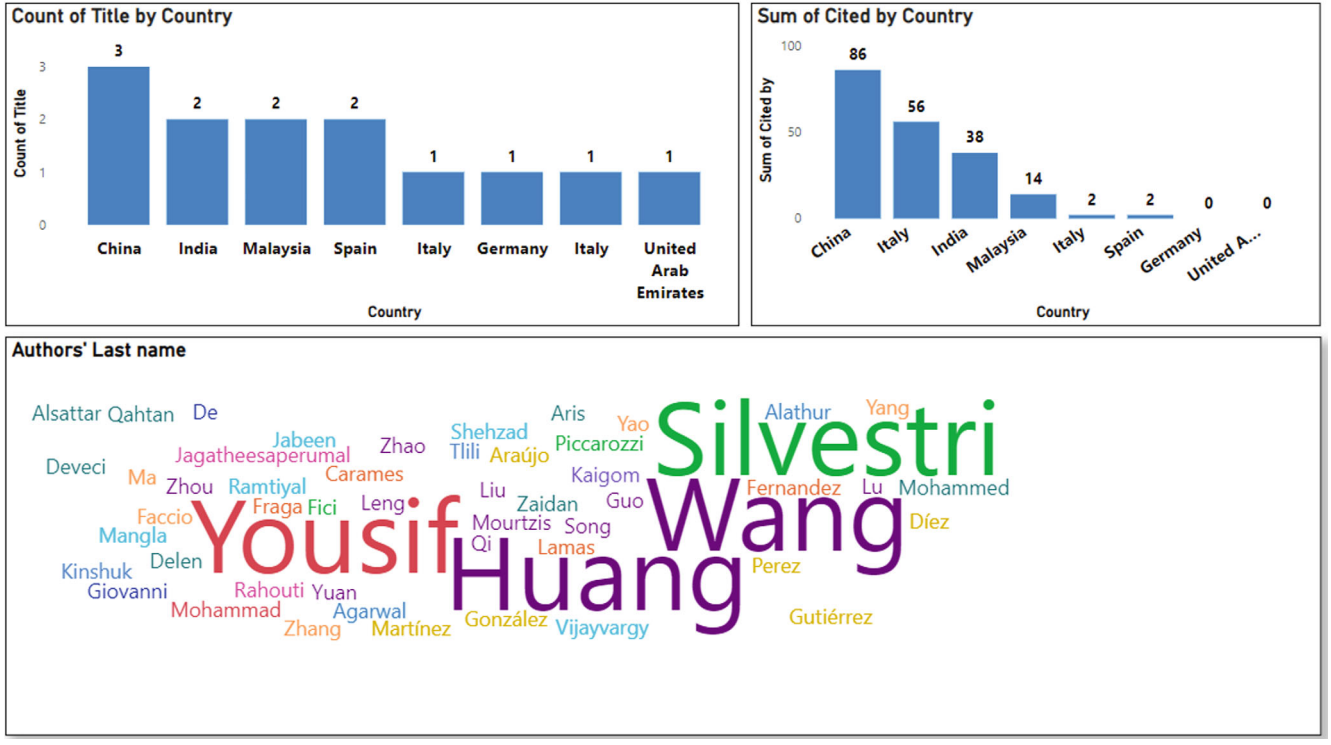


Fig. 3 Statistics of 'metaverse' literature in the context of Industry 5.0 according to geographical region, their citation, and authors' name

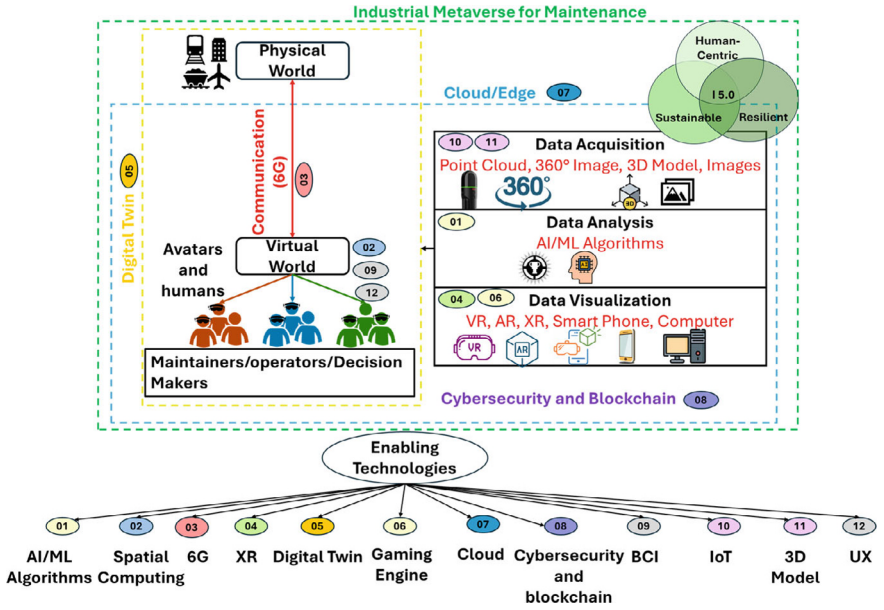


Fig. 4 Framework for industrial metaverse in maintenance adapted from [22]

The core of the framework is the physical world, representing the real-world environment where industrial operations take place. Following are the key components of the proposed IM framework for maintenance:

- *Digital Twin*: As the core of the framework, the digital twin creates virtual representations of physical assets, providing a digital counterpart for maintenance and asset management.
- *Data Acquisition*: Collecting data from the physical world using IoT sensors and devices (Like, LiDAR & 3D cameras) for feeding information into the industrial metaverse. This data can be a point cloud data of a railway system including whole infrastructure and vegetation.
- *Data Analysis*: AI/ML algorithms analyse acquired data to extract valuable insights and information. For example, extracting individual assets like, catenary wires from the collected railway point cloud data.
- *Data Visualisation*: The results of data analysis are presented through visualisation using VR/AR devices, dashboards, web apps., or other sources.
- *Decision Makers*: Humans (maintainers, operators, or any other user) and avatars (representing humans or robots) interact within the metaverse to facilitate decision-making.
- *Enabling Technologies*: A wide range of technologies, including AI/ML, spatial computing, 6G, XR, digital twins, gaming engines, cybersecurity, cloud computing, BCI, IoT, 3D modeling, and UX design, support the framework's operation.

Table 1 Literature related to metaverse in Industry 5.0

Ref	Technology/ Tools	Methodology	Outcome
[6]	NA	Literature review	It mentions that there is no definable link between metaverse and sustainability, but it has the potential to contribute to both social and environmental sustainability, but its impact is still uncertain and requires further research
[8]	XR, Opportunistic Edge Computing (OEC), Internet of Things (IoT) and Industrial IoT (IIoT) devices	Conceptual	It proposes the concept of a Meta-Operator, an industrial worker who utilises advanced XR devices to interact with industrial metaverse applications and their surroundings, following the principles of Industry 5.0
[13]	R language	Literature review	It provides a systematic literature review on metaverse implementation in industries considering its three core values, namely human-centric, sustainability, and resilience
[14]	Bitcoin network	Sensitivity analysis, correlation coefficient test and comparative analysis, MULTIMOORA method, LDFS-FWZIC Method	It provides a methodology integrating LDFS-FWZIC and MULTIMOORA methods to evaluate and rank 22 bitcoin networks based on privacy and anonymity attributes for supporting Industry 5.0's metaverse environment, with Built Bitcoin Client (BBC) identified as the most important attribute
[15]	Blockchain, bitcoin network	Review, multiple attribute decision-making (MADM), MULTIMOORA, sensitivity analysis	The paper proposes a comprehensive taxonomy of Bitcoin networks in terms of anonymity and privacy for supporting metaverse applications in Industry 5.0
[16]	DT	Conceptual	It emphasises on the significance of integrating digital twins with the metaverse in Industry 5.0 along with further recommendations and insights for optimising smart industries through immersive remote monitoring and control
[17]	Blockchain	Conceptual	It provides a socio-technically enhanced wisdom manufacturing architecture along with an advanced cyber physical system based smart manufacturing framework

(continued)

Table 1 (continued)

Ref	Technology/ Tools	Methodology	Outcome
[7]	NA	Conceptual	It provides a set of frameworks to analyse the metaverse technology from three perspectives to aid the smooth transition towards Industry 5.0. These perspectives are: • Environment, society and economy • Environmental, social and governance (ESG) • Sustainable development goals (SDG)
[4]	DT, CPS, and VR (Oculus Quest 2 glasses)	Experiment	It provides a framework based on hyperconnectivity where several enabling technologies are integrated to converge towards the industrial human-centered metaverse
[18]	Robotics	Conceptual	It identifies how the metaverse could enhance robotic applications in Industry 4.0 and Industry 5.0 by enabling more efficient workflows, collective decision-making, and human-centered robotised assistance
[19]	NA	Bibliometric analysis	It concludes that the metaverse has the potential to revolutionise digital transformation in Industry 5.0 by creating immersive experiences, collaboration, and enhancing human–machine interaction
[20]	Statisical analysis	Survey-based questionnaire technique and Structural Equation Modeling (SEM)	It highlights the metaverse's potential to revolutionise sustainable healthcare firms in Industry 5.0 by enabling more efficient and accessible healthcare services, improving patient outcomes, and reducing environmental impact
[21]	NA	Conceptual	It provides a reference model for industrial metaverse which establishes connection between three spaces i.e., physical space, cyberspace, and social space

By integrating various technologies and adopting a human-centric approach, the proposed framework can create a powerful tool for industries. The Industrial Metaverse can enhance efficiency, improve decision-making, and drive innovation by improving maintenance efficiency through predictive analytics. It can enhance remote collaboration between operators, improving communication and coordination. In addition to this, it can help in training and education to the engineers or workforce through virtual environments on complex tasks, reducing risks and improving safety.

To illustrate the framework's application, a case study has been conducted in railway sector. By creating digital twins of railway infrastructure and leveraging AR/VR technologies, maintenance teams can visualise assets in detail, identify potential

Table 2 Literature related to industrial metaverse in maintenance

Ref	Technology/Tools	Methodology	Outcome
[10, 11]	Hololens2, 5G, CAD, Unity, DT, granulator machine	Computer supported cooperative work (CSCW), MTO approach (Mensch) (Human), Technik (Technology) and Organisation (Organization), psychological stress analysis with PBA	The paper provides an XR-based demonstrator for multi-user collaboration in a remote-rendered digital twin of a granulator machine, highlighting human–computer interaction's role in facilitating human-to-human communication within potential industrial metaverse settings
[12]	IoT, Hololens2, Visual studio'19	Experiment, (Oriented brief-simultaneous localization and mapping) ORB-SLAM2	This paper provides an AR-driven industrial metaverse for the auxiliary maintenance of machine tools in IoT-enabled manufacturing workshops, enhancing efficiency and collaboration

issues, and simulate maintenance procedures. This can lead to more efficient maintenance planning, reduced costs, and improved overall reliability. The proposed framework demonstrates the potential of the Industrial Metaverse to revolutionise maintenance and asset management. By bridging the physical and virtual worlds and leveraging emerging technologies, industries can enhance efficiency, improve decision-making, and drive innovation. Figure 5 shows railway digital assets visualised and inspected by users in a metaverse environment.

The implementation of the Industrial Metaverse presents several challenges and risks, including data privacy concerns, infrastructure limitations, and the need for robust security measures. As sensitive data, such as maintenance records and real-time sensor data, is shared within the Metaverse, privacy protocols must be implemented to protect confidential information. Additionally, the infrastructure required to support the Metaverse, including high-speed networks and powerful computing resources, may pose limitations, especially in remote or less developed regions.



Fig. 5 Metaverse using VR (**a**) and AR (**b**) to inspect railway wheel and track

To mitigate security risks, robust cybersecurity measures and technologies, such as encryption, access controls, and blockchain technologies are essential to safeguard against cyberattacks and unauthorized access.

To enhance the generalisability of the framework beyond the railway industry, several adaptations can be considered. Firstly, the core components of the framework, including the digital twin, AI models, and virtual reality interfaces, can be adapted to suit the specific needs of different industries. For example, in the manufacturing industry, digital twins can be used to simulate production processes and optimise workflows. Secondly, the data integration and analysis modules can be generalised to accommodate diverse data sources, such as sensor data, IIoT devices, and historical records. Finally, the user interface can be customised to provide specific requirements of different user groups, ensuring a seamless and immersive experience. By addressing these challenges and incorporating these adaptations, the Industrial Metaverse can be a powerful tool for driving innovation and improving efficiency across various industries.

4 Conclusions

It has been concluded that few literature have been published on industrial metaverse in maintenance, highlighting the gap and its need. The analysis reveals key trends in the field, including the increasing focus on sustainability, digitalization, and human-centric approaches. The research also presents a framework for the Industrial Metaverse in maintenance, which leverages enabling technologies to create a human-centric, resilient, and sustainable collaborative environment. The framework incorporates key components such as digital twins, data acquisition and analysis, data visualization, decision makers, and enabling technologies. This framework needs to be validated in verified in future research study in real world application.

It has been concluded that by creating digital twins of railway infrastructure and utilising AR/VR technologies, maintenance teams can visualise assets in detail, identify potential issues, and simulate maintenance procedures.

Additionally, the Industrial Metaverse has the potential to significantly enhance industrial operations by improving efficiency, decision-making, and innovation. By bridging the physical and virtual worlds and leveraging emerging technologies, industries can revolutionise maintenance and asset management, leading to improved sustainability, resilience, and human-centric outcomes.

Acknowledgements This work has been carried out within the framework, 'AI Factory'. The authors would like to express their gratitude to 'Sweden's innovation agency (VINNOVA)' for their financial support for 'AI Factory'. We would also like to thank Swedish Energy Agency (Energimyndigheten) and JVTC (Luleå Railway Research Center) for their financial support. We also acknowledge the valuable support and resources provided by the eMaintenance LAB in conducting this research. In the last, we would like to thank Hanna Eriksson, Linnea Nordin, and Amanda Wennerberg Hafström from Teknikspränget (internship programme run by (IVA) Royal Swedish Academy of Engineering Sciences) to help in creating VR and AR environments for Metaverse.

References

1. Stephenson, N.: Snow Crash (1994)
2. Ritterbusch, G.D., Teichmann, M.R.: Defining the metaverse: a systematic literature review. IEEE Access **11**, 12368–12377 (2023)
3. Ball, M.: The metaverse: and how it will revolutionize everything (2022)
4. Martínez-Gutiérrez, A., et al.: Towards industry 5.0 through metaverse. Robot. Comput. Integrated Manuf. **89**, 102764 (2024)
5. Siemens and NVIDIA partner to build the Industrial Metaverse. Available: robotized automation
6. Piccarozzi, M., et al.: Metaverse: a possible sustainability enabler in the transition from Industry 4.0 to 5.0. Procedia Comput. Sci. **232**, 1839–1848 (2024)
7. De Giovanni, P.: Sustainability of the metaverse: a transition to Industry 5.0. Sustainability **15**(7), 6079 (2023)
8. Fernández-Caramés, T.M., Fraga-Lamas, P.: Forging the industrial metaverse-where Industry 5.0, augmented and mixed reality, IIoT, opportunistic edge computing and digital twins meet. arXiv Preprint arXiv:2403.11312 (2024)
9. Lyu, Z., Fridenfalk, M.: Digital twins for building industrial metaverse. J. Adv. Res. (2023)
10. Oppermann, L., Buchholz, F., Uzun, Y.: Industrial metaverse: supporting remote maintenance with avatars and digital twins in collaborative XR environments. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (2023)
11. Oppermann, L. et al.: Industrial metaverse? Human-centred design for collaborative remote maintenance and training using XR-technologies. In: International XR Conference (2023)
12. Liu, C., Tang, D., Wang, Z.: AR-driven industrial metaverse for the auxiliary maintenance of machine tools in IoT-enabled manufacturing workshop. In: 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE) (2023)
13. Tlili, A., Huang, R., Kinshuk, X.: Metaverse for climbing the ladder toward 'Industry 5.0'and 'Society 5.0'? Serv. Ind. J. **43**(3–4), 260–287 (2023)
14. Mohammed, Z.K., et al.: Bitcoin network-based anonymity and privacy model for metaverse implementation in Industry 5.0 using linear Diophantine fuzzy sets. Ann. Oper. Res. 1–41 (2023)
15. Mohammad Z.K., Bin Yousif, S., Bin Yousif, Y.: Modelling bitcoin networks in terms of anonymity and privacy in the metaverse application within Industry 5.0: comprehensive taxonomy, unsolved issues and suggested solution. J. Intell. Syst. **33**(1), 20230117 (2024)
16. Jagatheesaperumal, S.K., Rahouti, M.: Building digital twins of cyber physical systems with metaverse for Industry 5.0 and beyond. IT Professional **24**(6), 34–40 (2022)
17. Yao, X., et al.: Enhancing wisdom manufacturing as industrial metaverse for industry and society 5.0. J. Intell. Manuf. 1–21 (2022)
18. Kaigom, E.G.: Potentials of the metaverse for robotized applications in Industry 4.0 and Industry 5.0. Procedia Comput. Sci. **232**, 1829–1838 (2024)
19. Agarwal, Alathur, S.: Metaverse revolution and the digital transformation: intersectional analysis of Industry 5.0. Transform. Gov. People Process. Policy, **17**(4), 688–707 (2023)
20. Shehzad, N., et al.: Metaverse adoption as a cornerstone for sustainable healthcare firms in the industry 5.0 epoch. J. Enterp. Inf. Manag. **37**(4), 1254–1281 (2024)
21. Guo, J., et al.: Industrial metaverse towards Industry 5.0: connotation, architecture, enablers, and challenges. J. Manuf. Syst. **76**, 25–42 (2024)
22. Kour, R., et al.: Metaverse in industrial contexts—a comprehensive review. Front. Virtual Reality (2024)

Inter-Organizational and Multi-objective Operational Risk Management: A Case Study of Well Control Management for Land Drilling Operation in Iraq



Guicang Peng, Knut Erik Bang, and Tore Markeset

Abstract This paper addresses critical challenges of well control management in land rig drilling operations, focusing on the unique inter-organizational dynamics and multi-objective considerations that distinguish these operations from offshore drilling operations. We achieved this goal by examining data from an Iraqi oilfield and conducting a thorough review of the literature on well control management. Our work highlights major operational challenges that arise when different entities with diverse goals and safety cultures come together, such as unclear responsibilities and differing levels of safety program maturity. We examined these issues from a process safety perspective and offer both general and targeted solutions to improve well control management. The general recommendation is to implement effective process safety management and choose performance indicators that emphasize leading, not lagging, metrics. This means using measures that ensure all parts of an organization are working toward the same safety goals. The study offers specific recommendations about managing well control barriers—specifically, the physical systems, technical solutions, operational processes, and organizational structures that should be in place to keep wells under control. At the heart of these recommendations is a push for better cross organizational collaboration and a sharper focus on balancing the many diverse objectives that tend to pull risk managers in different directions. The insights gained from this work go beyond just onshore drilling operations. They offer guidance for many industries facing complex, dynamic, interconnected operations.

Keywords Operational risk · Onshore drilling operation · Well control management · Inter-organization · Multi-objective · Barrier management

G. Peng · K. E. Bang · T. Markeset (✉)
University of Stavanger, Stavanger, Norway
e-mail: tore.markeset@uis.no

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_8

1 Introduction

Effective well control is essential for protecting petroleum drilling operations. Although offshore drilling frequently dominates safety management research, land drilling operations pose unique problems that necessitate specialized safety management strategies [1].

The land rig drilling service business possesses lower entry barriers compared to its offshore counterpart and functions within a highly competitive, contract-driven landscape. This complicates the establishment of consistent safety management positions due to the brevity of contracts and the intricacy of subcontracting arrangements.

The challenge is compounded by the more localized nature of the inland region, where activities face various geophysical, political, and socio-environmental restrictions. These discrepancies hinder compliance with international standards and generally lead to a weakening of rigorous operating protocols observed in offshore environments, hence affecting safety and operational integrity.

Additionally, land drilling activities frequently need the relocation of rigs, which involves disassembling and transporting them in segments to a new site. Each rig relocation possibly introduces new hazards while exacerbating old ones, hence complicating the management of well control apparatus and the overall integrity of the rig compared to offshore drilling rigs.

This paper analyzes these problems using a field investigation done at an oilfield in Iraq in 2019 [2]. This research offers advice about two essential aspects: inter-organizational dynamics and multi-objective considerations for land drilling operations. This approach seeks to improve operational safety by transcending conventional personal safety measurements to include process safety management, with a special emphasis on barrier management tactics at physical, technical, operational, and organizational levels.

1.1 Methodology, Limitations and Structure

This research conducts a literature review and a comprehensive field study at an Iraqi oilfield, employing questionnaires, workshops, interviews, audits, and surveys across diverse stakeholder organizations to evaluate existing safety management practices and identify deficiencies in current well control practices.

The gathered data and information were examined utilizing a barrier management framework that assessed four principal barrier components: physical, technical, operational, and managerial and organizational barriers. The analysis concentrated on two principal categories of challenges identified in the field survey. The first aspect is focused on inter-organizational complexity, which included misaligned safety expectations and practices, fragmented responsibilities, communication obstacles, and disparate levels of safety maturity among various stakeholders. The second group analyzed multi-objective factors, encompassing the equilibrium between operational

efficiency and safety, cost management in relation to safety investment, regulatory compliance vs operational flexibility, and short-term performance contrasted with long-term safety culture development.

This observation from the field study serves as the foundation for formulating pragmatic recommendations that consider both theoretical frameworks and operational limitations of land drilling operations. Nonetheless, despite its comprehensiveness, the study has certain inherent flaws that warrant acknowledgment. The field study was performed in a singular geographic site in Iraq, perhaps constraining the generalizability of the findings to other areas. The study's temporal scope may fail to encompass long-term changes in land drilling activities. The conclusions may also be affected by certain regional regulatory and cultural circumstances.

This paper is organized as follows: Sect. 2 examines the results of the literature review. Section 3 delineates the methods and conclusions of our field investigation, presenting evidence for the stated issues, addresses the prevailing challenges in well control management pertinent to land drilling operations, synthesizing ideas from both the literature review and field study. Section 4 presents evidence-based proposals to tackle the highlighted difficulties, taking into account both theoretical frameworks and practical limitations, Sect. 5 ends the analysis and proposes avenues for future investigation.

2 Literature Review

The literature review focused on three primary domains: major accident risk management in drilling operations, process safety management methodologies, and barrier management strategies. The review encompassed scholarly articles, industry standards, and accident investigation reports from the preceding three decades (1990–2020), with a specific focus on literature released subsequent to the Deepwater Horizon incident in 2010. Essential databases, including Web of Science, Scopus, and OnePetro, were queried utilizing combinations of terms such as "well control," "drilling safety," "process safety," "barrier management," and "major accident risk." The inquiry was then narrowed to concentrate primarily on literature pertaining to inter-organizational dynamics and multi-objective factors in operational risk management.

Major accidents in the oil and gas sector, such as well blowouts, fires, explosions, extensive environmental pollution, and events leading to numerous injuries or fatalities, have emerged as a central concern in operational risk management. Since the 1990s, there has been an increasing acknowledgment that the management of large accident risks necessitates a separate methodology from personal safety management [3, 4]. This transition is predicated on the recognition that major accidents fundamentally differ from personal safety occurrences in terms of their causation, complexity, and possible ramifications.

Following the Deepwater Horizon incident, research has broadened to include organizational, technological, and human variables in drilling operational risk

management, transcending single-objective methodologies. Significant contributions encompass Vinnem [5], who examines the formulation of risk indicators for substantial hazards in offshore operations; recent investigations by Shafiee and Animah [6] have furthered this research by creating a comprehensive framework for risk assessment in intricate drilling operations, with particular emphasis on the difficulties associated with high-pressure high-temperature wells. Islamov et al. [7] have proposed a thorough methodology for assessing well control occurrences through a mixed machine learning approach.

Kongsvik et al. [8] advocate for a qualitative methodology regarding organizational safety indicators, recognizing the shortcomings of solely quantitative measures and arguing for the integration of qualitative data to enhance the understanding of organizational safety; Skogdalen and Vinnem [9] and Skogdalen's [4] research on safety performance indicators in deepwater drilling. These studies offer significant insights into the inter-organizational characteristics of drilling operational risks. Nevertheless, these studies are insufficiently in tackling the difficulties of deploying such systems across many organizations with possibly divergent objectives and priorities [10–12].

Many oilfield operators and drilling contractors continue to depend solely on human safety metrics, such as Lost Time Injury (LTI), to evaluate operational safety concerns. The 2010 Deepwater Horizon incident serves as another illustration. This disastrous incident, resulting in 11 deaths and a multi-billion-dollar environmental catastrophe, transpired on the same day the drilling rig was lauded for seven years of injury-free operation, highlighting the insufficiency of conventional safety metrics in addressing intricate operational risks [13].

The management of major accident risks, grounded in a process safety framework and employing a barrier management technique, has been a central theme in extensive research [14]. Specifically, in petroleum well drilling operations, there are four essential barrier factors for well control: physical, technical, operational, and managerial/organizational, as seen in Fig. 1. Each barrier aspect is essential for risk mitigation and operational safety, as elaborated below.

Physical barriers are tangible assets that organizations directly oversee to avert incidents or alleviate their effects. The literature highlights the significance of physical barriers, including blowout preventers (BOPs) and fire suppression systems, in managing major accident hazards [5].

Technical barriers are generally addressed by the engineering and technical departments, encompassing the design, installation, and maintenance of safety–critical systems. The literature highlights the significance of engineering solutions in averting significant mishaps, including the creation of sophisticated well control systems and automated safety monitoring technology [15].

Operational barriers encompass the procedures and practices integral to the daily activities overseen by the operational divisions. Literature indicates that conventional safety performance metrics, such as Lost Time Injury (LTI), frequently fail to adequately encompass the intricate dangers linked to drilling activities. This is especially pertinent in operational safety, where compliance with procedures and employee proficiency are essential.

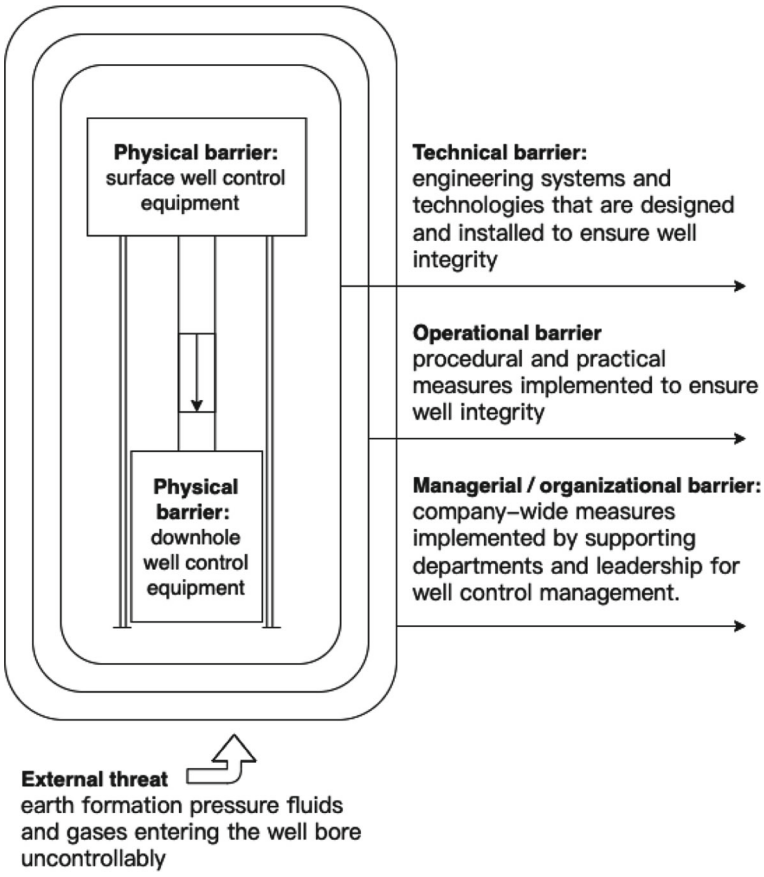


Fig. 1 Drilling well blowout barrier scheme

Managerial and organizational barriers include the comprehensive responsibilities of all supporting departments, such as executive leadership, administration, human resources, contracts, and procurement. Literature indicates that organizational characteristics significantly impact safety management, including aspects such as safety culture and the distribution of resources for safety initiatives.

3 Field Study and Findings

The field study was conducted on an oilfield in Iraq in 2019, where we employed the Participatory Action Research (PAR) methodology to ensure comprehensive coverage of the research objectives. PAR is a qualitative research methodology that involves researchers and participants collaborating to understand social issues and

take actions to bring about social change. This approach allowed us to integrate real-world experiences with theoretical insights, making our study both relevant and actionable.

The study focused on three primary areas: (1) drilling business environment challenges; (2) intra- and inter-organizational collaboration processes; and (3) technical and operational standardization. The research methods included questionnaires, workshops, interviews, audits, and surveys. Market surveys, document reviews, on-site audits, and stakeholder discussions were conducted across multiple organizations involved in drilling operations [2]. The field study identified challenges that can be put into two categories, namely: inter-organizational complexity and multi-objective considerations.

3.1 Inter-Organizational Complexities

In the field study, we observed that land drilling operations typically involve numerous organizations, including operators, contractors, and subcontractors, each with their own safety practices, cultures, and priorities. This complexity creates several challenges:

Misaligned Safety Expectations and Practices

The often temporal and transactional contracting relationships between organizations led to notable discrepancies in safety expectations and practices. This misalignment resulted in increased risks of miscommunication, ultimately compromising overall safety performance.

Fragmented Responsibility

The distribution of safety responsibilities across multiple organizations can result in gaps or overlaps in safety management. In several cases, we observed that no single entity assumed responsibility for emergency response coordination. This fragmentation left potential risks unaddressed, posing a significant threat to the operation's safety integrity.

Communication Barriers

The study also highlighted substantial communication challenges arising from differences in organizational structures, reporting systems, and safety terminology. For example, during a well control incident, delayed and unclear communication was observed between the operator and subcontractor, due to incompatible reporting systems exacerbating the situation.

Varying Levels of Safety Maturity

Organizations involved in the operation may have different levels of safety management maturity, leading to inconsistencies in safety performance across the project. For instance, we observed that contractors and subcontractors often have a less mature

safety program, which contrasts sharply with the more thorough and systematic approach taken by the operator.

3.2 Multi-objective Considerations

The field study also revealed the complex interplay of multiple, often competing objectives in safety management. The findings highlighted several key challenges faced by organizations:

Operational efficiency vVersus Safety

In the observed operations, there was a persistent tension between the need to maintain drilling efficiency and the imperative of adhering to comprehensive safety practices. For instance, during a high-pressure drilling phase, the drive to meet project deadlines led to the bypassing of certain non-mandatory safety checks. This decision, while boosting short-term productivity, significantly increased the risk of safety incidents.

Cost Management Versus Safety Investment

The study also identified financial pressures as a critical factor influencing safety investment decisions. In many cases, contractors opted to delay well control equipment upgrading, due to budgetary constraints. This decision underscored the ongoing tension between minimizing operational costs and the need for adequate investment in safety measures and training.

Regulatory Compliance Versus Operational Flexibility

Strict adherence to safety regulations may sometimes be perceived as limiting operational flexibility, creating a challenge in balancing compliance with practical operational needs. For example, the need to comply with exhaustive reporting requirements delayed critical on-the-fly decision-making during a complex drilling operation, illustrating the operational trade-offs inherent in regulatory compliance.

Short-Term Performance Versus Long-Term Safety Culture Development

The transient nature of many drilling projects often led to a focus on immediate safety performance metrics rather than the development of a long-term safety culture. The emphasis is often placed on meeting short-term safety targets to secure future contracts, with little attention paid to fostering a sustainable safety culture.

3.3 Observations Related to Barriers

Physical Barriers

The field study in Iraq further underscored the variability in the effectiveness of physical barriers across different organizations. It was observed that contractors often operated with outdated or poorly maintained equipment, leading to significant gaps in the physical safety measures. These findings highlight the need for stringent management of physical barriers, ensuring that all involved parties maintain and upgrade their assets to the highest standards.

Technical Barriers

The field study revealed inconsistencies in the maturity of technical barriers across different organizations. While operators generally maintained robust technical barriers, some contractors and subcontractors lacked the technical expertise or resources to implement equivalent measures. This disparity in technical capability often led to uneven safety performance across the project, highlighting the need for standardized technical requirements and improved collaboration between technical departments across organizations.

Operational Barriers

The field study revealed that operational barriers were frequently compromised by the pressures of maintaining efficiency and meeting tight deadlines. For example, during high-pressure drilling phases, certain safety procedures were bypassed to save time, significantly increasing the risk of incidents. This tension between operational efficiency and safety adherence underscores the necessity of reinforcing operational barriers through better training, stricter adherence to protocols, and continuous monitoring.

Managerial and Organizational Barriers

The field study highlighted significant challenges in this area, particularly in the alignment of safety cultures and the distribution of safety responsibilities. Fragmented responsibility across multiple organizations often led to gaps in safety management, particularly in emergency response coordination. Additionally, misaligned safety cultures between operators and contractors resulted in inconsistent safety practices, further complicating the safety management landscape.

4 Discussions and Recommendations

The insights derived from the literature review and field study underscore the critical need for a more nuanced and comprehensive approach to operational risk management in land drilling operations. The complexities inherent in these operations—ranging from the involvement of multiple organizations with differing safety cultures

to the dynamic and transient nature of drilling environments—highlight the challenges of current land drilling operational risk management practices. These findings necessitate a shifting of focus from traditional personal safety metrics to a more integrated approach that encompasses process safety, inter-organizational and multi-objective collaboration, as discussed below.

4.1 Process Safety Management Based on Barrier Management Strategy

The academic consensus strongly advocates that major accident risks in industries such as oil and gas should be managed through a process safety management system rather than relying solely on a personnel safety management approach [3, 9]. While personal safety is undeniably a crucial priority in any well-managed organization—and organizations that excel in personal safety often perform well in other operational areas—this focus alone is inadequate for preventing major accidents. These accidents, though rare, result from a complex series of failures across multiple layers of protection, as depicted in the Swiss Cheese Model proposed by Reason [3].

To effectively manage these risks, it is essential to move beyond a singular focus on personal safety and adopt a barrier management strategy that is inherently suited to process safety management. A barrier management strategy involves the identification, implementation, and maintenance of multiple layers of safety barriers—physical, operational, technical, and organizational—that work collectively to prevent incidents or mitigate their consequences. This approach ensures that all potential failure points are addressed, thereby reducing the likelihood of catastrophic failures.

For land drilling operations, this means shifting towards a more comprehensive safety management system that integrates both process and personnel safety. By recognizing and actively managing all layers of protection, from physical barriers to organizational processes, this strategy enhances the overall resilience of operations and significantly reduces the risk of major accidents. This integrated approach aligns with the complexities and dynamic nature of land drilling, ensuring that both immediate and long-term safety risks are systematically addressed.

4.2 Leading Versus Lagging Barrier Performance Indicators

The distinction between leading and lagging barrier performance indicators varies across industries and publications. The UK Health and Safety Executive [16] defines leading indicators as active monitoring systems for operational and organizational controls designed to prevent unwanted situations. In contrast, lagging indicators are reactive measures that reflect the outcomes of the risk control systems. This definition

draws a clear boundary between leading indicators, which measure the effectiveness of controls, and lagging indicators, which summarize past outcomes.

However, other perspectives, such as the one presented by the American Petroleum Institute [17], classify performance indicators along a spectrum from the most lagging (Tier 1) to the most leading (Tier 4). This gradation acknowledges that the distinction between leading and lagging indicators is not absolute; instead, it reflects a continuum.

In the context of well control barrier management, the definition of an "unwanted situation" is critical. The "blowout"—the uncontrolled flow of formation fluids to the surface without effective well control barriers— is the primary unwanted situation. Following the UK HSE's approach, indicators that measure the performance of well control barrier measures and provide sufficient time and space to prevent a blowout are considered leading indicators. Conversely, any measurement summarizing well control failures after an actual blowout would be classified as a lagging indicator. This distinction is vital for designing effective well control barrier management systems in land drilling operations.

4.3 Non-probabilistic Barrier Performance Indicators

As evidenced in the field study, land drilling operations are inherently dynamic, characterized by short-term durations, frequent changes in facilities, equipment, workforce, geolocations, and contract statuses. These ever-changing conditions make it impractical to rely on probabilistic safety performance indicators, which typically depend on large datasets collected within stable operational contexts over extended periods. Such probabilistic indicators often fail to capture the immediate risks present in a rapidly shifting environment.

For instance, consider a drilling site where the equipment is relocated every few weeks. The workforce might also be rotating, with varying levels of experience and familiarity with the specific site's operational challenges. In this scenario, relying on historical OSHA statistics from previous years or different locations would provide little insight into the current safety risks. The conditions on the ground may have changed significantly, rendering past data less relevant [18, 19].

Instead, short-term or just-in-time performance indicators are far more effective in such contexts. For example, real-time monitoring of equipment integrity through automated sensors can immediately alert the team to any developing issues, such as unexpected pressure changes in the well or vibrations indicating potential equipment failure. Similarly, daily safety briefings that assess the current competencies of the rotating workforce, combined with up-to-date hazard assessments based on the latest geophysical data, can provide a clearer and more actionable risk picture [20].

Another scenario could involve a sudden geopolitical shift or regulatory change affecting the drilling site's operational environment. In such cases, a just-in-time barrier performance indicator might include real-time updates on local security

threats or changes in environmental regulations that could impact drilling activities. By integrating this immediate data, the drilling operation can swiftly adapt its barrier management protocols to reflect the current risks, ensuring that decisions are based on the most relevant and up-to-date information.

This approach not only aligns with the dynamic nature of land drilling operations but also enhances the ability to respond proactively to emerging risks, thereby supporting more effective and timely safety management decision-making.

4.4 Inter-Organizational and Multi-objective Barrier Performance Indicators

The dynamic nature of land drilling operations, where most activities are outsourced to contractors, presents significant challenges that require a more comprehensive approach to safety management. Traditional intra-organizational barrier performance indicators, which are often focused on internal processes, are insufficient in this context. The complexities of land drilling operations, including the involvement of multiple organizations with varying safety practices, the transient nature of the workforce, and the constantly changing operational environment, demand the development of inter-organizational and multi-objective barrier performance indicators.

These indicators should be able to capture the intricate interplay between diverse organizational elements and the multiple objectives that influence safety performance. For example, in land drilling operations, the integrity of physical assets, workforce competencies, and organizational structures are in a state of constant flux, influenced by factors such as fluctuating contracts, geopolitical conditions, local regulations, and the evolving needs of the project.

To effectively manage safety in such a complex environment, it is crucial to develop indicators that not only measure individual organizational performance but also assess the collective impact of all involved entities. These inter-organizational and multi-objective barrier performance indicators should be developed to measure how well safety protocols are upheld across contractors, while meeting operational goals, ensuring adequate resource allocation for safety, tracking compliance with evolving regulations, and promoting sustainable safety practices.

By integrating these multi-objective considerations into inter-organizational barrier performance indicators, land drilling operations can achieve a more holistic view of safety management; the result is a more resilient well control management system that is better equipped to handle the unique challenges of land drilling operations.

4.5 Well Control Barrier Performance Management

Based on the above recommendations, considerations of well control barrier performance management are discussed below.

Physical Barrier Performance

Well control equipment certifications, third-party inspection certifications: It is critical that both contractors and operators collaborate to ensure all well control equipment meets industry standards and regulatory requirements, through rigorous certification processes. This collaboration should involve clear communication channels and shared responsibilities for arranging third-party inspections to verify compliance. By aligning the certification and inspection efforts across different organizations, the risk of equipment failure is minimized, and a consistent safety standard is maintained throughout the operation.

Well control spare parts and consumables' quantity and quality, storage, and maintenance performance: Effective inter-organizational coordination is essential for managing the inventory of well control spare parts and consumables. Contractors, operators, and procurement departments must work together to ensure the availability of high-quality materials, proper storage conditions, and timely maintenance. Joint audits and inspections should be conducted regularly to assess the condition and availability of these resources. This collaborative approach prevents potential shortages or failures and ensures all parties are prepared to address emergencies, maintaining operational continuity and safety.

Surface well control equipment integrity and reliability: The integrity and reliability of surface well control equipment should be a shared responsibility between contractors and operators. Regular monitoring, testing, and maintenance must be coordinated to ensure that equipment like blowout preventers (BOPs) and choke manifolds are functioning correctly. Joint maintenance schedules and emergency response plans should be established, with clear roles and responsibilities for each party. This inter-organizational approach ensures that equipment reliability is upheld across the operation, reducing the likelihood of uncontrolled releases and improving overall well pressure management.

Casing head and annular integrity and reliability: Ensuring the integrity and reliability of casing heads and annular preventers requires close cooperation between all parties involved in the drilling operation. Operators and contractors should jointly conduct inspections and tests, with each organization contributing their expertise to identify and address any potential issues. Collaborative decision-making processes should be in place to ensure that any signs of wear or potential failure are promptly addressed. This shared responsibility helps maintain wellbore stability and prevents blowouts, fostering a unified approach to safety management.

Drilling mud circulation and weighting system integrity and reliability: The effectiveness of the drilling mud circulation and weighting systems depends on the coordinated efforts of operators, contractors, and service providers. These parties must work together to perform regular inspections and maintenance of mud pumps,

tanks, and related equipment. By establishing joint protocols for monitoring system performance and addressing any detected issues, the inter-organizational team can ensure that the drilling mud systems remain reliable. This collaboration is crucial for controlling well pressure, preventing kicks, and ultimately avoiding blowouts, thereby enhancing the overall safety of the drilling operation.

Technical Barrier Performance

Casing program, drilling mud and cement program design: Developing and implementing well control-related specifications for casing programs, drilling mud, and cementing require close coordination between the operator's technical team and contractors. Collaboration ensures that the specifications meet the operational needs and safety standards of all parties involved. Joint planning sessions and reviews should be conducted to align the technical requirements and adjustments that may arise from changing downhole conditions. This collaborative approach helps to ensure that all aspects of the well design contribute effectively to well control, minimizing the risk of blowouts or other major incidents.

Well shut-in, well kill, and diverting procedures and specifications: Developing and implementing well shut-in, well kill, and diverting procedures require a coordinated approach between all technical teams involved in the operation. These procedures should be tailored to the specific downhole conditions, with input from both operators and contractors to ensure that they are effective and feasible. Regular drills and simulations should be conducted collaboratively, to ensure that all teams understand their roles and responsibilities during an emergency, enhancing the overall readiness of the operation.

Drilling program technical specifications: For example, flow check requirements during penetration and tripping in and out, downhole pressure balancing requirements by mud weighting and property adjustment. Specifications related to well control during drilling operations, such as flow checks and pressure balancing, should be developed in close collaboration between the operator and drilling contractor. This ensures that the procedures are practical, effective, and aligned with the operational capabilities of the involved parties. Regular reviews and adjustments should be made collaboratively, based on real-time data and operational feedback, to maintain the integrity of well control measures throughout the drilling process.

Blowout drill schedule and requirements: The scheduling and execution of blowout drills must involve all parties in the drilling operation, including operators, contractors, and any relevant third parties. These drills should be conducted regularly and reflect realistic scenarios that could occur in the specific drilling environment. Jointly planned and executed drills ensure that all teams are well-prepared, understand their roles, and can coordinate effectively in the event of a blowout, thereby enhancing the overall emergency preparedness of the operation.

Operational Barrier Performance

Well control related risk analyses, communication, and documentation: Effective communication channels must be established to ensure that risk analysis findings are documented comprehensively and shared across organizations in a timely manner.

This collaborative documentation process ensures that all parties have a clear understanding of the potential risks and the measures in place to mitigate them, facilitating a unified approach to managing well control risks.

Well control related training and certification: Operators and contractors must collaborate to ensure that all personnel, regardless of their organizational affiliation, receive consistent, high-quality training that meets industry standards. Periodic recertification and temporary training sessions should be planned jointly to address any gaps in knowledge or skills, ensuring that the workforce remains competent and prepared to handle well control challenges. This inter-organizational approach to training fosters a cohesive safety culture and enhances the overall effectiveness of well control measures.

Well control execution and recording: The execution and recording of well control precautions, such as flow checks and mud property adjustments, well shut-in, well kill, blowout diverting, and ignition should be standardized across all participating organizations. This requires close coordination between operators and contractors, to ensure that procedures are followed consistently and that records are accurately maintained. By establishing common protocols and monitoring compliance jointly, organizations can reduce the risk of operational discrepancies that could compromise well control, thereby ensuring a uniform approach to safety throughout the drilling process.

Blowout emergency responses and collaboration: In the event of a blowout or other well control emergency, a coordinated response is essential. Operators, contractors, and any other involved parties must have a clear, shared understanding of the emergency response procedures and their respective roles. Regular joint emergency drills should be conducted to practice collaboration under different criticality, ensuring that all parties can work together effectively during an actual emergency. This inter-organizational collaboration is crucial for minimizing the impact of a blowout and ensuring the safety of personnel and the environment.

Managerial and Organizational Barrier Performance

Well control requirements at the tendering and contracting stages: Operators and contractors must work together to ensure that the contractual terms reflect comprehensive well control standards that are mutually agreed upon and clearly understood by all parties. This collaboration should extend to defining specific responsibilities, compliance expectations, and performance metrics related to well control, thereby creating a solid foundation for safety management throughout the project.

Well control management commitment and resource allocation: Ensuring commitment to well control management requires alignment and coordination across all organizational levels, both within and outside the operator's and contractor's teams. This involves not only allocating sufficient resources—such as personnel, technology, and financial investment—but also ensuring that these resources are consistently supported by leadership at all levels. Inter-organizational meetings and resource planning sessions should be conducted to synchronize efforts and guarantee that well control initiatives are adequately funded and prioritized across the entire operation.

Well control work interface and report line definition: Clear definitions of work interfaces and report lines are essential for effective well control management. This requires a collaborative approach where operators, contractors, and other involved parties jointly define and agree on the reporting structures and communication channels. By establishing clear interfaces and report lines, organizations can ensure that information flows efficiently, responsibilities are clearly delineated, and any issues related to well control are promptly addressed by the appropriate parties.

Well control performance auditing, feedback, and continuous improvement: Auditing well control performance and implementing continuous improvement processes should involve all relevant organizations. A shared auditing framework should be established, with input from both operators and contractors, to assess well control performance objectively and comprehensively. Regular feedback sessions should be held to review audit findings and develop joint improvement plans. This collaborative approach to performance auditing ensures that lessons learned are shared across organizations and that continuous improvement efforts are aligned, leading to enhanced safety outcomes.

Well control contingency and change management: Effective contingency planning and change management for well control require close coordination between all parties. Joint development of contingency plans, including clear procedures for emergency decision-making, ensures that all organizations can respond swiftly and cohesively to unforeseen challenges. Regular joint drills and scenario planning exercises should be conducted to test these plans and refine the coordination mechanisms. This inter-organizational approach to contingency and change management enhances the overall resilience of the operation, ensuring that well control issues can be managed effectively even under the most challenging conditions.

5 Conclusion

The research findings reveal contrasts between traditional safety management approaches and the unique challenges present in land drilling operations. A fundamental distinction lies in the overall safety management approach. While traditional methods have predominantly focused on personal safety metrics such as Lost Time Injury (LTI) and single-organization management systems, our findings demonstrate that land drilling operations require a more comprehensive, integrated process safety management approach that effectively spans multiple organizations with varying safety cultures and operational priorities.

The nature and application of performance indicators represent another key area of contrast. Traditional safety management typically relies on probabilistic indicators based on long-term data collection in relatively stable operational environments. However, our research reveals that land drilling operations demand a different approach, requiring non-probabilistic, just-in-time indicators that can effectively capture and respond to the dynamic, short-term nature of these operations.

This distinction is particularly important given the frequent changes in personnel, equipment, and operational conditions characteristic of land drilling environments.

Organizational structure and dynamics present a third significant area of contrast. While offshore operations typically operate within more stable, hierarchical organizational structures with clearly defined lines of responsibility, land drilling operations function within complex inter-organizational networks characterized by frequently changing contractors and overlapping responsibilities. This complexity necessitates a more adaptable and collaborative approach to safety management than traditional methods provide.

These comparative findings clearly demonstrate that traditional approaches to well control management are insufficient for addressing the unique challenges of land drilling operations. The dynamic nature of land drilling, combined with its inherent inter-organizational complexity and multi-objective considerations, requires a more nuanced and integrated approach to safety management than conventional methods provide.

Key contributions of this research include:

Highlighting the need for a shift from personal safety metrics to a more comprehensive process safety management approach in land drilling operations.

Proposing a barrier management strategy that addresses physical, technical, operational, and managerial/organizational aspects of well control.

Advocating for the development of non-probabilistic, just-in-time performance indicators that are better suited to the dynamic nature of land drilling environments.

Emphasizing the importance of inter-organizational and multi-objective barrier performance management to capture the complex interactions between various stakeholders and operational goals.

Providing specific recommendations for improving well control barrier performance across all four barrier elements, with a focus on collaborative approaches between operators, contractors, and other involved parties.

The insights gained from this study extend beyond well control management in land drilling operations. In fact, the importance of inter-organizational and multi-objective considerations highlighted in this research is increasingly relevant across various sectors of industry today. As business environments become more complex and interconnected, organizations are increasingly relying on collaborative networks and partnerships to achieve their objectives. This trend, coupled with the increasingly dynamic and unpredictable nature of global markets and operational environments, requires a paradigm shift in how we approach operational risk management as a whole. Whether in manufacturing, logistics, healthcare, or any other sector involving multiple stakeholders and rapidly changing conditions, the principles of inter-organizational collaboration and multi-objective optimization in risk management are becoming critical. The framework and strategies proposed in this study for well control management can provide valuable perspectives for developing more robust, adaptive, and collaborative operational risk management approaches across various industries, helping organizations navigate the complexities of modern operational landscapes.

This study, while providing valuable insights, has several limitations that should be acknowledged, and which point to important directions for future research. The geographical scope was limited to a single oilfield in Iraq, the temporal nature of the 2019 field study provides only a snapshot of operations, and access to certain contractors was limited, potentially affecting the comprehensiveness of our inter-organizational analysis. Future research should address these limitations through longitudinal studies tracking the implementation of the proposed framework over time, comparative studies across different geographical regions and regulatory environments, and the development of standardized methodologies for real-time assessment of barrier effectiveness.

References

1. Theophilus, S.C., Esenowo, V.N., Arewa, A.O., Ifelebuegu, A.O., Nnadi, E.O., Mbanaso, F.U.: Human factors analysis and classification system for the oil and gas industry (HFACS-OGI). Reliab. Eng. Syst. Saf. **167**, 168–176 (2017). https://doi.org/10.1016/j.ress.2017.05.036.Publisher:Elsevier
2. Peng, G., Bang, K.E., Markeset, T.: Identifying challenges for major accident prevention in onshore drilling operations—a case study from a middle east onshore oilfield, November 1–5 (2020). https://doi.org/10.3850/978-981-14-8593-0
3. Reason, J.: Managing the Risks of Organizational Accidents. Ashgate Publishing, Hampshire, United Kingdom (1997)
4. Skogdalen, J.E.: Risk management in the oil and gas industry: integration of human, organizational and technical factors. PhD thesis, UiS, NO.142, University of Stavanger, Norway, Norway (2011). http://hdl.handle.net/11250/182335 Accessed 04 July 2024
5. Vinnem, J.E.: Risk indicators for major hazards on offshore installations. Saf. Sci. **48**(6), 770–787 (2010). https://doi.org/10.1016/j.ssci.2010.02.015.Publisher:Elsevier
6. Shafiee, M., Animah, I.: An integrated FMEA and MCDA based risk management approach to support life extension of subsea facilities in high-pressure–high-temperature (HPHT) conditions. J. Mar. Eng. Technol. **21**(4), 189–204 (2022)
7. Islamov, S., Grigoriev, A., Beloglazov, I., Savchenkov, S., Gudmestad, O.T.: Research risk factors in monitoring well drilling—a case study using machine learning methods. Symmetry **13**(7), 1293 (2021)
8. Kongsvik, T., Almklov, P., Fenstad, J.: Organizational safety indicators: some conceptual considerations and a supplementary qualitative approach. Saf. Sci. **48**(10), 1402–1411 (2010). https://doi.org/10.1016/j.ssci.2010.05.016.Publisher:Elsevier
9. Skogdalen, J.E., Vinnem, J.E.: Quantitative risk analysis of oil and gas drilling, using Deepwater Horizon as case study. Reliab. Eng. Syst. Saf. **100**, 58–66 (2012). https://doi.org/10.1016/j.ress.2011.12.002.Publisher:ElsevierLtd
10. Øien, K., Utne, I.B., Herrera, I.A.: Building safety indicators: part 1—theoretical foundation. Saf. Sci. **49**(2), 148–161 (2011). https://doi.org/10.1016/j.ssci.2010.05.012.Publisher:ElsevierIndiaPvtLtd
11. Øien, K., Utne, I.B., Tinmannsvik, R.K., Massaiu, S.: Building safety indicators: part 2—application, practices and results. Saf. Sci. **49**(2), 162–171 (2011). https://doi.org/10.1016/j.ssci.2010.05.015.Publisher:ElsevierIndiaPvtLtd
12. Wilpert, B.: Organizational factors in nuclear safety. In: PSAM 5: probabilistic safety assessment and management, Osaka, Japan (2000)
13. Deepwater Horizon Incident Joint Investigation Team: Marine Casualty, Explosion, Fire, Pollution and Sinking of Mobile Offshore Drilling unit Deep-water Horizon, with the Loss of Life

in the Gulf of Mexico. Technical report, Bureau of Ocean Energy Management (2010). https://www.bsee.gov/site-page/deepwater-horizon-joint-investigation-team-releases-final-report
14. Aven, T., Vinnem, J.E.: Risk Management: With Applications from the Offshore Petroleum Industry. Springer, London, UK (2007)
15. Kongsvik, T., Almklov, P., Haavik, T., Haugen, S., Vinnem, J.E., Schiefloe, P.M.: Decisions and decision support for major accident prevention in the process industries. J. Loss Prev. Process Ind. **35**, 85–94 (2015). https://doi.org/10.1016/j.jlp.2015.03.018.Publisher:Elsevier
16. HSE and Chemical Industries Association: Developing Process Safety Indicators. A Step-by step Guide for Chemical and Major Hazard Industries. Health and Safety Executive, United Kingdom (2006). https://www.hse.gov.uk/pubns/books/hsg254.htm
17. American Petroleum Institute: Process safety performance indicators for the refining and petrochemical industries. American Petroleum Institute (2010). https://www.api.org/oil-and-natural-gas/health-and-safety/refinery-and-plant-safety/process-safety/process-safety-standards/rp-754.
18. Lashari, S.E.Z., Takbiri-Borujeni, A., Fathi, E., Sun, T., Rahmani, R., Khazaeli, M.: Drilling performance monitoring and optimization: a data-driven approach. J. Pet. Explor. Prod. Technol. **9**(4), 2747–2756 (2019). https://doi.org/10.1007/s13202-019-0657-2
19. Svensson, I., Saeverhagen, E., Bouillouta, F.M.: Driving Rig Performance through Real-Time Data Analysis, Benchmarking, Dashboards and Developed Key Performance Indicators, pp. 021–011002 (2015). SPE-173413-MS. Accessed 18 Aug 2024
20. Yunus, M.N.M., Tarmizi, M.F.A.: Drilling minimum standard as regulatory benchmarking for drilling performance in Malaysia. In: Offshore Technology Conference Asia (2020). Kuala Lumpur, Malaysia. https://doi.org/10.4043/30384-MS

An Improved Framework to Transition from TBM to CBM in the Maintenance of Freight Locomotives



Mohamed Ali Mohamed Alaswed and Jyoti K. Sinha

Abstract This paper presents an improved CBM (Condition-based-Maintenance) framework for freight locomotives maintenance that can be implemented to either complement or completely replace any conventional TBM (Time-Based Maintenance) approach adopted by the rail freight industry today. The foundation and most significant component of this framework is the critical review of; the existing TBM regime, and the historical failure data using techniques such as Pareto Analysis and FMEA (Failure Mode & Effect Analysis). Once the critical review is completed, the improved framework experimentation and implementation stages can begin, which consist of the deployment of sensors and CBM techniques to detect anomalies on the locomotive systems and subsystems by fusion or separate monitoring of parameters such as vibration and temperature. An example of the framework application in the critical review stage is included in this paper using real locomotive failure data. The paper concludes with suggested future work to enable transition from TBM to CBM with focus on components with the greatest impact on cost and the organization's profitability at the time of this study.

Keywords Condition-based-maintenance · Time-based maintenance · Freight locomotives maintenance · CBM · TBM · Reliability · Maintenance

1 Introduction

Freight locomotives that are used to pull intermodal and bulk material cargo today still mostly follow a periodic Preventative Maintenance (PM) strategy which is mainly based on inspections, component cleaning, component replacements and overhauls

M. A. M. Alaswed (✉) · J. K. Sinha
School of Engineering, The University of Manchester, Manchester, UK
e-mail: mohamedalimohamed.alaswed@postgrad.manchester.ac.uk

J. K. Sinha
e-mail: jyoti.sinha@manchester.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_9

at fixed intervals also known as TBM or Time-Based-Maintenance [1]. Condition-Based Maintenance (CBM) as a strategy in the freight locomotives industry has existed for a long time and multiple attempts of its application already exist in both industry and academic literature. For example, prior study dating as far back as 1995 by K N Fry from British Rail Research investigated using CBM to improve the reliability of Class 47 locomotive reliability [2].

The main objectives of this work are to determine the best answer to the question of why has CBM not been adopted as a main strategy in freight locomotives maintenance until today? In other words, why is TBM still dominant in the maintenance of freight locomotives? And based on the assumption that CBM is the better approach, what is the best practical approach that freight locomotive businesses can take to expand the adoption of CBM maintenance in their already existing TBM regimes? The best way to answer these questions is to firstly consider the advantages and disadvantages of TBM in contrast to CBM and then consider similar studies that tried to answer the same questions. The literature review to identify similar studies was done through online searches of journals such as ScienceDirect, Google Scholar, Extrica, MDPI, Hindawi and Sage Journals for the years 1990–2022.

The advantages of TBM include being a tried and tested strategy that has yielded the reliability required by operators and customers of freight trains for many decades and the ease of maintenance planning and execution based on a fixed set of maintenance activities [3]. The disadvantages of TBM include sharply increased cost of maintenance as the number of maintenance activities increases, and locomotives must be made unavailable to perform maintenance that may not be truly required which increases downtime [4]. The biggest disadvantage of CBM on the other hand is introducing more complicated maintenance planning as execution is based mostly on predictions [5]. However, in a CBM regime only when the health of the locomotive begins to deteriorate, the locomotive is removed from service and maintenance is performed, and when done over the service life of the locomotive results in a sharply decreased maintenance cost [6]. Similar studies that discussed the issue of slow CBM adoption in industry includes one carried out by Rosmaini and Shahrul [6] who concluded that CBM is the favorable approach, but more research is needed to make CBM more realistic for industrial practice. Another study by Verhagen et al. [5] discussed the challenges to slow CBM adoption in the Aviation industry, and they concluded that even with big strides made in condition monitoring technology reseach, adoption of CBM is behind due to lack of practical frameworks that explain how to adapt current maintenance procedures for CBM. The work in this paper is intended as contribution to fill the gap in industrial application of CBM research in general and freight locomotives maintenance research in particular with the main argument being that CBM adoption in freight locomotives maintenance can be achieved with a clear and improved framework that demonstrates how to introduce CBM to an already existing and mostly TBM regime and that the lack of such framework is the main reason for the slow adoption of CBM in the locomotive maintenance industry.

The paper is presented as follows: basic introduction, improved maintenance framework methodology overview and descriptions of its three stages namely critical

review, experimentation and implementation stages respectively. The framework is demonstrated in Sect. 3; however, this demonstration only includes the critical review stage of the improved framework, which is the biggest limitations of this study. Finally, conclusions are drawn in Sect. 4.

2 Improved Framework Methodology

As it is the case with all TBM regimes, the biggest drawback lies in neglecting failure causes. In this paper, an improved framework to transition to CBM for freight locomotives is proposed, which can address the existing TBM model's limitations. The TBM strategy considered for improvement in this paper belongs to a diesel-electric locomotive which is equipped with a Diesel Engine that turns a high-power Traction Alternator, Fig. 1 shows this arrangement.

In most industries, the original version of the maintenance policy is derived from the OEM (Original Equipment Manufacturer) suggestions [6]. The original version of the locomotive maintenance regime in this paper is based on the manufacturer guidelines, which are mostly TBM. Later versions of the regime evolved with adjustments made by the locomotive operators and maintainers to address specific operating and performance requirements. The power generated by the locomotive Traction Alternator is fed to a group of electric traction motors, Fig. 2a shows this arrangement. Each traction motor is mechanically coupled to gears that rotate the locomotive wheels. Figure 2b shows the motor and wheelset arrangement. Each set of three motors and three wheelsets form what is known as "motor bogie" or truck, the locomotive considered in this paper has two bogies or trucks [7].

Figure 3 shows the proposed CBM Framework hierarchy with the critical review stage as its foundation. The improved framework is comprised of three main stages which are described in the following sections.

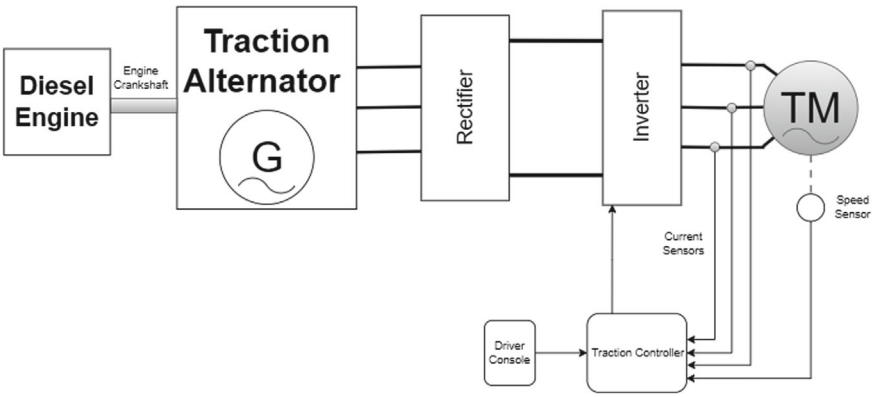


Fig. 1 Locomotive engine and traction alternator arrangement

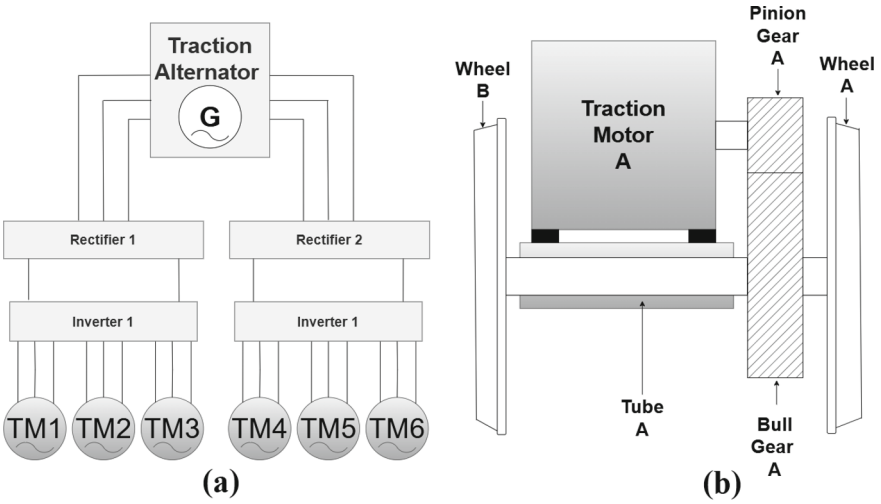


Fig. 2 **a** Alternator and motors arrangement, Fig. 2, **b** motor and wheelset arrangement

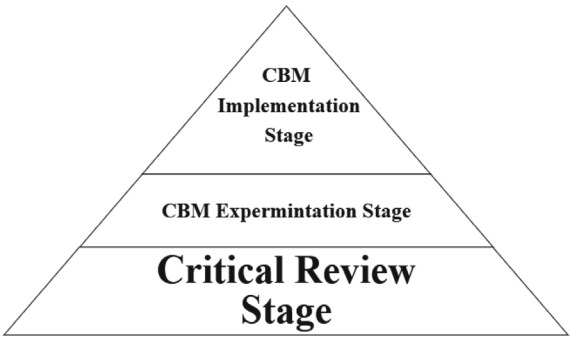


Fig. 3 Improved framework hierarchy

2.1 Critical Review Stage

In this stage the locomotive historical failure data is analyzed using the Pareto analysis technique which is also known as the 80–20 rule [8]. What follows is the critical review and analysis of the existing TBM regime in relation to the historical failure data represented in the Pareto analysis; doing this will enable understanding of how it has become the maintenance strategy adopted by the organization or workshop, the strengths and weaknesses of the current regime, the obstacles that could hinder the implementation of CBM.

The next step is the critical review using Failure Modes and Effects Analysis (FMEA) tool, which is the primary analytical tool in the proposed framework, because it is a systematic approach and can reveal all the failure modes, causes of failure, and the effects of failures better than all other techniques [3].

Other techniques are also considered in this framework, namely, Weibull Analysis because it is one of the best tools to reveal the characteristic life of failed components and to help in establishing a more precise failure rate baseline that can be used to perform other analytical techniques such as Fault Tree Analysis (FTA) and Reliability Block Diagram (RBD). FTA & RBD are effective in modelling the causes and effects of failures on the overall system or component reliability, which is something FMEA and Weibull cannot provide [3]. The final step of the critical review stage is to identify gaps in the existing maintenance regime where CBM can be implemented.

2.2 CBM Experimentation Stage

This is about experimenting with practical condition monitoring methods that can be performed by maintenance teams within locomotive workshop environments or onboard locomotives by fusion or separate monitoring of parameters such as vibrations, temperatures, pressures, voltage, current acoustic emissions, and other parameters. This includes performing in-service testing and review of results and findings.

2.3 CBM Implementation Stage

Selecting the most appropriate and effective condition monitoring methods based on results in experimentation stage. Integration of the most effective and easy to implement -into the existing maintenance framework. Cost benefit analysis to analyze the risks and the benefits of the improved maintenance model.

3 Framework Application

3.1 Historical Failure Data—Pareto Analysis

Pareto analysis of locomotive component failures data between September 2022 and December 2023 for the locomotive in this case study is performed. Figure 4 shows the top 20 locomotive failures during this period.

The Pareto analysis shown in Fig. 5 is based on the frequency of each locomotive component failure and shows that the first nine locomotive component failures contribute to approximately 80% of all failures with AC traction motor as the most dominant component failure (42%).

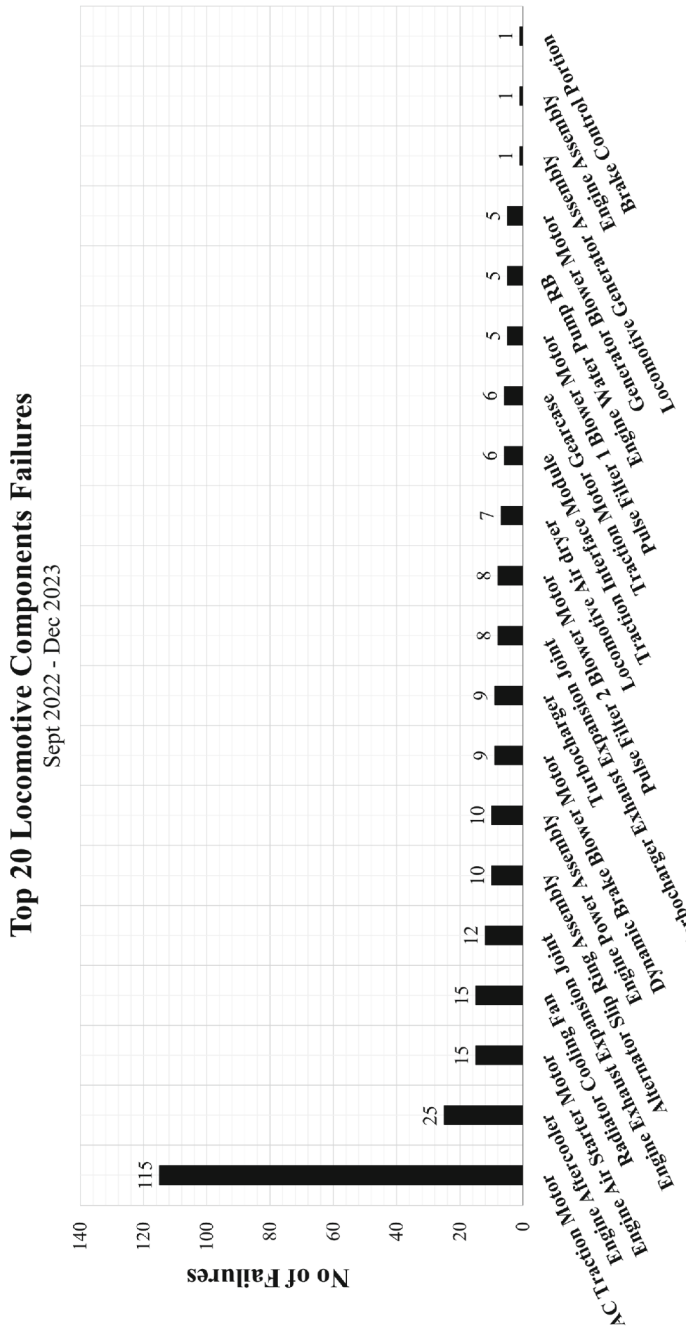


Fig. 4 Locomotive historical components failures—top 20

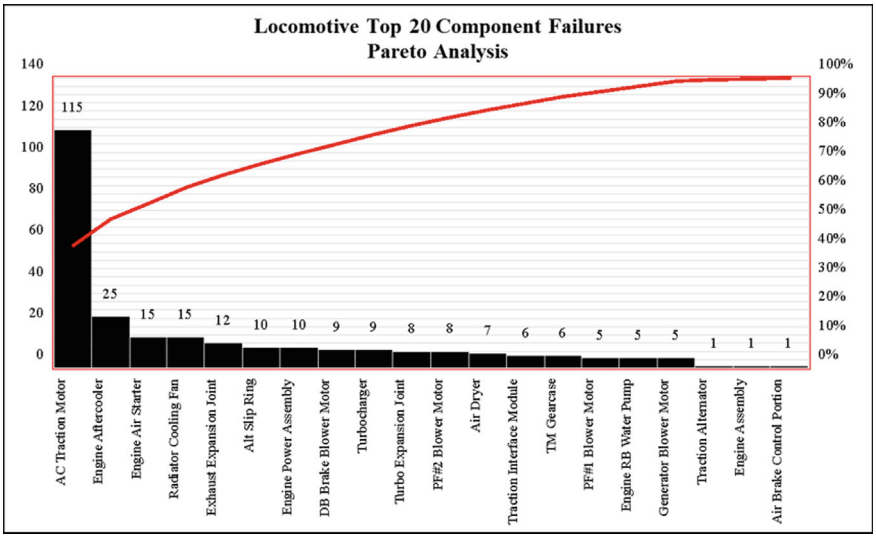


Fig. 5 Failures Pareto analysis

3.2 Critical Review of Current Maintenance Regime

The current maintenance regime for the locomotive in this paper starts with a daily exam, followed by a 30-day inspection, and continues at a 30-day interval that consists of Inspections (I), Cleaning (C), and parts Replacements (R) such as filters and seals. At 1095-days or 3 years of service life, the locomotive is stopped for its first major maintenance event, which includes major locomotive parts and systems inspections (I), Cleaning (C), Replacements (R) and overhauls (O). The electric AC traction motors that feature as the dominant failure in the Pareto analysis for example is one of the components that gets overhauled during the 1095-days or 3 years interval and get replaced during the 2190-days or 6 years interval.

At 3650-days or 10 years, the locomotive bogies and all their sub-assemblies are overhauled and finally at 4380-days of service-life, the main traction alternator is overhauled. Table 1 shows all 20 maintenance components represented in Fig. 4 with details of the type of TBM task that is carried out during each interval.

3.3 Failure Mode and Effects Analysis (FMEA)

The Failure Mode and Effect Analysis (FMEA) is used to critically evaluate the most dominant failure in the Pareto Analysis, which is the AC motor failures. A list of potential failure modes pertaining to each category are identified using the terminology adopted by the freight industry to identify traction motor failures. Rankings

Table 1 Current maintenance regime review

Locomotive component or system	Maintenance interval								
	Daily	30-day	92-day	184-day	365-day	1095-day	2190-day	3650-day	4380-day
AC traction motor	I	I	I	I	I	R	O	I	R
Engine aftercooler	I	I	I	I	I	I	R	I	R
Engine air starter motor	I	I	I	I	I	R	R	I	R
Radiator cooling fan	I	I	I	I	I	R	R	I	R
Engine exhaust expansion joint	I	I	I	I	I	I	R	I	R
Alternator slip ring assembly	I	I	I	I	I	I	I	I	I
Engine power assembly	I	I	I	I	I	I	R	I	R
Dynamic brake blower motor	I	I	I	I	I	I	R	I	R
Turbocharger	I	I	I	I	I	I	R	I	R
Turbo Exhaust expansion joint	I	I	I	I	I	I	R	I	R
Pulse filter 2 blower motor	I	I	I	I	I	I	R	I	R
Air dryer	I	I	I	I	I	I	I	I	I
Traction interface module	I	I	I	I	I	I	I	I	I
Traction motor gearcase	I	I	I	I	I	I	I	I	I
Pulse filter 1 blower motor	I	I	I	I	I	I	R	I	R
Engine RB water pump	I	I	I	I	I	I	R	I	R
Generator blower motor	I	I	I	I	I	I	R	I	R
Traction alternator	I	I	I	I	I	I	I	I	O
Engine assembly	I	I	I	I	I	I	I	I	R
Air brake control portion	I	I	I	I	I	I	R	I	R

Table 2 Failure modes and effects analysis

Component	Failure cat	Failure mode		Failure effect		Failure cause		S	O	D	RPN
Traction motor	Mechanical	FI	Locked motor rotor	El	Locked wheels	C1	DE bearing lubrication	8	10	5	400
						C2	DE bearing excessive loading (External factors)				
						C3	DE bearing misalignment				
						C4	NDE bearing lubrication				
						C5	NDE bearing misalignment				
						C6	NDE bearing excessive loading (External factors)				
		F2	Slipped pinion	E2	Non powered wheels	C3	DE bearing misalignment	7	3	5	105
						C9	Manufacturing defect				
		73	Noise or vibrations	E3	Wheel set excessive vibrations	C2	DE bearing excessive loading (External factors)	8	6	10	4S0
						C3	DE bearing misalignment				
						C5	NDE bearing misalignment				
				El	Locked wheels	C6	NDE bearing excessive loading (External factors)				

(continued)

Table 2 (continued)

Component	Failure cat	Failure mode		Failure effect		Failure cause		S	O	D	RPN
						C7	Loose motor mounting				
	Electrical	74	Grounded stator	E2	Non powered wheels	C10	Stator failure	7	3	3	63
		75	Grounded rotor	E2	Non powered wheels	C11	Rotor failure	7	2	5	70

are set for the severity (S), the occurrence (O) and the detection (D) to allocate Risk Priority Number (RPN) values based on Eq. 1 [9].

$$\mathrm{RPN} = \mathrm{S} \times \mathrm{O} \times \mathrm{D} \tag{1}$$

3.4 Identifying Gaps for the CBM

AC traction motors are rotating components, understanding the vibration signature of rotary systems is a good choice for CBM [3]. Analysis of the current maintenance model revealed corrective actions already in place that are performed to detect C1 & C4 (lubrication failures) at intervals before the motor is removed for overhaul. C2 & C6 (excessive loading), C3 & C5 & C6 (misalignment) do not have any actions in the current maintenance model to detect them; besides they can all be considered CCFs (Common Cause Failures) with high vibration as the CCF (Common Cause of Failure). Therefore, in this paper, it is proposed that regular onboard vibration measurements should be conducted and vibration trend analysis performed.

The optimal interval for performing onboard vibration monitoring and trend analysis depends on several factors, such as the size of the fleet, the availability of locomotives to perform monitoring, the type of measuring equipment to be used and where to install it on the locomotive, but to start with, the proposed interval for performing onboard vibration measurements and trend analysis in the improved maintenance model is 90 days. This interval is to be reviewed after performing at least 20 vibration measurements. The equipment to be used is to focus on the measurement of radial vibrations using accelerometers in two test points as shown in Fig. 6.

4 Conclusion

In this paper, an improved framework to transition from a mostly time-based maintenance model to a mostly condition-based maintenance model for frieght locomotives is proposed. The framework consists of three stages namely; the Critical Review Stage, the CBM Experimentation Stage, and the CBM Implementation Stage. Only the critical review stage is demonstrated in this paper focusing on the frequency of failures and the most dominant failure, which is the AC Traction Motor. The marriage between frequently performing critical reviews of the existing TBM model and historical failures is the secret recipe of the improved maintenance framework presented in this paper. As future work, the proposed framework can be extended to demonstrate the framework CBM experimentation and implementation stages and to include more locomotive components. Moreover, the Pareto analysis can be expanded to be based on cost to address the components with the greatest impact on the profitability of the organisation.

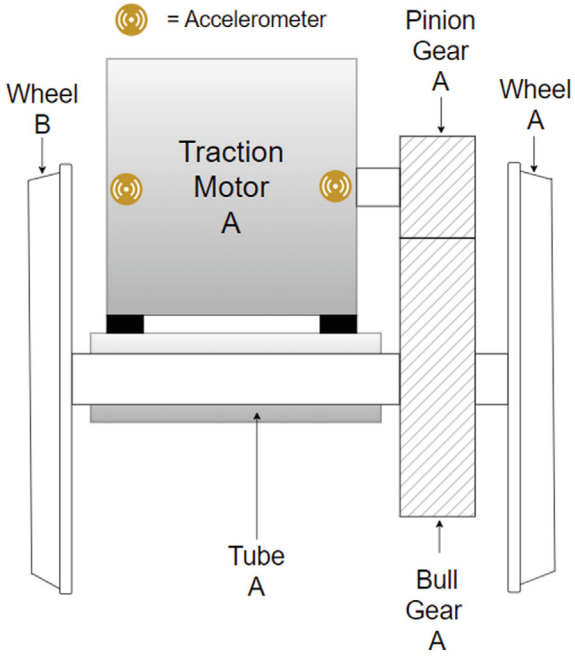


Fig. 6 Accelerometers proposed locations to be mounted on traction motors to perform onboard vibration monitoring and trend analysis

References

1. Barnes, A.: Increased locomotive performance using condition based maintenance. Birmingham (2014)
2. Fry, I.N.: Diesel locomotive reliability improvement by system monitoring. J Rail Rapid Transit **F** (1995)
3. Ben-Daya, M., Duffuaa, S. O., Raouf, A., Jezdimir, K., Ait-Kadi, D.: Handbook of Maintenance Management and Engineering, 1st edn. Springer Dordrecht Heidelberg, London New York (2009)
4. Gao, W., Zhang, Z., Zhou, Y., Jiang, L., Wang, W.: Maintenance optimization for a system with an implemented interval of replacement activities. Chengdu, China (2013)
5. Verhagen, W.J.C., Santos, B.F., Freeman, F., Kessel, P.V., Zarouchas, D., Loutas, T., Yeun, R.C.K., Heiets, I.: Condition-based maintenance in aviation: challenges. Areospace **10**(762) (2023)
6. Rosmaini, A., Shahrul, K.: An overview of time-based and condition-based maintenance. Comput. Ind. Eng. **1**(63), 135–149 (2012)
7. The railway technical website, "Bogies," (2019). [Online]. Available: http://www.railway-technical.com/trains/rolling-stock-index-l/bogies.html. Accessed 3 June 2024
8. Kunene, B.B., Telukdarie, A.: Analysis of failure mode effects of the traction motor. In: 2022 IEEE 28th International Conference on Engineering, Technology and Innovation (ICE/ITMC) and 31st International Association For Management of Technology (IAMOT) Joint Conference, no. 28th & 31st, pp. 1–6 (2022)
9. Andrews, J.D., Moss, T.R.: Reliability and Risk Assessment. The American Society of Mechanical Engineers, New York (2002)

Further Reading

10. FasTrak SoftWorks, I.: What is Time-based Maintenance (TbM)? FTMaintenance, 1 March 2020. [Online]. Available: https://ftmaintenance.com/maintenance-management/what-is-time-based-maintenance/. Accessed 10 July 2024
11. AAR, "Association of American Railroads," (2022). [Online]. Available: www.aar.org/wp-content/uploads/2023/06/AAR-Freight-Rail-Emissions-Chart-scaled.jpg. Accessed 13 Aug 2023
12. Federal of Railroad Adminstration , "Code of Federal Regulations," 30 May 2024. [Online]. Available: https://www.ecfr.gov/current/title-49/subtitle-B/chapter-II/part-229. Accessed 20 May 2024
13. ISO: Condition monitoring and diagnostics of machines Vocabulary. International Standards Organisation (2012)
14. CEN, *EN:133067:2017 Maintenance Terminology,* Brussels: European Committee for Standardization, 2018.
15. Europe's Rail, "Europe's Rail," 2018. [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8RAIL%20ii. Accessed 9 Apr 2024
16. SKF, "Principles of rolling bearing selection," 2024. [Online]. Available: https://www.skf.com/mena/products/rolling-bearings/principles-of-rolling-bearing-selection. Accessed 14 Aug 2023
17. SKF Group, Railway Technical Handbook, Volume 2, Drive Systems: traction motor and gearbox bearings, sensors, condition monitoring and services, volume 2 ed., SKF Group (2012)
18. IEEE Power systems reliability subcommittee: Report of large motor reliability survey of industrial and commercial installations Part I. IEEE Trans. Ind.y Appl. **IA-21**(4), 853–864
19. Europe's Rail: Shift2Rail (2015). [Online]. Available: https://projects.shift2rail.org/s2r_ip5_n.aspx?p=FR8RAIL. Accessed 20 Apr 2024
20. Transport General Authority: TGA (2023). [Online]. Available: https://tga.gov.sa/Download/Get/228/RegulationDocuments. Accessed 2 June 2024
21. Baasch, B., Sorribes-Palmer, F., Luber, B., Pietsch, L., Lorenzutti, C., Parida, R.D., Galipienzo, J., Groos, J., Plaza, D.B., Ayuso, F.P.: Deliverable D 3.5 technologies for wagon monitoring systems. In: European Union Funding for Research and Innovation (2018)
22. Banisoleiman, K.: The CPMPS diesel engine management support system. Sheffield (1990)
23. Bernal, E., Spiryagin, M., Cole, C.: Onboard condition monitoring snesors, systems and techniques for freight railway vehicles: a review. IEEE Sens. J. **19**(1), 1558–1748 (2019)
24. Cang, V.T., Tuan, D.D.: Reliability based calculation of repairing interval of locomotive wheel axles. Chongqing (2012)
25. Fumeo, E., Oneto, L., Anguita, D.: Condition based maintenance in railway transportation systems based on big data streaming analysis. Procedia Comput. Sci. **53**, 437–446 (2015)
26. Kahl, N., Thiery, F., Wohlrath, R., Rantatalo, M., Klien, C., Gribi, F.: Deliverable D1.4 wayside vehicle monitoring for condition-based maintenance. In: European Union Funding for Research and Innovation (2022)
27. Kothamasu, R., Huang, S.H., VerDuin, W.H.: System health monitoring and prognostics—a review of current paradigms and practices. Int. J. Adv. Manuf. Technol. **28**, 1012–1024 (2006)
28. Li, C., Luo, S., Cole, C., Spiryagin, M.: An overview: modern techniques for railway vehicle on-board health monitoring systems. Int. J. Veh. Mech. Mobil. **55**(7), 1045–1070 (2017)
29. Massaro, A., Cannella, E., Selicato, S., Dipierro, G., Saponaro, A., Trotta, M.G., Galiano, A.: Railway Components Wear: A Smart Platform for Full Traceability of Maintenance Activities. Italy, Rome (2021)
30. Mattos, I.G., Reis, D.A., Magalhaes, D.J.: Locomotive's spare parts optimization, Hong Kong (2008)
31. Milroy, I.P., Johnson, R.A., Gaertner, P.S.: Performance of AC diesel-electric locomotives in typical Australian tasks, Oakbrook, IL, USA, May (1996)
32. RK, M.: An Introduction to Preventitive Maintenance, Plant Engineering Series ed., Van Nostrand Reinhold, New York (1990)

33. Schneidhofer, C., Dubek, K., Dorr, N.: Robust sensors enabling condition-based maintenance of lubricated components in locomotives and wagons, Lisbon (2023)
34. Smith, A.M.: Reliability-Centered Maintenance. McGraw-Hill, New York (1993)
35. Sun, H., He, D., Zhong, J., Wei, Z., Lao, Z., Shan, S.: Preventitive maintenance optimization for key components of subway train bogie with consideration of failure risk. Eng. Failure Anal. **154**(107634) (2023)
36. Tanwar, M., Raghavan, N., Khanam, S.: Condition based maintenance policy for crankcase lubricating oil in diesel locomotives. In: 2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1593–1598 (2021)
37. Watkins, K.M.: The integration of AC drive technology to Amtrak's passenger locomotives, Pittsburgh, PA, USA (1993)

Application of Learning Curves in Maintenance 4.0: A Comprehensive Review



Prakash Naraian Sharma, Gunjan Soni, and M. L. Meena

Abstract Human learning and forgetting: How it happens and what it looks like are widely studied in a variety of disciplines. Consequently, there is a wealth of research on learning curves (LCs). LCs are a useful tool for managerial decision-making in maintenance because they may characterize how workers' performance improves as a result of practice or repetition. Human learning is crucial for labour-intensive industrial companies, particularly those with high labour costs. Furthermore, even in an industry that is not labour-intensive, technology and organizational learning are commonplace in many firms. Therefore, it should come as no surprise that research in this crucial field is still published in the maintenance literature, even in light of the abundance of earlier studies on learning curves. However, a thorough and methodical analysis of the literature research has never been conducted. The objective of this bibliography is to find out a broad summary of maintenance learning curves. It outlines this particular study field, summarizes its conclusions, identifies its main areas of application, and considers potential future research avenues. In order to accomplish this, a thorough literature study on the uses of learning curves in maintenance is presented in this paper. Initially, a framework comprising standard learning curve (LC) models is created. Following that, the essential qualities of learning curves and their uses in maintenance are noted. The literature is categorized using this methodology. In addition, this study discusses the most significant and instructive papers in each of the main categories and points out potential future research directions.

P. N. Sharma (✉) · G. Soni · M. L. Meena
Department of Mechanical Engineering, Malaviya National Institute of Technology Jaipur, Jaipur, Rajasthan, India
e-mail: pnsharma06@gmail.com

G. Soni
e-mail: gsoni.mech@mnit.ac.in

M. L. Meena
e-mail: mlmeena.mech@mnit.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_10

Keywords Learning curve · Learning · Experience curve · Skill requisition process · Learning progression · Learning path · Learning trajectory · Maintenance · Industry 4.0 · AR or augmented reality · Smart maintenance

1 Introduction

In recent years, achieving a balance between quality, cost, and productivity and, most importantly, the industry has achieved flexibility to meet the needs of the modern marketplace [1]. Competitiveness manufacturing is under pressure due to shorter product lifecycles systems that are moving away from mass production customization. Industry 4.0 is able to meet those requirements and services that make use of the Internet of Things (IoT) and smart systems to utilize knowledge to adapt to dynamic production changes-based plans [2]. The learning curve theory has been used in the past to a wide range of management decision-making areas, including quality, production planning, and control of inventory improvement [3, 4].

Learning curve history: Over centuries, the ideas of forgetting and learning have existed. German psychologist Hermann Ebbinghaus wrote the first academic paper on identifying the forms of forgetting curves and learning curves for individual recollections in 1885. The quantum of papers on the "learning" phenomena that have been written since then is astounding. The search yields nearly 2,929,345 articles in the Scopus database, of which over 2,759,493 are published in engineering publications, of which 1,640,307 are research articles, 930,649 are conference papers, 123,808 are book chapters, and 113,860 are review papers. Further details reveal that 2,808,532 articles have been published in the English language in the Scopus database.

1.1 Advantages of Applying Learning Curve Concepts in Smart Maintenance Management

The learning curve approach has numerous benefits when applied to maintenance management, as illustrated in Fig. 1. Managers who fully comprehend these benefits may discover useful and perceptive management implications. At the detailed planning and operational levels, benefits encompass enhanced accuracy in labor standards, improved realism in monitoring production targets, and refined forecasting of process availability, production output prediction [5], and identification of nonconforming units [6]. Improved lot-sizing models and inventory management [7], improved supplier selection [8] more precise vehicle routing algorithms [9] and better manual order picking processes [10, 11] are some of the advantages for management decision support systems. Applying learning curves—sometimes referred to as "experience" curves in this context—at the strategic level involves making decisions about the best time to setting competitive prices, introduce new products, deciding

how much money to invest to spur product and process innovations, choosing organizational design structures, and vertical integration [12].

1.2 Objective of the Research

The research aims to conduct a systematic literature review of previous studies on learning curves in manufacturing maintenance, synthesizing research outcomes and identifying Primary application domains to guide future learning curve research. While previous studies have mostly employed "narrative" or "meta" review approaches, this study utilizes a Denyer and Tranfield [13] "systematic" approach for more on literature review methodologies [14]. Important models of narrative literature reviews on LCs and their applications are provided by Yelle [4] and Jaber (2006) [4, 13]. Jaber (2006, chapter) and Yelle's [4] review covered various types of learning curves and provided a comprehensive overview of previous research in various areas [4]. Jaber (2006, chapter) further explained that Yelle et al. [4] focus on experience curves, including forgetting models [13]. A set of research articles examining the theory, models, and applications of LCs was presented by Jaber et al. [3, 15]. This study adopts a systematic method with a well-defined and rigorous process for literature selection. Ensuring a futuristic reliable review of learning curves in maintenance, our methodology involves a comprehensive and documented search using relevant keywords in a database. The criteria for including and excluding articles are meticulously documented, as well as the final selection process, to maintain transparency. Section 2 outlines the advanced review methodology employed in this research.

1.3 Framework of the Study

Section 2 introduces an outline for categorizing the review findings and offers a detailed explanation of the systematic search and evaluation methodology, with descriptive statistics for the literature sample collected. In literature review section, highlight the most significant and insightful articles within each major category of the framework. Finally, conclusion part outlines potential future research directions and discusses their possible impact.

Categorizing the bibliography search in maintenance on learning curves start with a brief overview of common models and types (Sect. 2), defining the typical features of learning curves and their submissions in maintenance.

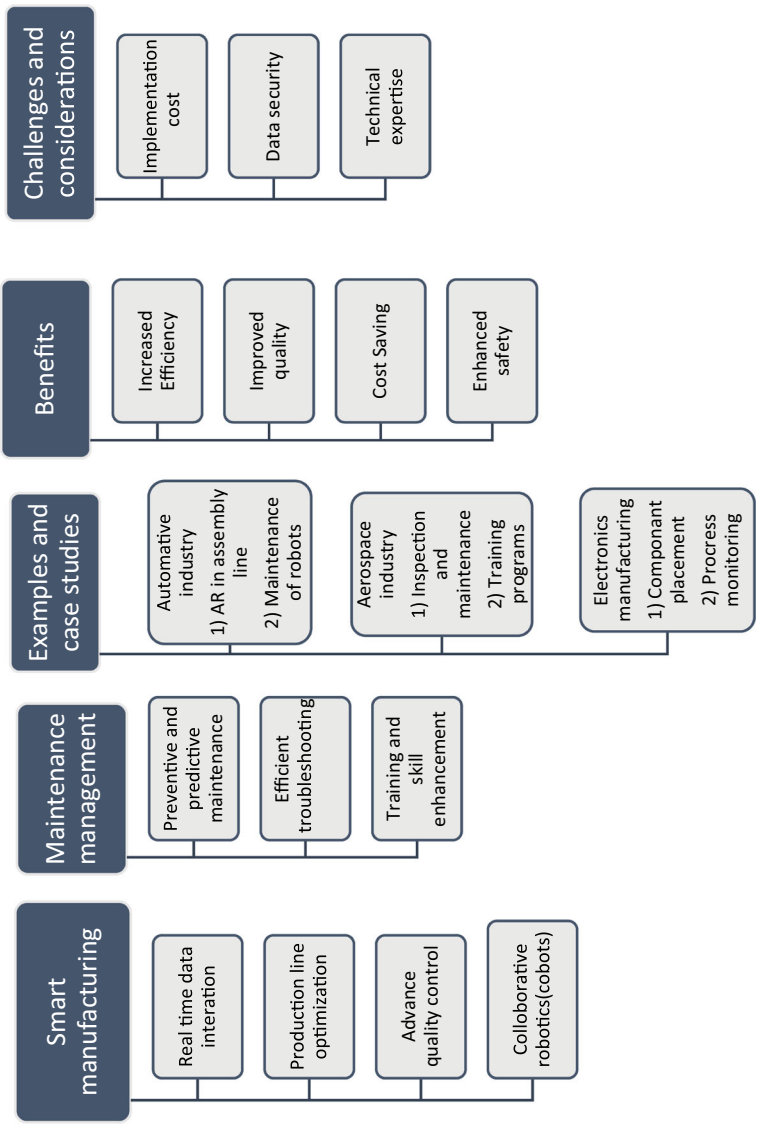


Fig. 1 Different characteristics of applying learning curve concepts in smart maintenance management

1.4 Different Types of Models of Learning Curve

Experiential learning, as an intrinsic mechanism of human development, has been extensively evidenced to enhance cognitive and operational performance through iterative practice and engagement with real-world tasks. Repetition of a task decreases the time needed to complete it. Experimental psychologists were the pioneers in studying this phenomenon, conducting experiments, gathering data, and modeling it with mathematical equations. A learning curve models the underlying trend that aligns with the distribution of observed data points, capturing the progression of performance or efficiency over time. However, the experiences of workers in real-world scenarios differ significantly from those in psychology experiments, where subjects were often exposed to listings of words or visual images.

Wright et al. [16] were the pioneering researchers to document learning phenomena within an industrial framework, demonstrating that with each doubling of production output, a steady and predictable percentage reduction in the amount of time needed to produce a unit was observed [3]. Authors developed one of the most widely recognized learning curves in industrial engineering literature, which applies to the performance of individuals, groups, or organizations. In the literature, learning curves are also referred to by various names, such as "progress functions," "improvement curves" and "startup curves," [16–18]. Following Wright et al. [16] work, various learning curve models emerged in maintenance literature [3]. This section delineates the predominant types, categorizing them into three distinct frameworks: Exponential, log-linear, hyperbolic. Log-linear models, Exponential models, Hyperbolic models, learning models for scheduling, learning and forgetting models.

1.5 Learning Curve Characteristics

The proposed framework for categorising literature on experience or learning curves in manufacturing smart maintenance covering the following characteristics (refer to Grosse et al. [12]), [10].

Studies classify into three categories: experimental, empirical, and theoretical/modeling. Experimental and empirical studies present real-world learning data, which can either be newly collected (from field studies or lab experiments) or previously published. For each study, the industry context (e.g., electronics or automotive) is identified, and it is noted whether curve fitting techniques were applied. Theoretical/modeling works, on the other hand, describe the different type of learning curve models used in literature part.

Articles categorize based on task duration into three groups: continuous learning, interrupted learning, or a combination of learning and forgetting. Additionally, the review tracks whether learning and forgetting are attributed to individuals, groups, or organizations, as indicated in the studies.

Furthermore, the assessment includes whether the articles specify the type of learning involved—motor or cognitive. From a production management perspective, various factors driving learning are examined, such as the number of units produced, setup times, defective items, assembly times, experience, and additional elements like problem identification, project completions, or shipments. In scheduling, learning effects are distinguished as sum-of-processing-time-based, position-based, group-dependent or job-dependent.

1.6 Usability of Learning Curves Areas

Maintenance research is divided into application areas using content category frameworks [19]. The areas are: Production Management: EPQ models balance inventory holding and setup costs to determine lot-sizes [20]. JELS models coordinate shipments to minimise costs for a VBS [21]. Learning curves, which describe how performance improves with experience, have applications across many fields. Key application areas include:

1. Manufacturing and Operations: Production Efficiency: Learning curves are used to predict how production time decreases as workers become more familiar with tasks.

Cost Reduction: Estimating cost reductions over time due to improved labor efficiency.

Process Optimization: Identifying when process improvements or automation can be introduced based on learning plateau points.

2. Human Resources and Training: Employee Training: Estimating how quickly employees will master new skills or technologies.

Skill Retention and Retraining: Assessing how long it takes for employees to relearn skills after periods of non-use (forgetting curves).

Workforce Planning: Predicting labor needs based on learning rates and how long it will take workers to become proficient.

3. Supply Chain and Logistics: Demand Forecasting: Using learning curves to predict future demand for logistics services as experience is gained in new markets or routes.

Sustainable Logistics: Applying learning curves to optimize resource use and minimize environmental impact as logistics processes improve over time.

4. Healthcare and Medical Fields: Surgical Procedures: Monitoring how quickly surgeons and healthcare workers improve with practice in performing new surgeries or procedures.

Treatment Efficiency: Estimating how quickly healthcare organizations can improve patient outcomes as staff members gain experience with new treatments.

5. Software Development and IT: Agile Development: Using learning curves to understand team efficiency improvements in iterative cycles.

System Maintenance: Estimating how quickly IT teams will adapt to new technologies or tools for system updates and troubleshooting.

6. Project Management: Project Timeline Estimation: Predicting how quickly teams can complete tasks based on learning from previous similar projects.

Risk Management: Anticipating risks associated with slower-than-expected learning in new or complex environments.

7. Education: Student Learning: Applying learning curves to evaluate how students' progress over time in mastering new topics or subjects.

Curriculum Design: Adjusting teaching methods based on predicted learning trajectories to maximize student engagement and knowledge retention.

8. Military and Aerospace: Weapon System Maintenance: Predicting maintenance and repair times as personnel gain experience with advanced technology.

Pilot and Operator Training: Modeling how training effectiveness improves over time in complex systems like flight simulators.

9. Energy and Sustainability: Renewable Energy Technology: Learning curves help predict cost reductions and performance improvements in solar, wind, and battery technologies.

Green Manufacturing: Identifying how experience improves the sustainability and efficiency of production processes, contributing to circular economies.

10. Business Strategy: Market Entry: Estimating how long it will take a company to become profitable in a new market based on operational learning.

Economies of Scale: Understanding how production scale and learning interact to reduce unit costs over time.

11. Robotics and Automation: AI and ML Model Training: Predicting the improvement of machine learning models as more data is gathered and processed.

Autonomous Systems: Analyzing how autonomous systems, such as robots, improve their performance through iterative learning.

The application of learning curves is broad and adaptable, making it a powerful tool in fields where experience plays a critical role in improving performance and efficiency.

1.7 Adaptive Assistance of Augmented Reality Work in Industry Fourth Evolution

In the new era of artificial intelligence, an augmented reality (AR)-based assistance system must account for maintainer or operators' characteristics to manage their

cognitive resources effectively, tailoring support to align with their specific profiles (Ariansyah et al. [22]). According to Wickens' Multiple Resources Model, cognitive resources are inherently limited, and their allocation is critical to performance. This underscores the importance of implementing strategies for optimal resource management to mitigate the risk of cognitive overload, which can lead to performance degradation. Nevertheless, cognitive workload management is not the sole approach under investigation. Numerous studies, including those outlined below, advocate for customizing the types of information presented in AR assistance systems to suit the specific use case and the user profile, thereby enhancing usability and effectiveness [23, 24].

2 History of Learning Curves

The first thorough study of learning curves was carried out by Yelle [4], who focused on approaches for curve fitting and parameter estimation [4]. Yelle [4] made significant contributions to the understanding of learning curves by differentiating between organizational learning and labor learning. The survey covered a wide range of learning curve applications, such as purchasing strategies, management control, investment decision-making, forecasting, cost-cutting projects, and non-manufacturing contexts. The field is rich, as Yelle noted when she said, "the possible areas for future research are indeed rich in the LC area." Towill [25] built upon this seminal work by compiling a comprehensive bibliography of LC literature from year 1970 to year 1984, incorporating book chapters, key textbooks, and studies on progress functions and their industrial applications [22, 25].

A survey by Gulledge and Khoshnevis [26] focused on production models that treated the production volume rate as a decision variable quantity and included learning. However, LCs were not a specific focus of their analysis. The literature on learning and forgetting curves, especially in relation to lot-sizing, was thoroughly reviewed by Jaber and Bonney [8]. Jaber updated this review in 2011 [3, 27]. Additionally, Jaber [28] examined several modeling techniques for learning and forgetting curves, going over applications like project management, lot-sizing, and dual resource-constrained systems [13].

Biskup [29] presented a comprehensive overview of the literature on scheduling, incorporating learning effects into the analysis [8], and Azzouz et al. [30] expanded upon the review by developing a classification scheme for scheduling problems that incorporate learning effects [29]. They also conducted a comparison of solution algorithms and various modeling approaches.

Grosse et al. [12] distinguished themselves from most learning curve reviews by performing a meta-analysis of empirical learning curve (ELC) data from standard operating procedures (SOPs) or industrial production processes (IPP). This meta-analysis stands out by gathering datasets from existing studies and applying established learning curves to offer insights into model effectiveness and application areas [11].

Additionally, three notable articles provided overviews of learning curves but were not full literature reviews. Badiru [31] and Hackett [32] using curve fitting methods evaluated well-known learning curves [30, 32], Conway and Schultz [33] examined the difficulties in forecasting future production costs using learning curves [31]. They emphasized the significance of precise data reporting and examined various factors, both in-production and pre-production that production rates and impact costs.

This study adds significantly to the body of knowledge by offering a thorough and organized analysis of learning curves, with a particular focus on maintenance contexts. By identifying key research streams, synthesizing findings, and critically evaluating the body of knowledge across these domains, it offers a thorough analysis. A distinctive feature of our approach lies in the application of a systematic literature review (SLR) and selection process for the literature, ensuring rigor and depth in the review, resulting in a large dataset of 457 works, setting this study apart. The following section explains the detailed methodology behind this systematic literature review.

3 Methodology and Materials

A systematic literature review adheres to a stringent and dependable search and evaluation process [14]. This systematic methodology allows readers to replicate the literature sample generation process and validate review findings. Systematic literature reviews are increasingly popular in closed loop supply chain, manufacturing production and operations management [34, 35].

Our literature search was confined to prominent journals in the field of maintenance, deliberately excluding those from unrelated disciplines such as psychology or medicine, owing to the substantial volume of literature associated with the selected keywords, as demonstrated in Fig. 2.

The number of articles having keyword “Learning curve” OR “Experience curve” that have been written since then is 30,161 in the Scopus database, of which 19,551 are research article, 4137 are conference papers, 3021 are review papers, and 1011 are book chapters, and 3446 are in the engineering area. Out of 30,161 publications, 28,244 are written in English language. If “maintenance” added in to the search database 38 articles are have been published so far.

4 Discussion

A number of articles on the well-liked frame work genres are covered in this area. These pieces were chosen by the writers following a careful evaluation and debate of every submission. The article’s contribution to learning curve theory and practice served as the main criterion. The review of chosen articles focuses on the “maintenance management” area since, as Fig. 2 demonstrates, this is where the majority of papers were assigned.

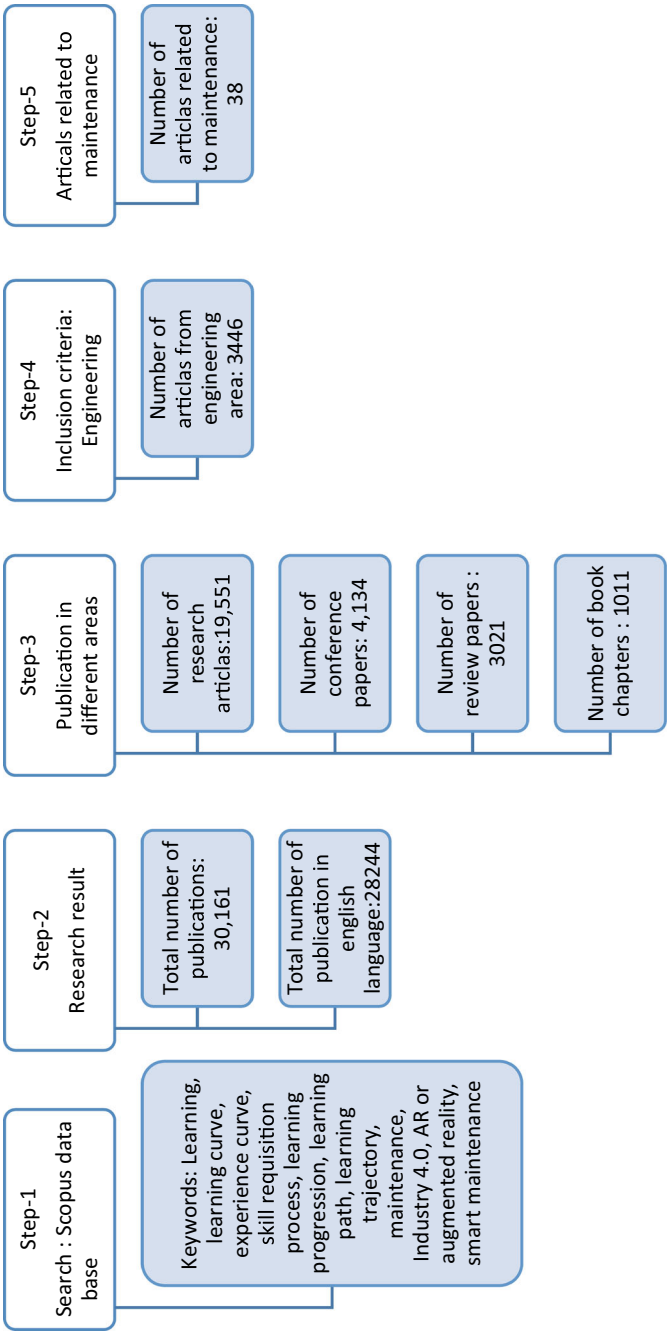


Fig. 2 Research results: Scopus data base

Concept of learning curves: Not many exploration articles examined the hypothetical thinking back to the state of expectations to learn and adapt in assembling. In 1986, Muth presented a hypothesis to make sense of the assembling progress capability (expectation to absorb information) based on an accidental search [36]. He demonstrated how random search theory makes sense of take-offs from the typical log-direct expectation to absorb information capability, specifically how the level of impact and abnormality of log Using this method, linear improvement rates can be achieved.

The origins of the S-Curve learning effect (CLE), specifically for crtitical production projects (CPP), were examined by Cochran and Sherman [37]. They called the first segment of the S-Curve structure a "bulge" because of several contributing causes. Lower rates of education cause the bulge because of the initial increase in workforce, costs associated with tools and equipment debugging and additional labor training, and a lot of changes to the product's design during startup. The authors cautioned against generalizing their results, as they were specific to the company in question.

Researchers used empirical data from various manufacturing units or industries to determine the various factors that influence learning and forgetting. The effort of one of the first comprehensive empirical studies was presented by Nadler and Smith [38]. Studies of the learning rates of various processes. They examined the methods by which nine products were made in three companies and discovered that machine-intensive processes, such as assembly lines, have a lot of learning content, whereas labor-intensive processes have a greater learning content. In several essays, Glover also gave comprehensive explanations of the learning curve fracturing idea (1965, 1966, 1967). These articles looked at how to choose employees and train them and keep an eye on how they're progressing with new procedures in the hopes of avoiding learning plateaus [16, 17, 39].

Bailey [40] explored the factors influencing the use of laboratory experiments, highlighting the impact of neglect. His findings notably revealed that the rate of neglect was not correlated with the learning rate. Additionally, Bailey [40] identified a connection between both the extent of prior knowledge and forgetting and the time elapsed since initial exposure to the material [17, 40].

5 Conclusion

This review synthesises research on learning curves in production management and maintenance, analysing 41 articles to map the field's development. Although learning curves have found widespread application in operations, their use extends beyond this domain, influencing various fields and decision-making processes to track performance improvements, the literature remains heavily focused on theoretical models, particularly the Wright Learning Curve, with limited empirical validation. Studies often emphasize repetition-driven learning, with task duration as a common metric, yet lack exploration into learning and forgetting dynamics.

Key findings reveal substantial attention to production management applications, but little focus on sustainability or social impacts. The field's expansion into learning logistics is similarly limited, with gaps in areas like vehicle routing, delivery efficiency, and supply chain sustainability. Empirical research on forgetting, disruptions and industry-specific knowledge loss is identified as crucial for deeper understanding and practical application, especially in areas like low-level assembly and part fabrication where data granularity is now possible due to improved cost-reporting systems.

Opportunities for future research include investigating deviations in learning curve trends to enhance production planning, understanding learning curve impacts at various production levels, and examining the role of the experience curve in competitive product strategy. The study also underscores the potential value of extending research to sectors like healthcare, where rapid technological changes accelerate learning demands. Although specific search criteria and scope limit the review, it provides a foundational analysis and identifies critical gaps for further investigation.

6 Research Gaps and Recommendations for Future Research

1. **Impact of Forgetting on LCs**

There is a significant gap in understanding the impact of forgetting on learning curves. No research to date has focused solely on how forgetting affects learning speed and retention in maintenance tasks, particularly in environments where tasks are infrequent, or skills are not regularly reinforced. Empirical studies exploring the causes and effects of forgetting are necessary to enhance the applicability of learning curves in such contexts.

2. **Predictability of LCs in Complex Environments**

Another critical gap lies in predictability of learning curves, especially the Wright Learning Curve (WLC), within real-world, complex environments. The literature lacks detailed examinations of deviations from trend lines and the underlying factors driving these variations. Research aimed at managing these deviations could improve the reliability of learning curves as decision-support tools for production planning and control.

3. **Application of LCs in Supply Chain Management and Sustainable Logistics**

Research on the use of LCs in supply chain management and logistics is conspicuously lacking, particularly regarding sustainability. Areas such as vehicle routing, delivery efficiency, and sustainable supply chain performance remain largely unexplored. Addressing this gap could foster more efficient and environmentally sustainable supply chain practices by incorporating learning curve insights into logistics management.

4. Application of LCs in Sustainable Manufacturing Maintenance

The application of learning and experience curves within manufacturing maintenance, particularly in the context of sustainability, remains significantly underresearched. Critical areas such as predictive maintenance efficiency, energy optimization within maintenance operations, and sustainable management of equipment lifecycles have received minimal scholarly attention. Bridging this research gap could drive advancements in manufacturing efficiency and environmental sustainability by embedding learning curve insights into maintenance frameworks. These unexplored areas offer substantial potential for advancing our understanding of learning curves, facilitating their integration into maintenance and operations management, and contributing meaningfully to the sustainability agenda in manufacturing.

References

1. Apostolopoulos, G., Andronas, D., Fourtakas, N., Makris, S.: Operator training framework for hybrid environments: an augmented reality module using machine learning object recognition. Procedia CIRP **106**, 102–107 (2022)
2. Mourtzis, D., Fotia, S., Boli, N., Pittaro, P.: Product-service system (PSS) complexity metrics within mass customization and Industry 4.0 environment. Int. J. Adv. Manuf. Technol. **97**, 91–103 (2018)
3. Jaber, M.Y., El Saadany, A.M.: An economic production and remanufacturing model with learning effects. Int. J. Prod. Econ. **131**(1), 115–127 (2011)
4. Yelle, L.E.: The learning curve: Historical review and comprehensive survey. Decis. Sci. **10**(2), 302–328 (1979)
5. Smunt, T.L.: Incorporating learning curve analysis into medium-term capacity planning procedures: a simulation experiment. Manage. Sci. **32**(9), 1164–1176 (1986)
6. Smunt, T.L.: A comparison of learning curve analysis and moving average ratio analysis for detailed operational planning. Decis. Sci. **17**(4), 475–495 (1986)
7. Lapré, M.A., Mukherjee, A.S., Van Wassenhove, L.N.: Behind the learning curve: linking learning activities to waste reduction. Manage. Sci. **46**(5), 597–611 (2000)
8. Jaber, M.Y., Bonney, M.: The economic manufacture/order quantity (EMQ/EOQ) and the learning curve: past, present, and future. Int. J. Prod. Econ. **59**(1–3), 93–102 (1999)
9. Azadegan, A., Dooley, K.J.: Supplier innovativeness, organizational learning styles and manufacturer performance: an empirical assessment. J. Oper. Manag. **28**(6), 488–505 (2010)
10. Zhong, H., Hall, R.W., Dessouky, M.: Territory planning and vehicle dispatching with driver learning. Transp. Sci. **41**(1), 74–89 (2007)
11. Grosse, E.H., Glock, C.H.: The effect of worker learning on manual order picking processes. Int. J. Prod. Econ. **170**, 882–890 (2015)
12. Grosse, E.H., Glock, C.H., Müller, S.: Production economics and the learning curve: A meta-analysis. Int. J. Prod. Econ. **170**, 401–412 (2015)
13. Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. **14**(3), 207–222 (2003)
14. Abernathy, W.J., Wayne, K.: Limits of the learning curve. Harv. Bus. Rev. **52**(5), 109–119 (1974)
15. Mohamad, Y.: Jaber, Ryerson University, Toronto, Canada, Learning and forgetting Models and Their Applications, Handbook of Industrial and Systems Engineering, 10.1201_9781420038347-43_chapterpdf"
16. Wright, T.P.: Factors affecting the cost of airplanes. J. Aeronaut. Sci. **3**(4), 122–128 (1936)

17. Glover, J.H.: Manufacturing progress functions II. Selection of trainees and control of their progress. Int. J. Prod. Res. **5**(1), 43–59 (1966)
18. Baloff, N.: Startup management. IEEE Trans. Eng. Manage. **4**, 132–141 (1970)
19. Steedman, I.: Some improvement curve theory. Int. J. Prod. Res. **8**(3), 189–206 (1970)
20. Glock, C.H., Grosse, E.H., Ries, J.M.: Reprint of "decision support models for supplier development: systematic literature review and research agenda." Int. J. Prod. Econ. **194**, 246–260 (2017)
21. Glock, C.H., Grosse, E.H., Ries, J.M.: The lot sizing problem: a tertiary study. Int. J. Prod. Econ. **155**, 39–51 (2014)
22. Ariansyah, D., Erkoyuncu, J.A., Eimontaite, I., Johnson, T., Oostveen, A.M., Fletcher, S., Sharples, S.: A head mounted augmented reality design practice for maintenance assembly: toward meeting perceptual and cognitive needs of AR users. Appl. Ergon. **98**, 103597 (2022)
23. Glock, C.H.: The joint economic lot size problem: a review. Int. J. Prod. Econ. **135**(2), 671–686 (2012)
24. Mompeu, G., Danglade, F., Mérienne, F., Guillet, C.: Methodology for augmented reality-based adaptive assistance in industry. Comput. Ind. **154**, 104021 (2024)
25. Towill, D.R.: Transfer functions and learning curves. Ergonomics **19**(5), 623–638 (1976)
26. Gulledge Jr, T.R, Khoshnevis, B.: Production rate, learning, and program costs: survey and bibliography. Eng. Cost Prod. Econ. **11**(1), 223–236 (1987). https://doi.org/10.1016/0167-188x(87)90048-6
27. Towill, D.R.: Management systems applications of learning curves and progress functions. Eng. Costs Prod. Econ. **9**(4), 369–383 (1985)
28. Jaber, M.Y.: Learning and forgetting models and their applications. In Badiru A.B. (ed.), Handbook of Industrial and Systems Engineering, pp. 1–27. Boca Rotan, FL, CRC Press (2006)
29. Biskup, D.: A state-of-the-art review on scheduling with learning effects. Eur. J. Oper. Res. **188**(2), 315–329 (2008)
30. Azzouz, A., Ennigrou, M., Ben Said, L.: Scheduling problems under learning effects: classification and cartography. Int. J. Prod. Res. **56**(4), 1642–1661 (2018)
31. Badiru, A.B.: Computational survey of univariate and multivariate learning curve models. IEEE Trans. Eng. Manage. **39**(2), 176–188 (1992)
32. Hackett, E.A.: Application of a set of learning curve models to repetitive tasks. Radio Electron. Eng. **53**(1), 25–32 (1983)
33. Conway, R.W., Schultz, A.: The manufacturing progress function. Deptartment of Industrial and Engineering Administration, Cornell University (1958)
34. ÜBER DAS GEDÄCHTNIS., "coo-31924003974312-1-1723537194," May 1885, Accessed 21 Aug 2024. [Online]. Available: https://books.google.co.in/books?id=-4_XwPeyP68C
35. Hochrein, S., Glock, C.H.: Systematic literature reviews in purchasing and supply management research: a tertiary study. Int. J. Integr. Supply Manag. **7**(4), 215–245 (2012)
36. Hochrein, S., Glock, C.H., Bogaschewsky, R., Heider, M.: Literature reviews in supply chain management: a tertiary study. Manag. Rev. Q. **65**, 239–280 (2015)
37. Muth, J.F.: Search theory and the manufacturing progress function. Manage. Sci. **32**(8), 948–962 (1986)
38. Cochran, E.B., Sherman, H.A.: Predicting new product labour hours. Int. J. Prod. Res. **20**(4), 517–543 (1982)
39. Nadler, G., Smith, W.D.: Manufacturing progress functions for types of processes. Int. J. Prod. Res. **2**(2), 115–135 (1963)
40. Bailey, C.D.: Forgetting and the learning curve: a laboratory study. Manage. Sci. **35**(3), 340–352 (1989)

A Review on Cable-Driven Soft Robotic Systems Used for Surgical Applications



Samaksh Judson, Chetan Jalendra, and B. K. Rout

Abstract For the last four decades or so, the use of soft cable-driven surgical robots for surgery has excited the scientific world and has elicited breakthroughs in all domains pertaining to it. However, these robots are not being used extensively in all types of surgeries. To better understand the research gaps and limitations of all the technologies which have kept these robots away from practical applications, is explored in this paper. Here, all the relevant technologies and novelty related to the types of materials, kinematics and dynamic modelling, actuation techniques, and control scheme, available in literature are analyzed, keeping in mind the practicality and benefits of their implementation. By enabling intra-body operations without making multiple or large incisions, robotic technology can enhance surgery by augmenting precision, spatial flexibility, and dexterity. Furthermore, with the extensive literature survey, the areas where strong emphasis should be given are highlighted, keeping prevailing designs in mind to improve this niche technologies for unlocking of its true potential.

Keywords Continuum robotics · Kinematics · Dynamics · Sensors · Actuation · Control · Soft materials · Minimally invasive surgery

1 Introduction

Robots commonly used in mainstream applications, which are typically made using inflexible materials that exhibit minimal elastic deformation. While robots that are rigid, hard, and inflexible exhibit traits of accuracy and strength, they do not possess the versatility of living organisms. As the domain of robotics continues to grow

S. Judson · B. K. Rout
BITS Pilani, Pilani, Rajasthan, India

C. Jalendra (✉)
Department of Technology, Jodhpur Institute of Engineering and Technology, Jodhpur, Rajasthan, India
e-mail: jalendrachetan@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_11

and progress beyond its traditional confines of industrial automation and manufacturing, it is increasingly being applied in areas such as healthcare, field exploration, and human–robot collaboration. This has led to the development of more flexible and specialized electro-mechanical systems that are better suited to these new applications.

Quite recently, the robotics field has witnessed the emergence of a novel paradigm known as "soft robotics", which has been developed to integrate human–robot interaction [1] into the domain of biomedical robotics [2]. Through the analysis and replication of the anatomical, physiological, and behavioral characteristics present in biological entities, scientists have been attempting to create soft robots capable of executing tasks with comparable agility and adaptability. Moreover, soft robots possess various benefits over traditional robots, such as their ability to interact safely with humans, their adaptability to wearable devices, and their utilization of uncomplicated gripping systems. Due to their distinct characteristics and benefits, they possess a considerable scope of utilisation [3].

Soft robotic surgery has experienced notable progress and achievements throughout its evolutionary history [4]. During the 1970s, laparoscopic surgery surfaced as an innovative method, fundamentally transforming the realm of minimally invasive interventions. The decade of the 1980s marked the commencement of robotic surgery, thereby establishing the groundwork for procedures that involve robotic assistance. The advent of the da Vinci Surgical System during the 2000s has significantly advanced the field by providing improved precision and control. Concurrently, the notion of Natural Orifice Translumenal Endoscopic Surgery (NOTES) was introduced, affording surgeons the capability to conduct medical procedures via inherent bodily apertures. The year 2007 witnessed the pioneering of translumenal endoscopic surgery, which represented a notable stride forward in the field. During the 2010s, there was an incorporation of pliable and supple materials in NOTES procedures. This integration facilitated surgical interventions that were more flexible and secure.

This chronological progression of research can be corroborated in Fig. 1 as it offers us conclusive evidence how the scientific world has taken to the emergence of soft robotics. We can observe that over the past couple of decades, the number of publications in the domain of soft surgical robotics has been consistently increasing making this an ideal research topic.

In order to successfully perform robotic surgeries, the implements in question must possess several desirable traits which contribute to the safety and reduce the overall complexity of the process [5]. While the ideal properties required may vary depending on the specific application and the desired functionalities, some properties that are commonly desired in soft robots include precision, biocompatibility, versatility and ease of sterilization and maintenance.

While other actuation methods, such as shape memory actuation [6] or fluid elastomer actuation [7], have their own merits, cable-driven actuation [8] stands out due to its ability to optimally balance precision, maneuverability, and compactness all the while being the most economic option on the market in terms of installation and maintenance. The use of cables or wires to transmit forces from external actuators

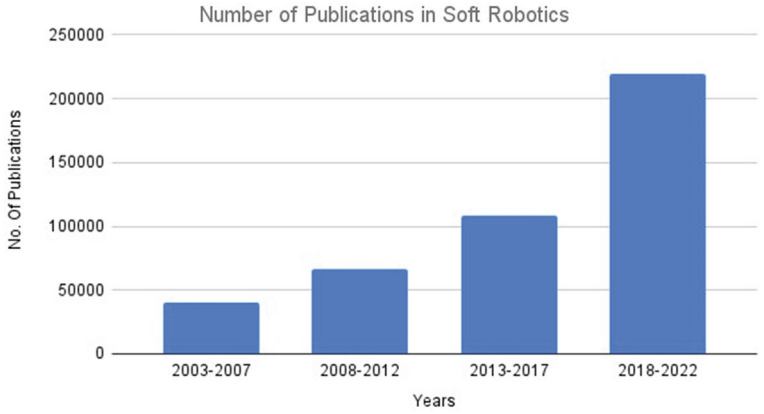


Fig. 1 Number of publications related to soft robotics for surgical applications over the past 20 years

to the soft robot's end effector offers better control and dexterity, closely mirroring the natural movement of a surgeon's hand [9].

The primary objective of this review is to analyze the latest advancements and challenges in cable-driven soft robotics for minimally invasive surgery, with a focus on materials, actuation, control, and clinical impact. As depicted in Fig. 2, the relevant topics covered include materials, kinematics and dynamics, actuation strategies, and sensing and control, all of which are crucial for the advancement of this field. The paper is organized in the following ways: Section 2 critically reviews the materials used for the soft surgical robots. Likewise, Sect. 3 discusses kinematic and dynamic modelling used for soft flexible surgical robots. Moreover, in Sects. 4 and 5, control schemes and actuation strategies that have been developed for soft robot designs are reviewed. Consequently, a thorough examination of these fundamental components with reference to their implementation in current market models, yields a holistic comprehension of the technological progress and obstacles encountered in the realm of soft surgical robotics, thereby promoting additional investigation and ingenuity in this domain.

Case in point, the cable-driven robot developed by J. Shanghai Jiao Tong University for Robot-assisted laparoscopic radical prostatectomy can bend up to an angle of 270° [10]. This helps in overcoming problems related to the accessibility of the blind area typically experienced when performing similar procedures with surgical robots equipped with rigid manipulators. The size of the incision required to conduct the operation is significantly reduced due to the endoscope and surgical instrument integration. The utilization of a master–slave control system in soft robotics is a convenient approach for surgeons, as it effectively reduces the dependence on their manual dexterity and skillset.

In cable-driven soft surgical robots the choice of materials used in these robotic systems plays a fundamental role in shaping their capabilities and ensuring their successful integration into surgical applications. In the next section, the diverse

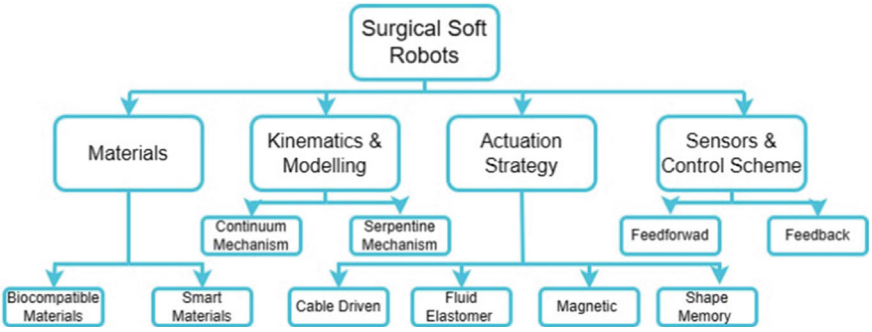


Fig. 2 Outline of the current review study

materials employed in cable-driven soft surgical robots are discussed, exploring their unique characteristics, fabrication methods, and the impact of material choices on the overall performance and functionality of these advanced medical devices. The subsequent section delves into the varied materials utilized in cable-driven soft surgical robots, examining their distinct characteristics.

2 Methodology

To carry out this literature review, we conducted a systematic search across multiple academic databases, including IEEE Xplore, PubMed, and Scopus, focusing on publications related to cable-driven soft robotics for minimally invasive surgery. The selection criteria included studies published within the past two decades to capture both foundational research and recent advancements. Keywords such as "cable-driven soft robotics," "surgical applications," "actuation strategies," "kinematics and dynamics," and "sensing and control" were used to ensure a comprehensive scope. Articles were screened based on their relevance to materials, actuation methods, control systems, and clinical implications in surgical contexts.

3 Materials

3.1 Biocompatible Materials

The implementation of cable-driven actuation involves the utilization of pliable tubes and pulleys to transmit positional information and exertion of force from external actuators to distant locations. The present actuation mechanism exhibits the ability to operate effectively within limited spaces and execute intricate object manipulation

tasks in challenging surroundings. The cables utilized in the required applications are composed of a variety of materials, including stainless steel, polyethene, and super elastic materials such as NiTi. These materials are selected for their exceptional strength, ductility, and corrosion resistance. The polycrystalline state has exhibited a recoverable strain of up to 11 percent, encompassing both elastic and super elastic contributions. This is attributed to the silicone's ability to retain its elastic properties and durability, as demonstrated in a previous study [11].

Soft robots are capable of addressing the issue of elastic property mismatch—which arises due to a difference between the elastic modulus of rigid robotic materials and natural organisms comprising of biological materials through mechano-compatibility, which is achieved by utilizing materials with elastic moduli that fall within the range of those found in biological materials. Due to the materials' high compliance, impact stresses are greatly minimized as contact time is increased due to elastic deformation of material. For this purpose, elastomers [12] have been identified as a viable option for the development of soft robots due to their exceptional mechanical compliance and high strain limit. Additionally, elastomers are commonly filled with conductive materials such as carbon nanofibers or particles to enhance their electrical and thermal conductivity. Soft polymer composites are utilized to actuate electro-active polymers (EAP) such as PVDF and PVDF/CNF.

Silicone based rubbers such as ECOFLEX and have also been used in robots assisting in surgeries involving cardiac ablation. While said robot has only been tested on swine [13], conclusive test results, both in vivo and in vitro, only add to the cadence of the product [14].

3.2 Smart Materials

Heng et al. [15] explores options of adapting smart materials originally intended for use in human–machine interfaces (such as e-skin) for pressure sensing capabilities in soft robotics intended for MIS. These materials can potentially enable the detection of subtle changes in the environment, such as variations in tissue stiffness or the presence of abnormal tissue. By integrating these materials into the robot's instruments or surface, the robot can provide real-time feedback to the operator regarding tissue properties or the presence of abnormalities, enhancing the operator's ability to make informed decisions during the procedure. The predominant force-sensing mechanisms utilized in electronic skins are resistive and capacitive sensing [16]. Piezoresistive electronic skins that are bulky, such as sensors based on sponges, primarily operate by detecting changes in the number of conductive pathways [17] or alterations in the shape of the sensing material due to pressure. Capacitive sensors exhibit superior linearity and lower power consumption in comparison to resistive sensors [16], which modulate their resistance in response to an applied force. However, resistive pressure sensing holds promise for seamless integration with end effector materials, owing to its uncomplicated mechanism and facile data collection approach.

Moreover, the present study investigates the electromechanical coupling of liquid metal embedded elastomers (LMEE) with a fluid dispersion of less than half of the total volume. As such, LMEE can exhibit either conductive or insulating properties. Two typical examples of strain gauges are the Whitney Strain Gauge-Mercury, which is packed in silicon rubber, and the LMEE, which is built of galinstan/silicone [12].

Additionally, the authors also tout the thermal sensing capabilities of conductive nanomaterial-filled polymers, promising their viable candidature for flexible PTC temperature sensors. Here, the volume varies alongside the progression of the material through the melting point of crystalline regions, ultimately varying the resistance [18]. NAVISTAR Thermocool is another similar material of choice when it comes to endoscopic surgery performed via soft surgical robot teleoperation.

Recent works by Iida [19] and Laschi [20, 21] also mention advances in the development of soft robot designs using granular materials, electro adhesives and chemical adhesives. These materials help with the creation of biologically inspired tissue-like continuous materials which possess consistent mechanical properties and can be kinematically analyzed via the continuum method.

The choice of materials employed in cable-driven soft surgical robots has a notable influence not solely on their mechanical characteristics but also on their kinematic conduct and design deliberations. The subsequent segment delves into the kinematic modelling of cable-driven soft surgical robots, examining the complex correlation between material characteristics and the consequent robotic motion, along with the design principles and further considerations.

4 Kinematics and Dynamics

The design mechanisms used to simulate soft robots for minimally invasive surgery are serpentine, continuum [22] and soft mechanisms. Out of these, the serpentine mechanism was one of the earliest to be implemented but has since given way to continuum mechanisms due to their increased compliance and flexibility [23].

To date, soft robotics research within the realm of continuum mechanics has primarily centered around two areas of study: kinematics analysis utilizing the piecewise constant curvature method [24], and static analysis [25, 26]. Duriez [27] proposed a model base on finite element method to address the challenge occur while controlling robots with nonlinear deformations. This approach represents a recent advancement in the field. The utilization of Lagrange multipliers was incorporated in the model at the actuation zones as well as position of the end-effector. The study evaluated the viability of utilizing the Feedforward Neural Network learning approach. The accuracy of the method was assessed using data produced by the mechanical model of the manipulator [28]. The results indicated that the approach achieved an accuracy rate of approximately ninety percent. The methodology employed by the researchers was deemed appropriate for addressing the calculation of inverse

kinematics elements. However, it was observed that each of these models necessitated training to be employed with every individual soft manipulator in order to accommodate for subtle variations between each manipulator.

The latest, most computationally efficient work belongs to Wang et al. [29], wherein the piecewise constant curvature method is built upon, and the Geometrically exact Cosserat rods theory is used to incorporate properties of viscosity and elasticity of the material—while ignoring friction which causes shearing forces—to include out plane bending due to torsion. The paper is able to fit values derived from changes in the elastic and gravitational potential energy density into the governing equation for the dynamic modeling of a manipulator to produce accurate results.

Soft robots pose challenges in describing their state and controlling their movement because they lack joints, setting them apart from traditional rigid robots. To facilitate modeling, the bending shape of a cantilever beam under pure bending has often been approximated as a circular arc. However, since silicone-based soft robots have a much lower elastic modulus compared to cantilever beams, they experience significantly greater and more intricate deformations. To simplify the modeling process, one can assume that each infinitesimally small segment of a structure undergoing slight deflection can be approximated as a circular arc with its own curvature and curvature plane. By defining the parameters for each segment, the overall deformation can be represented. The piecewise constant curvature approach [22] is based on the idea that a soft manipulator can be divided into multiple smaller sections, each behaving according to the principles of rigid-body forward, inverse, and differential kinematics. These segments are derived using arc geometry, where each segment is treated as a planar curve with uniform curvature throughout.

According to the proposed model, the centerline of the soft manipulator is defined by a series of discrete nodes denoted by $\{o_i\}$. Each node corresponds to the centroid of a cross section of the manipulator. The determination of the position coordinates of individual nodes necessitates the establishment of both the global frame and the local frame $\{x_{i-1}, y_{i-1}, z_{i-1}\}$ as shown in Fig. 3.

The z_{i-1} axis is observed to be perpendicular to the cross-sectional plane that corresponds to the i-th section. The manipulator's base is associated with the local frame $\{x_0,y_0,z_0\}$, and is stationary in the global frame. The variables ${}^{i-1}P_i$ and ${}^{i-1}R_i$ denote the local position and orientation, respectively, of crosssection i. These variables are expressed relative to the local frame that corresponds to the section labelled $i-1$. The homogeneous transformation matrix T can be utilized to express the global position ${}^{0}P_i$ and global orientation ${}^{0}R_i$ of cross-section i in a unified manner [23]:

$$ {}^{0}T_i = \begin{bmatrix} {}^{0}R_i & {}^{0}P_i \\ 0 & 1 \end{bmatrix} = \begin{cases} {}^{i-1}T_i & (i=1) \\ {}^{0}T_{i-1} \cdot {}^{i-1}T_i & (2 \le i \le N) \end{cases} \tag{1} $$

The transformation matrix, denoted as ${}^{i-1}T_i$ in Eq. (1), is utilized to express the correlation between the local coordinate frames of nodes $i-1$ and i. The N cross-sectional sequences, indexed by i from 1 to N, represent the 3-D morphology of a

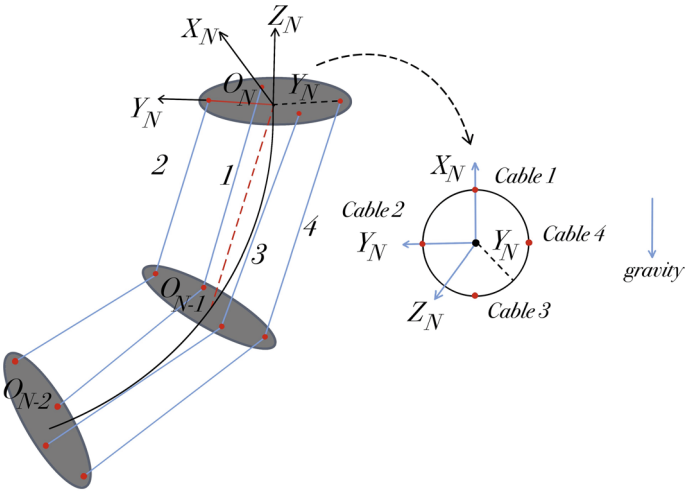


Fig. 3 Representation of local and global frames

compliant manipulator. Global angular velocity ω_i and linear velocity v_i is obtained by computing the derivative of ${}^0R_i$ and ${}^0P_i$ with respect to time [30].

The determination of the global linear acceleration a_i and global angular acceleration α_i of cross-section i involves the computation of the derivative of ω_i and v_i with respect to time. These values, alongside viscous and elastic forces, force due to gravity, driven forces and moments of these forces which were derived via analysis of the potential energy density. This includes the elastic potential, gravitational potential and kinetic energy density alongside external force.

Through the analysis of various force expressions, the governing equations of a soft manipulator can be simplified into a general form of manipulator dynamics as shown in Eq. (2) [29]:

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + F(q, \dot{q}) + G(q) = J(q) \tag{2}$$

The inertia matrix of a soft manipulator is denoted by $M(q)$, while the Coriolis force resulting from the angular velocity is represented by $C(q, \dot{q})$. The damping force $F(q, \dot{q})$ is attributed to the viscosity of the material. The conservative force $G(q, \dot{q})$ is composed of both the gravity and elastic force of the viscoelastic material. The Jacobian matrix $J(q)$ is also present in the system. The aforementioned equation can be utilized to determine joint parameter q_i in the presence of an external torque f.

While the piecewise constant curvature theory could be used to efficiently model a soft manipulator, it did not take torsion into account. This led to the points being modelled being limited to their initial planes. While, the method discussed above serves to amend that and allows the model to account for out-of-plane bending, it has limitations with regard to the consideration of frictional forces between the soft

material and actuating cables. To alleviate these problems, the concept of frictional manipulation [31] may be used in conjunction with this model to provide increase the accuracy of locomotion. The ubiquity of this method, which enables its utilization in friction modulation mechanisms alongside soft manipulator modelling, can be achieved through its implementation in task space as opposed to actuator space [24].

Therefore, examination of the prevalent methodologies pertaining to the kinematic modelling of cable-driven soft surgical robots generates significant insights into the potential improvements that can be made to the overall motion and dexterity of soft robots. Since, the selection of an appropriate actuation strategy is also paramount in achieving the desired kinematic behavior. The subsequent section will explore the diverse actuation strategies utilized in cable-driven soft surgical robots, analyzing their benefits, difficulties, and appropriateness in accomplishing accurate and regulated motion as prescribed by the kinematic prerequisites.

5 Actuation Strategies

Advancements in imaging and mechatronics have led to the development of various serpentine structures that are simple and efficient for single-port access and navigation along complex and curvy anatomical paths. The actuation strategy utilized in flexible surgical robots is contingent upon the specific clinical application for which it has been developed. The characterization of actuation strategies is typically contingent upon various factors, including but not limited to the joint structure, kinematics, dynamics, and back-drivability of a given robotic manipulator. This section presents an in-depth analysis of existing prototypes, with a focus on the metrics under consideration while placing an emphasis on identifying technology that is compatible with cable-driven soft robots [32].

5.1 Cable-Driven Actuation

The utilization of cable-driven mechanisms in flexible surgical robots has been introduced by Hirose's biomechanical snake-like system [33]. This class of robots is ideal for transoral and transanal procedures. This actuation method involves the use of cables or wires to transmit forces from external actuators to the surgical instrument or robot's end effector. Cable-driven actuation offers several advantages, including improved dexterity, reduced size of the instrument, and enhanced precision [8]. Here's how cable-driven actuation is used in minimally invasive surgery:

Cable Routing: The cables are directed from the external actuators, which are frequently situated outside the patient's body, towards the end effector or surgical instrument of the robot. The cables have been engineered to offer the requisite flexibility and strength to ensure accurate manipulation and efficient force propagation. The implementation of the STAR system [34] has incorporated cable-driven actuation

for the control of the soft robotic manipulator's movement. The precise manipulation of tissues during surgical procedures is facilitated by the cables' path along the flexible body of the robot, which enables consistent and controlled movements.

Cable and Pulley Actuation: Cable-driven actuation systems frequently utilize pulleys or sheaves to alter the direction and guide the trajectory of the cables. The utilization of pulleys facilitates the alteration of cable direction, thereby enabling the surgeon to govern the motion and dexterity of the surgical apparatus. Wire-actuated systems are a frequently employed substitute for motor-actuated joints in the realm of soft robots. Robots that possess flexibility are operated through cables and/or wires, resulting in distinct benefits such as remote actuation, high flexibility, and lightweighted designs. Kim et al. [35], introduced a hybrid instrument designed for minimally invasive surgery. The instrument utilized a cable-driven mechanism to enable the end-effector's dexterous movements, while a flexible pulley-based transmission was employed to compensate for elongation during the surgical operation.

Tension Control: Actuation systems that utilize cables are equipped with mechanisms that regulate cable tension. Through the manipulation of tension in various cables, the surgeon is able to govern the placement and motion of the instrument's end effector, emulating the actions of their own hand with heightened accuracy. In addition, the utilization of cables facilitates the manipulation of the tool's terminal actuator [36]. Through the application of varying tensions to the cables, the surgical instrument can achieve flexion, extension, rotation, or articulation in diverse directions, thereby affording the surgeon a broad spectrum of motion and maneuverability. The incorporation of tension control in cable-driven soft surgical robots, such as the da Vinci Surgical System [37], has been shown to enhance surgical precision and reduce the likelihood of errors, ultimately leading to improved patient outcomes.

A recent application of the tension control aspect of cable driven actuation was observed in a surgical endoscopic robot by the Jiang Toh University [38] which employed Bowden Cables—that work on this principle—to actuate their soft manipulator mechanism. Nevertheless, cable-based actuation techniques continue to face significant challenges due to high friction force and backlash hysteresis, which are the primary underlying issues when compared to alternative methods.

5.2 Other Actuation Techniques

While these actuation techniques may not primarily contribute to Cable-Driven Actuation, they possess certain unique advantages [36] which may be used to ascertain possible improvements to cable-driven designs.

Shape Memory Actuation: The implementation of Shape Memory Actuation has the potential to serve as a cable actuator, whereby the contraction force is regulated by temperature modulation, as opposed to the utilization of pulling force from DC actuators. The rotational movement of the grippers is facilitated by the conversion of the translational motion of Shape Memory Actuation cables [39]. Copaci et al. [40] proposed various gripper configurations based on the arrangement of shape

memory alloy (SMA) wires on the grippers. These configurations include parallel SMA grippers, grippers with a twist, net SMA grippers, and spiral SMA grippers, all of which utilize the aforementioned principle. Yuan et al. [6] have designed a single-port surgical robot that is actuated by SMA. The system exhibits the ability to execute movements that are adaptable to different spatial configurations.

Pneumatic Actuation: Pneumatically actuated soft robots employ a pneumatic drive system that is categorized into two distinct types, namely, positive and negative pressure drives [41]. The technique of positive pressure driving involves the operation of a soft actuator through the manipulation of a cavity by means of compressed gas, resulting in the expansion of the main body. In contrast, the opposing approach involves regulating the movement and deformation of soft actuators by means of vacuum-induced reduction of the cavity, resulting in air being drawn out. Furthermore, based on structural distinctions, pneumatic actuators may also be classified into two primary categories: pneumatic networks and fiber-constrained actuators. While the former category relies on a specific arrangement of layers such that inflation causes bending in a specific direction, the latter relies on embedded fibers in the elastic cavity to physically affect the nature of deformation.

Fluidic Elastomer Actuator: This refers to a particular category of actuator that comprises of layers of soft elastomer material that are segregated by a pliant yet comparatively inflexible restraint. Fluidic elastomer actuators have been observed to exhibit a wide range of motion primitives, such as extension, contraction, twisting, and bending [7].

Magnetic Actuator: Among the various actuation strategies that are employed in robots, magnetic actuation is particularly advantageous for the development of microscale robots, including bioinspired microrobots, tether less microrobots, cellular microrobots, as well as electromagnetic soft robots and medical robots [42].

Miscellaneous: Other actuation techniques involve liquid crystal elastomers [43], neuro-stimulation-based actuators [44], photo-responsive actuators [45], etc.

From the above text, it is clear that the utilization of actuation strategies in cable-driven soft surgical robots facilitates the production of the intended motion and forces essential for executing surgical procedures in an efficient manner. However, the overall efficiency and accuracy of robots employing this actuation method can be further improved if they were to make use of the alternative actuation techniques that have been discussed.

The additional utilization of shape memory actuation has the potential to improve a robot's dexterity and adaptability by enabling precise shape manipulation and deformation and the implementation of pneumatic actuation has the potential to yield a substantial force output, swift actuation velocities, and compliance, thereby offsetting the restricted force transmission capabilities of cables. Similarly, the utilization of magnetic actuation has the potential to facilitate actuation without physical contact and enable wireless operation, thereby enhancing maneuverability and decreasing interference with metallic instruments. Moreover, the incorporation of alternative actuation methods, such as hydraulic or electroactive polymers, may improve the robot's abilities by providing supplementary force generation, biomimetic characteristics, etc. Therefore, it is evident how the integration of cable-driven actuation

with other actuation methods enables soft surgical robots to exploit the benefits of each technique, while reducing the limitations of cable-driven actuation.

Since sensing and control systems are essential components in the monitoring and regulation of the actuation process, the next section delves into the varied sensing techniques and control methodologies utilized in cable-driven soft surgical robots, enabling accurate manipulation in the course of surgical procedures.

5.3 Sensing and Control

Since the primary objective of a surgical robot is to replicate the effects of a surgeon's hand with utmost precision, the control system design incorporates biological principles into robotic systems. The premise that human motor performance is accomplished through the interplay of two primary components, namely reactive and anticipatory, is widely supported in the literature as evidenced by [46] with the former being equated to feedback and the latter, feedforward control. Feedforward and feedback control are crucial control strategies utilized in surgical soft robotics to attain accurate and precise manipulation of the robot's movements.

5.4 Feedforward Control

Feedforward control is a control methodology that utilizes predetermined inputs to generate the desired robot movements. The process entails the prediction of necessary control signals, which are then administered directly to the robot's actuators, based on a predetermined task or trajectory. The inputs may be obtained from mathematical models, motion planning algorithms, or expert knowledge [47]. Feedforward control is a frequently employed technique in surgical soft robotics to achieve accurate execution of predetermined trajectories or movements. Through the utilization of predetermined actuation signals, the robot is capable of executing a predetermined path or motion with precision, without necessitating constant input from sensors.

5.5 Feedback Control

The implementation of feedback control, or closed-loop control, entails the ongoing monitoring and modification of the robot's actions in response to sensory feedback obtained in real-time. The employed control methodology leverages data obtained from sensors to quantify the current state or performance of the robot and subsequently contrasts it against a predetermined reference or setpoint. According to this comparison, the feedback control mechanism calculates corrective measures to regulate and sustain the intended behavior. The implementation of feedback control is

critical to surgical soft robotics, particularly when the robot is required to interact with its surroundings or when the environment is subject to unpredictable changes. The acquisition of sensory feedback from diverse sources, including force sensors vision systems and position encoders, enables the system to make ongoing adjustments to its behavior in order to attain the intended surgical results [48]. Position sensors, such as encoders or inertial measurement units (IMUs), are utilized in the tracking of the position and orientation of robotic arms, surgical tools, and anatomical structures. The implementation of these sensors allows for accurate manipulation management and spatial perception in surgical procedures. The utilization of force sensors is employed to quantify the magnitude of force exerted on surgical instruments or the interaction forces that occur between the robot and the patient's tissues via haptic feedback [49]. The utilization of cameras and imaging sensors is a common practice in surgical robotics to facilitate visual feedback of the surgical site. The implemented sensors possess the capability to acquire high-resolution images or video, thereby enabling surgeons to obtain real-time visual feedback of the surgical site [50]. The implementation of feedback control mechanisms allows for the robot to effectively respond to variations, counteract disruptions, and guarantee precise and secure handling of biological materials.

Surgical soft robots utilize a variety of feedback control strategies, including position control [51], force control [52] and compliance control [53], to achieve precise manipulation. The control scheme employs sensor feedback to quantify the state of the robot and modulate the control signals correspondingly [54]. These are used in tandem with feedforward control techniques such as trajectory control, which relies on pre-established trajectory data to publish a preliminary path from which deviations are rectified using the feedback control strategies mentioned above. Hybrid control schemes are frequently utilized in surgical soft robots, which involve the integration of various control strategies [42]. The aforementioned schemes employ a control architecture that follows a hierarchy, wherein diverse control strategies function at varying levels. At a lower level, position control can be employed to achieve accurate positioning, whereas higher-level control techniques, such as force or compliance control, guarantee secure tissue interaction.

The investigation of advanced control techniques such as model-based control, learning-based control, and adaptive control is underway to improve the performance and adaptability of surgical soft robots. A recent study by Zhang et al. [55] made use of these concepts to model a human–robot shared control. The idea behind this was to complement the high-level precision of the robotic manipulator with the human mind's superior cognitive ability and enhance the overall efficiency of surgical procedures. The Learning from Demonstration (LfD) paradigm was used to train the robot for trajectory planning. It involved robots acquiring novel skills via the emulation of human-provided demonstrations [56]. The utilization of Continuous Hidden Markov Model (CHMM) was also implemented in the development of a collaborative control structure for haptic guidance while performing robotic surgery [57]. The data for LfD was generated via a surgical simulator as it proved to be more computationally efficient and cost effective. A sim-to-real adaptation software was then applied to the data to bridge the domain gap between the simulation and

real-time environment and a trajectory transfer algorithm was applied. The training data was then learned by the robot using Dynamic Motion Primitives (DMP) which use nonlinear differential equations to model dynamic systems. Furthermore, a role adaptation framework was developed and integrated to enable adjustability of the relative roles of the human operator and the surgical robot, the general equation for which takes the following form as shown in Eq. (3) [58]:

$$U(t) = \alpha(t)M(t) + (1 - \alpha(t)A(t) \tag{3}$$

This framework has been referred to as a confidence based shared control approach. The robot's overall control input, denoted as $U(t)$, is a result of merging two types of control commands: manual control commands from a human operator, represented as $M(t)$, and autonomous control commands, represented as $A(t)$. The combination is achieved using complementary scales, where the weight of the manual control commands is represented by $\alpha(t)$, which ranges from 0 to 1, and the weight of the autonomous control commands is represented by $1 - \alpha(t)$. The control commands include profiles for position, velocity, and acceleration, as well as force and torque. Various innovative techniques have been proposed to establish the function $\alpha(t)$ as illustrated. The determination of the function $\alpha(t)$ in these methods is reliant upon an independent variable satisfying particular performance criteria during the real-time robotic control task. The temporal evolution of $\alpha(t)$ is subject to various factors, including but not limited to the precision of tracking [59], distance from obstacles and/or desired locations [60], accuracy of forecasting human intent in robot control [61], degree of manipulation accuracy [62], and confidence in the autonomous controller of the robot [63].

Several robot systems including the da Vinci robot system [64], the Raven surgical robot [65] and the surgical endoscopic robot developed by the Jiang Toh University [38] which has been referenced earlier make use of teleoperation for minimally invasive surgeries. While the current testing of this approach has been limited to surface-level procedures, the implementation of a robust framework for controlling human–robot cyborgs is of significant importance in the context of flexible access surgery. The potential for intraoperative analysis and dynamics monitoring during intrabody navigation and manipulation can be realised through the integration of visual and haptic feedback systems. Teleoperation in minimally invasive surgery offers a significant benefit in terms of reducing orthopedic injuries and exposure to hazards in the operation room [66]. Additionally, the implementation of optimal control and feedback systems has the potential to reduce or eradicate these shortcomings. Haptic feedback has been observed to be a potential tool for improving surgical performance in both endoscopic and endovascular systems [62].

In light of previous research [67], it has been established that surgical procedures, including resecting, suturing, ablating, and retracting, typically necessitate the use of multiple surgical tools simultaneously. As a result, there is a need for additional clinical investigations to determine the efficacy of coordinating and manipulating these various tools within a single automation procedure. There remains a need for the development of smart surgical instruments that are lightweight, portable, ergonomic,

and capable of sustained operation. In addition, the majority of presently available soft surgical robots lack the ability to grasp multiple end effectors for concurrent accessibility.

6 Challenges and Research Gaps

Despite impressive developments in the field of cable-driven soft surgical robots, there are still a number of research gaps that need for additional study and creative thinking. This section discusses the research gaps that have been identified in the study, bringing light on topics including novel actuation strategies, sophisticated sensing methods, and enhanced control algorithms to maximize the use of these robots in surgical applications.

6.1 Materials

- The usage of biocompatible materials and their properties were discussed.
- The selection of intelligent materials with built-in sensing capabilities for improved haptic or tactile sensing in modern robots was also highlighted.

6.2 Kinematics and Dynamics

- The need to improve soft robot dynamics and modeling was emphasized via touching upon the limitations of high computation methods.
- A novel study that built upon Serpentine and Continuum Mechanics to account for out-of-plane deformation was thoroughly described.
- The concept of frictional manipulation was proposed to mitigate the limitations of this model. This would not only help with locomotion but also help to account for out of plane bending of the manipulator improving the overall accuracy of the model (Table 1).

6.3 Actuation Strategies

- Cable-driven actuation methods were analyzed based on their working principle and real-life implementation in soft robots.
- The working of alternative techniques like shape memory, pneumatic, fluidic, and magnetic actuation were also discussed.

Table 1 Prevalent commercial soft robots used for surgical applications

Surgical ‘	Place of origin	Actuation strategy	Control scheme	Primary functions
DLR Mirosurge [68]	Germany	Cable-driven	Feedback Control	Minimally Invasive Surgery
i-Snake [69]	United Kingdom	Cable-driven	Feedback Control	Endoscopic Procedures
Robotic Catheter [70]	United States	Hydraulic/ Pneumatic	Hybrid Control	Catheter Insertion and Control
STAR [34]	United States	Pneumatic Actuators	Feedback Control	Soft Tissue Manipulation
Bitrack [71]	Spain	Pneumatic	Feedback Control	Minimally Invasive Surgery
Revo-i [72]	South Korea	Cable-driven	Feedback Control	Minimally Invasive Surgery
Hinotori [73]	Japan	Cable-Driven	Feedback Control	Minimally Invasive Surgeries
RAVEN II [65]	United States	Cable-driven	Feedback Control	Robot-Assisted Surgery
Versius [74]	United Kingdom	Motor-driven	Hybrid Control	Minimally Invasive Surgery
da Vinci [37]	United States	Cable-driven	Feedback Control	Minimally Invasive Surgery

- This knowledge was used to highlight the potential advantages to using the alternative actuation techniques in tandem with cable-driven actuation so as to improve force output, maneuverability and other desirable traits.

6.4 Sensing and Control

- Feedforward and Feedback control strategies were discussed along with their role in robotics so that an improved model of control could be proposed.
- Shared control approach was proposed to combine human operators and autonomous systems and its working was discussed in detail.
- Advantages of shared control strategies over traditional feedforward and feedback schemes was explained and existing soft robot technology that could benefit from employing such an approach was also identified.

The distribution of study focus within the paper’s subjects on cable-driven soft surgical robots can be seen by analyzing the citations as shown in Fig. 4. The proportion of citations for each topic offers useful perceptions into the relative weight and priority given to various research issues.

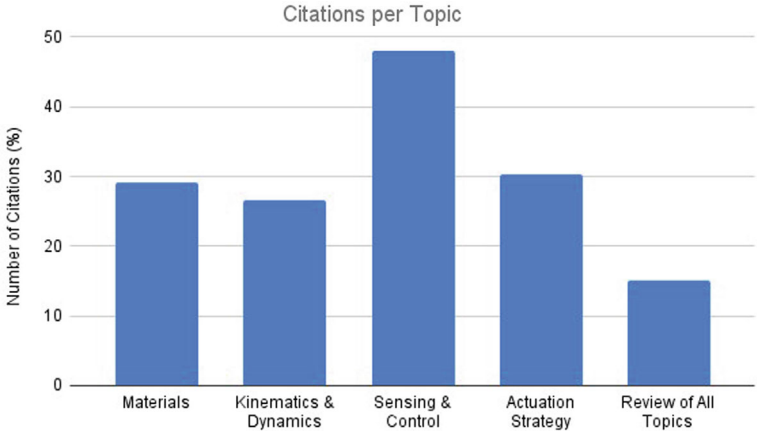


Fig. 4 Percentage of citations contributing to research in each topic

Soft robotics relies on materials, which accounted for 29% of citations. This shows how materials and building methods affect cable-driven soft surgical robots' efficacy and safety. The materials section's many references show the researchers' focus on innovative materials like elastomers, hydrogels, and shape memory polymers to increase these robots' flexibility, biocompatibility, and durability.

The kinematics and dynamics portion had about 26% of citations, emphasizing the importance of understanding cable driven soft surgical robot motion and behavior. Kinematic modelling, analysis, and design are essential for precise and controlled robotic motions. The many references on this subject demonstrate the researchers' dedication to studying kinematics and dynamics and their effects on surgical applications.

Cable-driven soft surgical robots' motion-generating procedures were cited 30% of the time. The references cover shape memory alloys, pneumatic or hydraulic systems, and cable-driven systems. The researchers' focus is on actuation approaches that improve surgical work precision, force generation, and adaptability.

The final section, comprising sensing and control, was referenced by 48% citations, underlining their importance in enhancing cable-driven soft surgical robot functionality and performance. The references in this section emphasize employing cutting-edge sensing technologies, feedback systems, and control algorithms to monitor surgical procedures in real time, improve perception, and manipulate precisely. This emphasizes reliable sensing and control systems for user interaction, safety, and precision.

Since there were a few papers which have been included in more than one category, the total will exceed a hundred percent. Over 15% of references consisted of review papers which touched upon each of the topics that has been discussed and around 10% of all citations consisted of literature that involved more than one topic such as materials and sensors, kinematics and control scheme, etc.

The tabular representation of prevalent commercial soft robots functions as an initial reference for subsequent inquiry and examination. The table serves the purpose of facilitating the identification of significant research patterns, prevalent actuation strategies, and specific surgical procedures that are targeted by soft surgical robots, through the consolidation of information pertaining to such robots. This information can be utilized by researchers to obtain valuable insights into the prevalent methodologies and technologies utilized in the field. It can also aid in identifying areas for improvement or research gaps, and potentially serve as a source of inspiration for novel research ideas.

Cable-driven soft surgical robots are diverse, and the researchers' citations reflect their efforts to thoroughly address materials, kinematics and dynamics, actuation techniques, and sensing and control systems. This study provides a complete overview of the field and promotes cable-driven soft surgical robot design by focusing on each component.

Advancements in cable-driven soft robotics for minimally invasive surgery significantly impact healthcare management and clinical operations. These robotic systems can streamline surgical procedures, reducing operating times and enhancing precision, which leads to improved patient outcomes and shorter recovery periods. This efficiency can also lower operational costs by decreasing bed occupancy and the need for post-operative care. However, the adoption of robotic-assisted surgery necessitates new training programs for surgical teams, requiring hospitals to invest in developing their skill sets. Moreover, the increased accuracy and reduced physical strain on surgeons may lower fatigue levels and human error. To fully benefit from these advancements, healthcare institutions must also invest in the necessary infrastructure and maintenance to support the technology.

7 Conclusion and Future Work

This paper has reviewed the literature and state-of-the-art technologies with a comprehensive analysis of cable-driven soft robots in the context of surgical applications. These robots hold great promise in revolutionizing surgery by enabling minimally invasive procedures and enhancing surgical precision. The review has highlighted the significance of cable-driven actuation in achieving the desired flexibility, dexterity, and maneuverability required for surgical tasks.

This paper delves into various crucial aspects of robotics, including kinematics, materials, actuation, and control. The objective of this framework is to optimize efficacy and accuracy through the amalgamation of data from diverse sources and offering a cohesive methodology for management. Finally, the paper consolidates the information with the help of graphs illustrating the number of publications in the field related to soft robotics for surgical applications over the past two decades. The proliferation of soft robotics literature can be attributed to a variety of factors. The field has garnered attention due to its potential to address the limitations of conventional inflexible robotics in specific applications, including but not limited to intricate

surgical procedures, interactions between humans and robots, and the development of soft grippers for fragile objects. The growth of soft robotics research has been propelled by increased funding and research initiatives, as well as collaborations among academia, industry, and government entities.

References

1. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. Foundations and Trends® in Human-Computer Interaction **1**(3), 203–275 (2007)
2. Hu, C., Shi, Q., Liu, L., Wejinya, U., Hasegawa, Y., Shen, Y.: Robotics in biomedical and healthcare engineering. J. Healthcare Eng. **2017**, 1–2 (2017)
3. Ashuri, T., Armani, A., Hamidi, R.J., Reasnor, T., Ahmadi, S., Iqbal, K.: Biomedical soft robots: current status and perspective. Biomed. Eng. Lett. **10**, 369–385 (2020)
4. Banerjee, H., Tse, Z.T.H., Ren, H.: Soft robotics with compliance and adaptation for biomedical applications and forthcoming challenges. Int. J. Robot. Autom. **33**(1) (2018)
5. Han, E.S., Advincula, A.P.: Safety in minimally invasive surgery. Obstet. Gynecol. Clin. North Am. **46**, 389–398 (2019)
6. Yuan, X., Liu, D., Gong, M.: Design and research on a shape memory alloy-actuated single-port laparoscopic surgical robot. In: 2014 IEEE International Conference on Mechatronics and Automation, pp. 1654–1658 (2014)
7. Majidi, C.: Mechanics of fluid-elastomer systems in soft robotics. In: Robotic Systems and Autonomous Platforms, pp. 425–448, Elsevier (2019)
8. Qian, S., Zi, B., Shang, W.-W., Xu, Q.-S.: A review on cable-driven parallel robots. Chinese J. Mech. Eng. **31** (2018)
9. Gifari, M.W., Naghibi, H., Stramigioli, S., Abayazid, M.: A review on recent advances in soft surgical robots for endoscopic applications. Int. J. Med. Robot. Comput. Assisted Surg. **15** (2019)
10. Li, Y., Miyasaka, M., Haghighipanah, M., Cheng, L., Hannaford, B.: Dynamic modeling of cable driven elongated surgical instruments for sensorless grip force estimation. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), IEEE (2016)
11. Han, R., Li, Y., Zhu, Q., Niu, K.: Research on the preparation and thermal stability of silicone rubber composites: a review. Compos. Part C Open Access **8**, 100249 (2022)
12. Gariya, N., Kumar, P.: A review on soft materials utilized for the manufacturing of soft robots. Mater. Today Proc. **46**, 11177–11181 (2021)
13. Wang, H., Zhang, R., Chen, W., Wang, X., Pfeifer, R.: A cable-driven soft robot surgical system for cardiothoracic endoscopic surgery: preclinical tests in animals. Surg. Endosc. **31**, 3152–3158 (2016)
14. Zhang, W., Deng, Z., Yuan, H., Luo, S., Wen, H., Liu, T.: Preparation and properties of silicone rubber materials with foam/solid alternating multilayered structures. Polym. J. **53**, 619–631 (2021)
15. Heng, W., Solomon, S., Gao, W.: Flexible electronics and devices as human–machine interfaces for medical robotics. Adv. Mater. **34**, 2107902 (2022)
16. Shintake, J., Piskarev, Y., Jeong, S.H., Floreano, D.: Ultrastretchable strain sensors using carbon black-filled elastomer composites and comparison of capacitive versus resistive sensors. Adv. Mater. Technol. **3**, 1700284 (2017)
17. Heng, W., Yang, G., Pang, G., Ye, Z., Lv, H., Du, J., Zhao, G., Pang, Z.: Fluid-driven soft CoboSkin for safer human–robot collaboration: fabrication and adaptation. Adv. Intell. Syst. **3**, 2000038 (2020)
18. Yang, J.C., Mun, J., Kwon, S.Y., Park, S., Bao, Z., Park, S.: Electronic skin: recent progress and future prospects for skin attachable devices for health monitoring, robotics, and prosthetics. Adv. Mater. **31**, 1904765 (2019)

19. Iida, F., Laschi, C.: Soft robotics: challenges and perspectives. Proc. Comput. Sci. **7**, 99–102 (2011)
20. Cianchetti, M., Laschi, C., Menciassi, A., Dario, P.: Biomedical applications of soft robotics. Nat. Rev. Mater. **3**, 143–153 (2018)
21. Laschi, C., Wood, R.J.: Smarter materials for smarter robots. Sci. Robot. **6** (2021)
22. Xu, F., Wang, H., Chen, W., Miao, Y.: Visual servoing of a cabledriven soft robot manipulator with shape feature. IEEE Robot. Autom. Lett. **6**(3), 4281–4288 (2021)
23. Peters, B.S., Armijo, P.R., Krause, C., Choudhury, S.A., Oleynikov, D.: Review of emerging surgical robotic technology. Surg. Endosc. **32**, 1636–1655 (2018)
24. Webster, R.J., Jones, B.A.: Design and kinematic modeling of constant curvature continuum robots: a review. Int. J. Robot. Res. **29**, 1661–1683 (2010)
25. Giorelli, M., Renda, F., Ferri, G., Laschi, C.: A feed-forward neural network learning the inverse kinetics of a soft cable-driven manipulator moving in three-dimensional space. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE (2013)
26. Jones, B.A., Gray, R.L., Turlapati, K.: Three dimensional statics for continuum robotics. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2659–2664 (2009)
27. Duriez, C.: Control of elastic soft robots based on real-time finite element method. In: 2013 IEEE International Conference on Robotics and Automation, pp. 3982–3987 (2013)
28. Giorelli, M., Renda, F., Ferri, G., Laschi, C.: A feed-forward neural network learning the inverse kinetics of a soft cable-driven manipulator moving in three-dimensional space. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5033–5039 (2013)
29. Wang, H., Wang, C., Chen, W., Liang, X., Liu, Y.: Threedimensional dynamics for cable-driven soft manipulator. IEEE/ASME Trans. Mechatron. **22**(1), 18–28 (2017)
30. Rone, W.S., Ben-Tzvi, P.: Continuum robot dynamics utilizing the principle of virtual power. IEEE Trans. Rob. **30**(1), 275–287 (2014)
31. Vikas, V., Cohen, E., Grassi, R., Sozer, C., Trimmer, B.: Design and locomotion control of a soft robot using friction manipulation and motor–tendon actuation. IEEE Trans. Rob. **32**, 949–959 (2016)
32. El-Atab, N., Mishra, R.B., Al-Modaf, F., Joharji, L., Alsharif, A.A., Alamoudi, H., Diaz, M., Qaiser, N., Hussain, M.M.: Soft actuators for soft robotic applications: a review. Adv. Intell. Syst. **2**, 2000128 (2020)
33. Owen, T.: Biologically inspired robots: snake-like locomotors and manipulators by shigeo hirose oxford university press, oxford, 1993, 220ipages, incl. index/i (£40). Robotica **12**, 282–282 (1994)
34. Shademan, A., Decker, R.S., Opfermann, J.D., Leonard, S., Krieger, A., Kim, P.C.W.: Supervised autonomous robotic soft tissue surgery. Sci. Transl. Med. **8** (2016)
35. Lee, C., Kim, M., Kim, Y.J., Hong, N., Ryu, S., Kim, H.J., Kim, S.: Soft robot review. Int. J. Control Autom. Syst. **15**, 3–15 (2017)
36. Li, C., Gu, X., Ren, H.: A cable-driven flexible robotic grasper with lego-like modular and reconfigurable joints. IEEE/ASME Trans. Mechatron. **22**(6), 2757–2767 (2017)
37. Azizian, M., Liu, M., Khalaji, I., Sorger, J., Oh, D., Daimios, S.: The da vinci surgical system. In: Handbook of Robotic and Image-Guided Surgery, pp. 39–55, Elsevier (2020)
38. Li, R., Chen, F., Yu, W., Igarash, T., Shu, X., Xie, l.: A novel cabledriven soft robot for surgery. J. Shanghai Jiaotong Univ. (Science) (2022)
39. Ikuta, K.: Micro/miniature shape memory alloy actuator. In: Proceedings., IEEE International Conference on Robotics and Automation, vol. 3, pp. 2156–2161 (1990)
40. Copaci, D.-S., Blanco, D., Martin-Clemente, A., Moreno, L.: Flexible shape memory alloy actuators for soft robotics: modelling and control. Int. J. Adv. Rob. Syst. **17**, 172988141988674 (2020)
41. Su, H., Hou, X., Zhang, X., Qi, W., Cai, S., Xiong, X., Guo, J.: Pneumatic soft robots: challenges and benefits. Actuators **11**, p92 (2022)
42. Wang, H., Zhu, Z., Jin, H., Wei, R., Bi, L., Zhang, W.: Magnetic soft robots: design, actuation, and function. J. Alloy. Compd. **922**, 166219 (2022)

43. Guo, Y., Zhang, J., Hu, W., Khan, M.T.A., Sitti, m.: Shapeprogrammable liquid crystal elastomer structures with arbitrary threedimensional director fields and geometries. Nature Commun. **12** (2021)
44. Vėbraitė, I., Hanein, Y.: Soft devices for high-resolution neuro-stimulation: the interplay between low-rigidity and resolution. Frontiers Med. Technol. **3** (2021)
45. Wu, J., Ye, W., Wang, Y., Su, C.-Y.: Modeling of photo-responsive liquid crystal elastomer actuators. Inf. Sci. **560**, 441–455 (2021)
46. Wolpert, D.M., Diedrichsen, J., Flanagan, J.R.: Principles of sensorimotor learning. Nat. Rev. NeuroSci. **12**, 739–751 (2011)
47. Shao, X., Wang, S., Yang, L., Liu, N.: Research on feedforward control based on robot dynamics parameters identification. In: 2020 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 973–978 (2020)
48. Roels, E., Terryn, S., Brancart, J., Sahraeeazartamar, F., Clemens, F., Assche, G.V., Vanderborght, B.: Self-healing sensorized soft robots. Mater. Today Electronics **1**, 100003 (2022)
49. Gao, H., Ai, X., Sun, Z., Chen, W., Gao, A.: Progress in forcesensing techniques for surgical robots. J. Shanghai Jiaotong Univ. (Science) **28**, 370–381 (2023)
50. Ferrara, M., Soliman, M.K.: Haptics and vision systems for surgical robots. In: Digital Surgery, pp. 87–93. Springer International Publishing (2020)
51. Even, S., Zheng, T., Lin, H., Ozkan-Aydin, Y.: Stable Real-Time Feedback Control of a Pneumatic Soft Robot (2023)
52. Haidegger, T., Benyó, B., Kovács, L., Benyó, Z.: Force sensing and force control for surgical robots. In: IFAC Proceedings Volumes, vol. 42, no. 12, pp. 401–406 (2009)
53. Khan, S.G., Herrmann, G., Grafi, M.A., Pipe, T., Melhuish, C.: Compliance control and human–robot interaction: part 1—survey. Int. J. Humanoid Rob. **11**, 1430001 (2014)
54. Santina, C.D., Bianchi, M., Grioli, G., Angelini, F., Catalano, M., Garabini, M., Bicchi, A.: Controlling soft robots: balancing feedback and feedforward elements. IEEE Robot. Autom. Mag. **24**, 75–83 (2017)
55. Zhang, D., Wu, Z., Chen, J., Zhu, R., Munawar, A., Xiao, B., Guan, Y., Su, H., Hong, W., Guo, Y., Fischer, G.S., Lo, B., Yang, G.-Z.: Human-Robot Shared Control for Surgical Robot Based on Contextaware Sim-to-Real Adaptation (2022)
56. Zhang, D., Li, Q., Zheng, Y., Wei, L., Zhang, D., Zhang, Z.: Explainable hierarchical imitation learning for robotic drink pouring. IEEE Trans. Autom. Sci. Eng. **19**, 3871–3887 (2022)
57. Power, M., Rafii-Tari, H., Bergeles, C., Vitiello, V., Yang, G.-Z.: A cooperative control framework for haptic guidance of bimanual surgical tasks based on learning from demonstration. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 5330–5337 (2015)
58. Saeidi, H., Opfermann, J.D., Kam, M., Raghunathan, S., Leonard, S., Krieger, A.: A confidence-based shared control strategy for the smart tissue autonomous robot (star). In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1268–1275 (2018)
59. Chipalkatty, R., Droge, G., Egerstedt, M.B.: Less is more: mixed-initiative model-predictive control with human inputs. IEEE Trans. Rob. **29**(3), 695–703 (2013)
60. Karras, G.C., Bechlioulis, C.P., Fourlas, G.K., Kyriakopoulos, K.J.: A mixed-initiative formation control strategy for multiple quadrotors. Robotics **10** (2021)
61. Dragan, A.D., Srinivasa, S.S.: A policy-blending formalism for shared control. Int. J. Robot. Res. **32**, 790–805 (2013)
62. Rassi, I.E., Rassi, J.-M.E.: A review of haptic feedback in tele-operated robotic surgery. J. Med. Eng. Technol. **44**, 247–254 (2020)
63. Saeidi, H., Wagner, J.R., Wang, Y.: A mixed-initiative haptic teleoperation strategy for mobile robotic systems based on bidirectional computational trust analysis. IEEE Trans. Rob. **33**(6), 1500–1507 (2017)
64. Kang, C.M., Chi, H.S., Kim, J.Y., Choi, G.H., Kim, K.S., Choi, J.S., Lee, W.J., Kim, B.R.: A case of robot-assisted excision of choledochal cyst, hepaticojejunostomy, and extracorporeal roux-en-y anastomosis using the da vinci surgical system. Surg. Laparosc. Endosc. Percutaneous Tech. **17**, 538–541 (2007)

65. Lum, M.J.H., Friedman, D.C.W., Sankaranarayanan, G., King, H., Fodero, K., Leuschke, R., Hannaford, B., Rosen, J., Sinanan, M.N.: The RAVEN: design and validation of a telesurgery system. Int. J. Robot. Res. **28**, 1183–1197 (2009)
66. Choi, P.J., Oskouian, R.J., Tubbs, R.S.: Telesurgery: past, present, and future. Cureus (2018)
67. dos Santos, B.M., Fogliatto, F.S., Zani, C.M., Peres, F.A.P.: Approaches to the rationalization of surgical instrument trays: scoping review and research agenda. BMC Health Services Res. **21** (2021)
68. Hagn, U., Konietschke, R., Tobergte, A., Nickl, M., Jörg, S., Kübler, B., Passig, G., Gröger, M., Fröhlich, F., Seibold, U., Le-Tien, L., Albu-Schäffer, A., Nothhelfer, A., Hacker, F., Grebenstein, M., Hirzinger, G.: DLR MiroSurge: a versatile system for research in endoscopic telesurgery. Int. J. Comput. Assisted Radiol. Surg. **5**, 183–193 (2009)
69. Berthet-Rayne, P., Gras, G., Leibrandt, K., Wisanuvej, P., Schmitz, A., Seneci, C.A., Yang, G.-Z.: The i2snake robotic platform for endoscopic surgery. Ann. Biomed. Eng. **46**, 1663–1675 (2018)
70. Gopesh, T., Wen, J.H., Santiago-Dieppa, D., Yan, B., Pannell, J.S., Khalessi, A., Norbash, A., Friend, J.: Soft robotic steerable microcatheter for the endovascular treatment of cerebral disorders. Sci. Robot. **6** (2021)
71. Khandalavala, K., Shimon, T., Flores, L., Armijo, P.R., Oleynikov, D.: Emerging surgical robotic technology: a progression toward microbots. Ann. Laparoscopic Endoscopic Surg. **5**, 3–3 (2020)
72. Chang, K.D., Raheem, A.A., Choi, Y.D., Chung, B.H., Rha, K.H.: Retzius-sparing robot-assisted radical prostatectomy using the revoi robotic surgical system: surgical technique and results of the first human trial. BJU Int. **122**, 441–448 (2018)
73. Hinata, N., Yamaguchi, R., Kusuhara, Y., Kanayama, H., Kohjimoto, Y., Hara, I., Fujisawa, M.: Hinotori surgical robot system, a novel robot-assisted surgical platform: preclinical and clinical evaluation. Int. J. Urol. **29**, 1213–1220 (2022)
74. Alkatout, I., Salehiniya, H., Allahqoli, L.: Assessment of the versius robotic surgical system in minimal access surgery: a systematic review. J. Clin. Med. **11**, 3754 (2022)

Potential of Artificial Intelligence in Millets Supply Chain Risk Management



Hemant Kumar, Rishi Kumar Jha, and Om Ji Shukla

Abstract Artificial intelligence (AI) is the most frequently used technology in big organized agriculture sector for optimizing farming efficiency and boosting supply chain management towards the maximum distribution of food grains among the users within minimum time and cost along with Machine Learning (MI), robotics and Internet of Thing (IoT). But now the time has come to think about shifting the dependency from traditional food to ancient nutritious food like Millet. But the risk is also involved in Millets Supply Chain (MSC) unlike other supply chain management. The revolution of AI technology also impacted the production- processing distribution and consumption of Millets for low-cost AI solutions. This AI sensor will definitely accelerate the demand and supply of millets suitability in nearby market i.e. soil-crop health monitoring; Increased efficiency of production on low cost; Improved decision-making on the involved factors like irrigation, fertilization, pest control, harvesting and finally processing; Enhanced automated customer service requirement; and easy accessibility of markets, etc. In the entire process, various risks have been involved from production to distribution of millets food grains among consumers in the form of the Millets Supply Chain towards fulfilling the requirement of customers. The social engagement in cultivation, processing, logistics, transportation, distribution, and finally consumption of millet to fulfil the economic and nutritional needs without affecting the environment and integrating them has also been discussed in term of Sustainable supply chain risk management (SSCRM). Attempt to identify the most critical barriers in MSC using DEMETAL decision modelling and minimised its impact using AI. Author established the cause-and-effect relationship in Millet Supply Chain Risk Management (MRSCM) for identified Risk in this research.

H. Kumar (✉) · R. K. Jha · O. J. Shukla
Department of Mathematics and Computing Technology, National Institute of Technology Patna, Patna, India
e-mail: hemantk.phd22.ma@nitp.ac.in

H. Kumar
Department of Mechanical Engineering, National Institute of Technology Patna, Patna, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_12

Keywords Artificial intelligence (AI) · Millets Supply Chain Risk Management (MSCRM) · Critical barriers · MCDA · DEMETAL and cause-effect relationship

1 Introduction

The global population's rapid growth has increased the need for food grains, leading to malnutrition and hunger. India ranks 111th out of 125 countries and the young population prefers working life partners contributes to malnutrition [1]. Millet is a leading alternative, but challenges include the availability of fertile land, irrigation facilities, cropping patterns, environmental and climate issues, trained farmers, pesticide and fertilizer use, quality production, and traditional processing facilities [2–5]. Advanced technology like AI, MI, robotics, and IoT can help overcome these challenges [2, 5–7]. This research explores the potential role of Artificial Intelligence (AI) in the agricultural sector, particularly in the production of nutritional smart food grains like millet. The DEMATEL method of Multi-Criteria Decision Analysis (MCDA) was applied to identify key risk factors for improving the distribution of millet food grain [8]. AI has been used to enhance cultivation and production efficiency, sustainability, and resource management. It has helped monitor crop health, weather forecasting, and yield forecasting, reduced labour costs, and improved logistic management. AI also provides real-time virtual training, advice, and counselling to farmers, promoting sustainable farming practices and optimizing resources, raw materials, seeds, and fertilizer allocation [2, 9–15]. This research highlights the importance of AI in reducing environmental impact and promoting sustainable farming practices.

Nowadays, Emerging technologies such as Cyber-Physical Systems (CPS) and Industrial AI are crucial in enhancing control of the agriculture supply chain, particularly in decision modelling for millet risk supply chain management. Industrial AI involves using artificial intelligence techniques to enhance industrial processes in millet supply chain management, including predictive analytics and optimization algorithms to analyse historical data and predict risks related to millet cultivation and supply chain disruptions. Emerging technologies such as AI and CPS significantly enhance decision modelling in supply chain management by enabling scenario analysis, risk assessment, and real-time decision support. Integration of CPS and AI allows for real-time data processing and decision support, leading to better management of supply chain risks and improved efficiency. The Green Revolution in India has significantly enhanced rice and wheat yields, thereby enhancing the food grain supply and promoting sustainability [16]. The government introduced welfare policies and revamped the public distribution system (PSD) in 1992 and 1997, with joint responsibility from the central and state governments through the Food Corporation of India (FCI) [4–6, 16, 17]. However, hunger and malnutrition persist in the Indian population. To address this, India has declared 2023 as the International Year of Millets (IYoM-2023), supported by 72 other countries [5, 16, 18]. The department is promoting millet production and consumption through "Shree Anna" [5].

Millet food grains compete with wheat, maize, and paddy in the agricultural business, fulfilling the demand and supply for human food and animal feed. The supply chain involves multiple businesses, including investment, resources, manpower, raw materials, technologies, and real-time information [2, 5, 12, 15, 16, 18]. A sustainable millet supply chain is achieved through advanced technology and artificial intelligence, ensuring efficient logistic movement and minimal risk impact on cropping, distribution, and market availability [6, 7, 7, 12, 19, 20]. Risk is a significant aspect of life, affecting project planning, execution, and increased costs. Supply chain risk management (SCRM) is crucial in managing uncertainties and unrealistic situations in logistics and transportation [21]. AI utility can help manage SCRM effectively and efficiently, minimizing costs and material losses [2, 20]. Millets Supply Chain Risk Management (MSCRM) aims to minimize these risks, ensuring a feasible and efficient solution for logistics and transportation [18]. This research identifies risk barriers in the millet supply chain, from farm cultivation to consumption [5]. Challenges include cultivation and production risks, pest and fertilizer use, crop disease outbreaks, environmental conditions, climate change, poor infrastructure, mismanagement of logistic and transport facilities, easy market availability, insurance, and government policy risks, timely payment and credit accessibility, and cultural and social risk barriers. These risks are collected from the literature review [2, 3, 12, 17, 22, 23]. With the help of experts and academicians have identified seven risk barriers in the Millets Supply chain, which include:

The DEMETAL modelling of MCDA was utilized to validate the most significant risk barriers for achieving the goal [25, 26]. The objectives of the research are:

1. This research aims to identify factors influencing Millet's supply chain risk management (MSCRM) systems in cultivating countries and integrate them using artificial intelligence.
2. The goal is to establish a cause-effect relationship among identified drivers to understand their interrelationship impact on MSCRM for improvements.

The study systematically identifies risk-driving factors in Millet's supply chain using artificial intelligence (AI). It develops decision modelling using DEMATEL methodology, prioritizing and categorizing them into cause-effect groups [13, 17, 22, 24]. Sensitivity analysis ensures robustness. AI optimizes supply chain operations, reduces MSC costs, and improves supply chain resilience, thereby minimizing risk [7, 25–28].

2 Methodology

Research Methodology refers to the continuous application of various traditional statistical analysis methods to make suitable decisions for specific case studies. The research methodology has evolved to an Advanced Research Methodology, utilizing multi-criteria decision-making (MCDM) for assessing and analysing supply chain factors across various industries [7]. Multi-Attribute Value Theory

(MAVT), weighted product model (WPM), Analytic Hierarchy Process (AHP), decision-making trial and evaluation laboratory (DEMATEL), Interpretive structural modelling (ISM), and weighted sum model (WSM) are the most popular. DEMATEL modelling involves structured questionnaires for expert opinions, aggregation and normalization, and matrix algebra analysis for understanding complex factors, impact determination, and recommendations, with validation and sensitivity analysis for robustness. This research evaluates risks in Millet's food grains industry using a literature review and an academician's expertise [3, 6, 8, 13, 17, 22, 24, 26, 28]. It identifies seven influencing risk barriers in MSCRM and proposes the MCDM model [7, 15]. The interrelationships among MSCRMs impact decision priorities for real-time supply and demand fulfilment, preventing delays and grain loss. Past studies mainly analyse risk factors in makhana, sugarcane, meat/halal, e-waste, and other industries without discussing technology advanced risk factors [3, 7, 8, 13, 15, 17, 22, 24, 26, 28]. This research utilizes a decision-making and trial evaluation laboratory (DEMATEL) to establish a complex cause-and-effect relationship among risk barriers in MSCRM, thereby minimizing the research gap. The research methodology of DEMATE-MCDA modeling is presented in diagrammatic representations in (Fig. 1).

DEMATEL for MSCRM Analysis the following:

1. Solves complex, intertwined problems.
2. Verifies inter-relationship between factors.
3. Offers specific chart for inter-relationship.
4. Based on technical experts' opinions.
5. Clarifies key factors and specifies priorities.

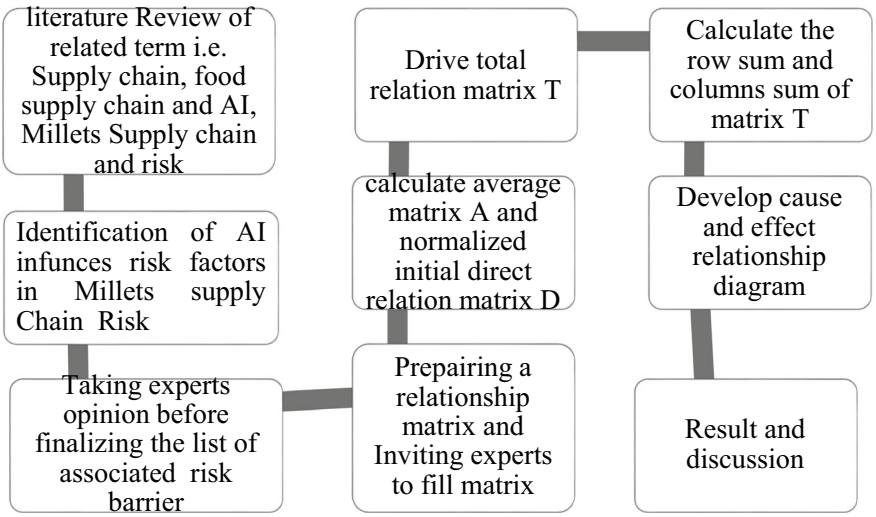


Fig. 1 Flow chart of the methodology. *Source* Author's own work

6. Helps identify practical solutions and cluster of complicated problems.

The sum of the row is denoted by D and the sum of the column is denoted by R through given Eqs. 1–3. Adding D and R represented the relative importance of each of the criteria however subtraction of D and R divides the criteria into two groups in the form of cause group as well as effect group respectively. The sensitivity analysis can be done as follows:

When $(D - R)$ is positive the particular pointed out criteria belonging to the cause group and
When $(D - R)$ is negative the particular pointed out criteria belonging to the effect group

Hence casual diagram can be obtained by mapping the entries of the matrix in a given data set of the $(D + R, D - R)$.

$$T = \left[t_{ij}\right]_{n\times n}, \quad i, j = 1, 2, \ldots, n \tag{1}$$

$$D = \left[\sum_{j=i}^{n} t_{ij}\right] = [t_i]_{nx1}, \tag{2}$$

$$R = \left[\sum_{I=i}^{n} t_{ij}\right] = [t_{iJ}]_{1xN} \tag{3}$$

where I is the identity matrix

Average matrix $A = [a_{iJ}]$ can be constructed as $a_{iJ} = \frac{1}{H}\sum_{k=1}^{H} x_{ij}^{k}$.

Based on the above DEMATEL modeling discussion proposed MSCRM has been analyzed for complex systems with interdependent factors, often used in decision-making and problem-solving processes for which data collection involves focusing on expert opinions from academicians, traders, researchers, and faculty members in the food and millets supply chain. A structured questionnaire is prepared, and seven segregated factors are evaluated using scales. Individual responses are aggregated into a collective dataset, and the total influence matrix is calculated using matrix algebra techniques. The results are interpreted to understand key factors and their interrelationships.

3 Result and Discussion

The DEMATEL method of MCDM has been used to identify the most recognized most influential critical risk barriers of millets supply chain risk management (MSCRM) in the agro-food industry. Nowadays, it is very difficult to control these risk barriers of MSCRM through transportation and logistics starting from source

(farmer) to destination (consumers). After a rigorous literature review around Seventeen critical risk barriers have been investigated and finally a segregated total of seven of the critical risk barriers in the millets supply chain with the help of academician and expert opinion. Accordingly, a relationship matrix has been prepared, inviting a total of ten such experts to fill the matrix. Views of expert opinion have been taken in Integer score from scale 0–3 taken average of all the values and prepared initial direct relationship matrix is the DEMATEL method [8, 20] to examine the inter-relationships among every such critical risk barrier. Further normalized the values of the direct-relation matrix then calculate the total effect matrix T. The results are presented in Tables 1, 2, 3 and 4 as:

Where $T = D(1 - D)^{-1}$.

Cultivation and processing Location; supply and demand management and Delivery time are the major risk characteristics perceived as important by both the producer and the supplier of Millets having easily accelerated by the uses of Artificial Intelligence algorithms, sensor and drown technology for soil testing, moisture contains, environmental and climatic impact along with demand and supply forecasting, the requirement of irrigation and fertilizer including pesticide, etc. even delivery time will be also minimized.

DEMETAL Modelling Results

The result Table 4 is clear that 3 barriers are under the category of cause and 4 barriers are in effect category-based result the cause-and-effect relationship diagram established the values of Ri + Cj versus Ri − Cj and concludes the factors above Ri + Cj is the cause. Three barriers of cause category are barriers are cultivation and processing Location (CB1) supply and demand management (CB3) and delivery time (CB6) are major causes that affect the Millets Supply Chain and required to overcome this risk. four barriers of effect category are warehouse and stocking (CB2), environmental impact (CB4), transporter goodwill (CB5), and uses of advanced technology (CB7). This cause-and-effect barrier impact on MSCRM will be minimized by AI and its related components in every step of supply chain management including transportation, logistics, and distribution. The integration of AI and DEMATEL modeling in millets supply chain risk management enhances decision-making, improves millet risk management strategies, and promotes a resilient and efficient of millet supply chain. When supply chain manager able to minimized the effect of explained barriers as well as uses of advances technology-AI in each and every process of cultivation, processing, logistic and transportation the millet supply chain enhanced the market potentials resultant burden on traditional food will be automatically minimized. The optimization of millets supply chain risk management also addressed the malnutrition with contribution of (Table 5).

The cause-and-effect relationship diagram analyzes the sum values of columns and sum values of rows for MSCRM. The sum and difference values of row sum (Ri) and column sum (Ci) are depicted above. The negative values of (Ri-Ci) represent the effect relationship and its positive values represent the cause relations of associated critical risk barriers in the millet supply chain (Fig. 2).

Table 1 Lists of critical barriers in the millets supply chain

S. No.	Critical Barriers (CBs)	Description
1.	Cultivation and Processing Location (CB1)	AI can minimize risks in millet cultivation and processing by monitoring soil health, controlling pests, and optimizing water management. This, combined with sensors, drones, and smart irrigation systems, boosts productivity [7, 15, 21]
2.	Warehouse and Stocking (CB2)	AI enhances warehouse maintenance and inventory management for millet storage and management. It uses real-time monitoring, automated inventory tracking, pest detection, and warehouse optimization to minimize risks, improve capacity, and optimize millet cultivation and processing [7, 15, 21]
3.	Supply and Demand Management (CB3)	AI can help optimize the millet supply chain by minimizing risks, managing market demand, and ensuring quality control, contributing to transparency, sustainability, and resource management, despite the complexity of the industry [7, 15]
4.	Environmental Impact(CB4)	Artificial intelligence (AI) can significantly improve supply chain risk management (SCRM) by integrating climate and weather data, addressing environmental challenges, and optimizing resource allocation and compliance with regulations [7, 8, 13, 15]
5.	Transporter Goodwill(CB5)	Goodwill is vital for agro-based companies' relationships with transporters, ensuring trust, quality assurance, and punctuality. It's vital for global business opportunities and customer retention in the competitive logistics industry [14, 17, 21–23]
6.	Delivery Time (CB6)	Direct delivery channel strategy, utilizing AI techniques such as drop and GPS-based tracking, has been effectively employed to minimize disruption risk and maximize customer satisfaction. The real-time deliveries prevent shipment quality and quantity loss, thereby enhancing sustainability and building trust in millet food logistics [15, 22–24]
7.	Uses of Advance Technology during Logistic (CB7)	Advanced technology and artificial intelligence enhance millet supply chain efficiency through real-time tracking, quick delivery, and accurate logistics, reducing risks like traffic jams, accidents, theft, and environmental losses [3, 5, 21–23]

Source: author's own work

Limitation and Future scope

These studies contribute to filling a gap in the knowledge of Risk patterns of the Millets supply chain with some limitations open for further research i.e. Study on Risk in Millets Food Supply Chain Management.

Table 2 The matrix of initial direct relations A

A = initial direct relation matrix								
	CB1	*CB2*	*CB3*	*CB4*	*CB5*	*CB6*	*CB7*	SUM
CB1	0	2.7	2.2	2.2	1.9	2.2	1.6	12.8
CB2	2.1	0	1.3	1.9	1	1	1.8	9.1
CB3	2.8	1.5	0	1	1.3	1.9	1.7	10.2
CB4	1.4	1.8	1	0	1	1.5	1.2	7.9
CB5	1.5	1.2	1.4	1.4	0	2	1	8.5
CB6	1.4	1.7	1.6	1.2	2	0	1.7	9.6
CB7	1	1.4	1.3	1.9	1.8	1.3	0	8.7
SUM	10.2	10.3	8.8	9.6	9	9.9	9	66.8

Source Author's own work

Table 3 The matrix of Normalized initial direct relations A'

Normalized initial direct-relation matrix								
		CB1	*CB2*	*CB3*	*CB4*	*CB5*	*CB6*	*CB7*
$D =$	*CB1*	0	0.211	0.172	0.172	0.148	0.172	0.125
	CB2	0.164	0	0.102	0.148	0.078	0.078	0.141
	CB3	0.219	0.117	0	0.078	0.102	0.148	0.133
	CB4	0.109	0.141	0.078	0	0.078	0.117	0.094
	CB5	0.117	0.094	0.109	0.109	0	0.156	0.078
	CB6	0.109	0.133	0.125	0.094	0.156	0	0.133
	CB7	0.078	0.109	0.102	0.148	0.141	0.102	0

Source Author's own work

Table 4 Total effect matrix T

Total relationship matrix									
		CB1	*CB2*	*CB3*	*CB4*	*CB5*	*CB6*	*CB7*	Ri
T =	*CB1*	1.160	1.105	1.093	1.106	1.102	1.111	1.114	7.791
	CB2	0.787	0.827	0.792	0.793	0.798	0.811	0.781	5.588
	CB3	0.875	0.920	0.913	0.916	0.903	0.900	0.889	6.315
	CB4	0.702	0.697	0.695	0.710	0.696	0.693	0.694	4.887
	CB5	0.758	0.768	0.745	0.749	0.762	0.744	0.753	5.278
	CB6	0.854	0.843	0.828	0.847	0.821	0.862	0.828	5.883
	CB7	0.778	0.768	0.752	0.750	0.744	0.769	0.765	5.326
	Ci	5.915	5.928	5.817	5.871	5.826	5.889	5.822	

Source Author's own work

Table 5 Prominent and influence of each barrier

The sum of influences given and received among seven criteria			
Dimension criteria	Ri + Ci	Ri − Ci	Identify
Cultivation and processing location (CB1)	13.7064	1.8762	**Cause**
Warehouse and stocking (CB2)	11.516	−0.3394	Effect
Supply and demand management (CB3)	12.1322	0.4974	**Cause**
Environmental impact (CB4)	10.7588	−0.984	Effect
Transporter goodwill (CB5)	11.1042	−0.548	Effect
Delivery time (CB6)	11.7716	0.0058	Cause
Advance technology (CB7)	11.1482	−0.4964	Effect

Source Author's own work

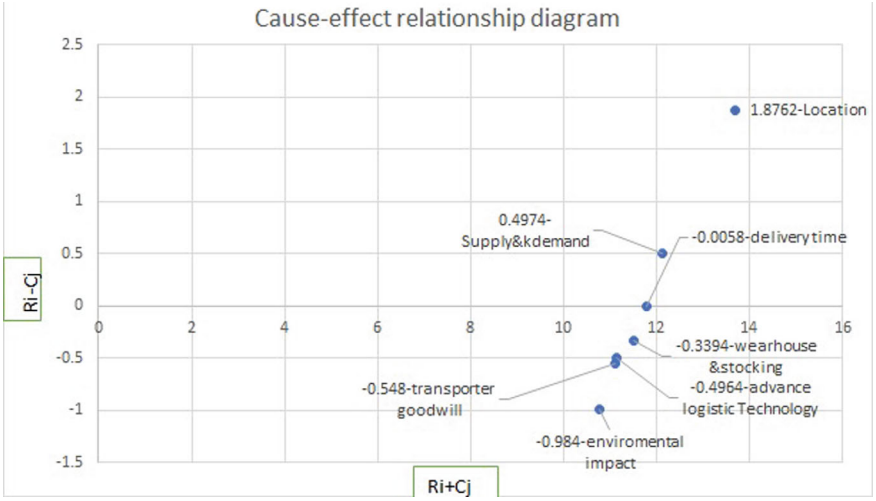


Fig. 2 Cause-effect relationship diagram. *Source* Author's own work.

1. Identifies maximum risk factors in Millets supply chain in India.
2. Determines multidimensional approach to overcome risk.
3. Uses advanced technology like AI, Machine learning, Internet of Things, and block chain to optimize risk.
4. Focuses on self-nutritious food products.

References

1. Global Hunger Index-2023: 18th edition of the GHI "the power of youth in shaping food systems". https://www.globalhungerindex.org/

2. Pandey, A., Bolia, N.B.: Millet value chain revolution for sustainability: a proposal for India. Socio-Economic Plann. Sci. **87**(Part B), 101592 (2023). https://doi.org/10.1016/j.seps.2023.101592
3. Vishwakarma, A., Dangayach, G.S., Meena, M.L., Gupta, S.: Analysing barriers of sustainable supply chain in apparel & textile sector: a hybrid ISM-MICMAC and DEMATEL approach. Cleaner Logistics Supply Chain **5**, 100073 (2022). https://doi.org/10.1016/j.clscn.2022.100073
4. Food Corporation of India—Study to Assess Existing Processes and Suggest Improvements in Supply Chain Management of FCI (2013)
5. 24th Standing Committee Report on Consumer Affairs Food and Public Distribution (2022–2023): Seventeenth Lok Sabha, Ministry of Consumer Affairs, Food and Public Distribution, (Department of Food And Public Distribution), Lok Sabha Secretariat, March 2023
6. Gao, H., Dai, X., Wu, L., Zhang, J., Hu, W.: Food safety risk behavior and social co-governance in the food supply chain. Food Control **152**, 109832 (2023). https://doi.org/10.1016/j.foodcont.2023.109832
7. Wahyuni, H.C., Rosid, M.A., Azara, R., Voak, A.: Blockchain technology design based on food safety and halal risk analysis in the beef supply chain with FMEA-FTA. J. Eng. Res. https://doi.org/10.1016/j.jer.2024.02.002
8. Benabdallaha, C., El-Amraouia, A., Delmottea, F., Frikhab, A.: Evaluation on risks of sustainable supply chain based on integrated rough DEMATEL in Tunisian dairy industry. Int. J. Supply Oper. Manage. **9**(3), 338–359 (2022). ISSN-Print: 2383-1359. ISSN-Online: 2383-2525
9. Shahbaz, M.S., Rasi, R.Z.R.M., Bin Ahmad, M.F., Rehman, F.: What is supply chain risk management? A review. Adv. Sci. Lett. **23**, 9233–9238 (2017)
10. Douglas, L.: The eight essential supply chain management processes. Supply Chain Manage. Rev. 18–26 (2004)
11. The State of Food Security and Nutrition in the World 2023: Food and Agriculture Organization of the United Nations, International Fund for Agricultural Development |United Nations Children's Fund, World Food Programme|World Health Organization, Rome. ISSN 2663-8061 (2023)
12. Rishabh Rathore, J., Thakkar, J., Jha, J.K.: Evaluation of risks in foodgrains supply chain using failure mode effect analysis and fuzzy VIKOR. Int. J. Qual. Reliab. Manage. (2022) ahead-of-print. https://doi.org/10.1108/IJQRM-02-2019-0070s
13. Indian Institute of Millets Chapter. https://www.millets.res.in/aicrp.php. Accessed 21 June 2021
14. Kashyap, A., Kumar, C., Kumar, V., Shukla, O.J: A DEMATEL model for identifying the impediments to the implementation of circularity in the aluminum industry. Decis. Anal. J. **5**, 100134 (2022). https://doi.org/10.1016/j.dajour.2022.100134
15. Kumar, V., Shukla, O.J.: Advancing sustainable e-waste management systems integrated with artificial intelligence in developing countries. Int. J. Intell. Enterp. **1**(1) (2024). https://doi.org/10.1504/IJIE.2024.10063448
16. Rishabh Rathore, J., Thakkar, J., Jha, J.K.: A quantitative risk assessment methodology and evaluation of food supply chain. Int. J. Logistics Manage. **28**(4), 1272–1293 (2017) © Emerald Publishing Limited 0957-4093. https://doi.org/10.1108/IJLM-08-2016-0198
17. Salehi, R., Asaadi, M.A., Rahimi, M.H., Mehrabi, A.: The information technology barriers in supply chain of sugarcane in Khuzestan province, Iran: a combined ANP-DEMATEL approach. In: Information Processing in Agriculture. China Agricultural University. Production and hosting by Elsevier BV (2020)
18. Venkatesh Bhat, B., Dayakar Rao, B., Tonapi, V.A.: The Story of Millets. Karnataka State Department of Agriculture, Bengaluru (2018)
19. The Story of Millets: Millets were the first crops Millets are the future crops Published by: Karnataka State Department of Agriculture, Bengaluru, India with ICAR-Indian Institute of Millets Chapter, Hyderabad, India (2018)
20. El Ayoubi, M.S., Radmehr, M.: Green food supply chain management as a solution for the mitigation of food supply chain management risk for improving the environmental health level. Heliyon-9 (2023). https://doi.org/10.1016/j.heliyon.2023.e13264

21. Wang X.: Food supply chain safety risk evaluation based on AHP fuzzy integrated evaluation method. Int. J. Security Appl. **10**(3), 233–244 (2016). ISSN: 1738-9976 IJSIA
22. Gandhia, S., Mangla, S.K., Kumar, P., Kumar, D.: Evaluating factors in implementation of successful green supply chain management using DEMATEL: a case study. Int. Strategic Manage. Rev. **3**, 96–109 (2015). Holy Spirit University of Kaslik. Hosting by Elsevier BV
23. Shieh, J.-I., Wu, H.-H., Huang, K.-K.: A DEMATEL method in identifying key success factors of hospital service quality. Knowledge-Based Syst. **23**, 277–282 (2010)
24. Das, S.: Supply chain of millets: an FPO perspective (With a Special Reference to Odisha). J. Global Values **XI**(2). ISSN: (P) 0976-9447, (e) 2454-8391 (2020). Impact Factor 7.712(SJIF)
25. Kumar, V., Shukla, O.J.: Major Challenges to Formal E-waste Management Systems in Developing Countries, Garg 2021, pp. 136–147 (2023). https://doi.org/10.46254/an13.20230032
26. Paul, A., Shukla, N., Paul, S.K., Trianni, A.: Sustainable supply chain management and multi-criteria decision-making methods: a systematic review. Sustainability **13**, 7104 (2021). https://doi.org/10.3390/su13137104
27. Shahbaz, M.S., Rasi, R.Z.R.M., Bin Ahmad, M.F., Rehman, F.: What is supply chain risk management? A review, Adv. Sci. Lett. **23**, 9233–9238 (2017)
28. Kumar, V., Shukla, O.J., Tripathi, S.: Analysis of impetus for formal E-waste management system in India. In: Handbook of Evidence Based Management Practices in Business, Routledge, pp. 542–551 (2023)

Sequential Hybrid Solution to Alleviate Congestion Utilizing Distributed Generations and FACTs Devices in the Power Network



Mohammad Ibrar, Charu Sharma, and Pawan Sharma

Abstract The incorporation of renewable energy sources into the power grid is pivotal in advancing a sustainable future. According to recent studies, Distributed generation resources can significantly reduce congestion in power networks. However, if these distributed resources are not planned smartly, they can sometimes aid in congestion and losses in the system. To address this issue, this paper employs a sequential approach to identify the optimal locations and sizes of distributed resources to mitigate congestion. This paper utilizes DC Optimal Power Flow (DC-OPF) methodology to compute Locational Marginal Pricing (LMP). Subsequently, Locational Marginal Prices (LMPs) are employed to ascertain the optimal location and size of the distributed source. The scheme is applied on three bus systems on two different simulation platforms, i.e. on the power world simulator and ETAP.

Keywords Transmission congestion · Congestion management · Renewable Energy Sources · Flexible alternating current system · Optimal power flow · Locational marginal price · Deregulated power market

1 Introduction

The electrical transmission network is crucial for delivering power from generators to consumers. With increasing demand, these networks must often operate at or near their maximum capacity. The modern power system has evolved through strategic reforms due to economic and geographical challenges, but deregulation and private entities exacerbate congestion issues. Additionally, the growing integration of renewable energy sources (RESs) into power grids results in balancing issues and technical problems, such as power network overloading. Transmission congestion (TC) occurs when the system cannot transfer power because transmission lines exceed their designed limits [1, 2]. This congestion can lead to various inefficiencies,

M. Ibrar · C. Sharma · P. Sharma (✉)
Department of Electrical Engineering, The Arctic University of Norway, Tromsø, Norway
e-mail: pawan.sharma@uit.no

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_13

reducing system reliability and security [2, 3]. It also makes the system less adaptable to changes in supply and demand, especially during demand surges. The system's ability to efficiently adjust energy supply in response to demand fluctuations is called system flexibility.

This flexibility allows the power system to make necessary adjustments without shedding load or causing system damage during unexpected events or faults [4]. Congestion disrupts price stability and competition in deregulated markets due to limited capacity and thus reduces system flexibility [5, 6]. Rich literature is available on congestion management (CM) techniques and researchers worldwide present various techniques to avoid congestion [1–7]. Authors in [3] have classified these congestion management techniques under various criteria. Authors in [2, 3] have classified various congestion management methods as cost-free and non-cost-free. Another way of classification is based on the nature of the method applied and is named as technical and non-technical methods of congestion mitigation [2, 3].

Non-technical methods for managing congestion involve strategies that do not require physical upgrades or modifications to the power system infrastructure. Non-technical methods are further classified into market-based and non-market methods. These approaches focus on operational and market-based solutions to alleviate congestion. These solutions include zonal and nodal pricing, market-splitting, counter trading, auctioning, and load curtailment techniques. Nodal prices are geographically variable and referred to as "Locational Marginal Prices" (LMP) [5, 6] under nonuniform pricing. The locational Marginal Price of a bus indicates the cost of supplying additional power, including energy costs, losses, and congestion charges [6]. It is determined through a constrained economic dispatch process. Authors in reference [8] have discussed the LMP-based congestion management system based on the Optimal Power Flow (OPF) solution, determining congestion-relief prices, and communicating these prices to market participants through auctions, bids, and contracts. Similarly, in [9] authors have utilized a DC load flow approach to calculate optimal bus prices and congestion costs. In [10] authors reported that buses with identical LMPs have been grouped into zones in the zonal pricing system. These zones are decided based on their effects on a transmission constraint by using congestion distribution factors (CDFs). According to [2, 11], when CDFs are computed by considering the real and reactive power they are called real power congestion distribution factors (PTCDFs) and reactive power transmission congestion distribution factors (QTCDFs) respectively [2].

On the other hand, technical methods for congestion management use engineering solutions to improve power system efficiency and reliability by upgrading, expanding, or optimizing existing infrastructure. FACTS devices like variable series capacitors and static phase shifters, manage grid congestion by dynamically adjusting power line parameters to optimize power flow and increase system efficiency [2, 3, 12]. Authors in [3, 10] have categorized congestion management techniques as generation side, transmission side, and end-user side methods. According to [13, 14] Generation Rescheduling (GR) adjusts generator outputs to manage congestion in electricity markets, aiming to minimize financial discrepancies but potentially increasing operational costs. Recently demand response management methods are also utilized to

limit congestion in the given network [15]. Further, in [12] Anusha Pillay et. al have classified congestion removal techniques into conventional techniques and intelligent optimization techniques. These intelligent optimization techniques are further divided into swarm-based and evolution-based [3, 12].

Generally, each congestion management technique has its advantages and disadvantages. In some instances, when congestion is not effectively relieved by a single CM method, a combination of methods is preferred to alleviate congestion. Literature is available on various hybrid techniques [2, 13, 16]. For example, Norway's approach integrates Generation re-scheduling (GR) with varied durations to optimize power plant operations and forecast market prices [14]. Further, these congestion management techniques vary from one country to another. For example: Nord Pool handles congestion through market mechanisms, utilizing Available Transfer Capability (ATC) for capacity limits [11–17]. The US model allows market participants to make decisions based on prices and incentives, with regulators ensuring fair practices and using price signals to manage congestion. Nowadays Renewable Energy Sources (RES) are gaining popularity and are emerging as effective, profitable, and viable options for congestion mitigation [2, 18].

Distributed Generators (DGs), significantly reduce power flows on congested transmission lines in smart grids, enhancing grid reliability and efficiency by improving voltage stability and minimizing losses. According to authors in [18], optimal DG placement is crucial for managing power demands and system losses across different load scenarios. In [19], the authors stated that in congested systems with high Locational Marginal Prices (LMPs), DGs can locally meet energy demands, lowering energy costs and easing strain on the grid.

However, careful sizing and placement of DGs are essential to maximize benefits and mitigate operational risks—methods based on Locational Marginal Prices (LMPs) [20]. Optimal DG placement using contribution factors from the bus impedance matrix (Z-bus) is explored in [21] to manage congestion in competitive power systems. This method identifies optimal DG locations to maximize system efficiency and social welfare. A probabilistic approach is adopted in [22, 23] to forecast transmission line overloads and strategically place distributed generation (DG) units to relieve congestion, thereby ensuring optimal planning and scheduling of Energy Storage Systems (ESSs) with renewable energy sources (RES) integration. Electric Vehicles (EVs) are pivotal in managing congestion within distributed power markets and networks. Presenting both challenges and opportunities. Junjie et al. [24] discuss the impact of EV charging on distribution systems, introducing a Fleet Officer (FO) role to coordinate with EV owners and Distribution System Officers (DSOs). The FO schedules EV charging, verified by DSOs through load flow calculations, ensuring grid stability. Market operators use this data for Optimal Power Flow (OPF) to detect and resolve congestion caused by EV charging. EVs, equipped with Vehicle-to-Grid (V2G) technology, address network congestion and support renewable energy storage [25]. Lopez et al. [26] propose integrating V2G-enabled EVs into power systems, allowing energy trading and congestion management through bid-based auctions. EVs adjust battery levels to alleviate line congestion, contributing to system stability using Power Distribution Factors (PDFs). Authors in [27, 28] focus

on increasing transmission capacity for accommodating high levels of wind energy. Ming et al. [7] propose using the Monte Carlo technique to calculate Transmission Congestion (TC) indices, assigning responsibility to failed components. This method tests system flexibility optimizing stability and resilience.

Based on the extensive literature review above, it is found that hybrid congestion management (CM) methods will be more beneficial in the future. Therefore, this paper focuses on a hybrid approach that comprises systematic placement of DR in the given network considering system requirements. First, a solar source is placed in the system and then a wind source is placed in the network. Since the system considered is very small, so wind source is placed instead of solar, and the last both sources are placed to increase system the system's flexibility.

The paper is organized as follows. Section 2 describes the methodology and Algorithm. Section 3 describes the case studies. Section 4 contains Results and discussion, and Sect. 5 concludes the paper.

2 Methodology and Algorithm

The paper employs DC Optimal Power Flow (DC-OPF) methodology as described in [26] to compute Locational Marginal Pricing (LMP). Like conventional OPF, DC-OPF also considers constraints from generator and load bids, including network constraints. However, it only considers the active power [27]. Since OPF is focused on finding an optimal solution according to an objective function, it is subject to a set of nonlinear equality and inequality constraints. The basic objective function gives the total operation cost and is defined as given [7].

$$\min \sum_{i=1}^{Ngen} C\left(i, p_g(i)\right) \tag{1}$$

Here, $C\left(i, p_g(i)\right)$ is the cost for the real power from generation from the i-th unit and Ngen is the number for the generating units. Usually, we use linear programming to handle control systems but here we are working with the quadratic programming to include the dynamics of the system. The energy cost function is quadratic and here it is defined as follows:

$$C_i\left(P_{G_i}\right) = a_i + b_i \times \left(P_{G_i}\right) + c_i \times \left(P_{G_i}\right)^2 \tag{2}$$

Here P_{G_i} Is the output power generation on the i-th generator where "a, b, c" are fuel cost coefficients, where "a" is a constant while "b" and "c" are linear and quadratic product coefficients respectively. Usually during calculations "a" is omitted as it does not have any effect with the change in power generated.

$$\sum_{j=1}^{Nbus}\left[p_{DR}(j) - p_d(j)\right] - \sum_{i=1}^{Nbus}\left[p_g(i)\right] = 0 \quad (\gamma \geq 0) \tag{3}$$

Here in Eq. (3), *Nbus* represents the number of buses, p_g denotes the generation at a bus, p_d indicates the load at a bus while p_{DR} represents the generation from distributed resources (DR). The symbol γ signifies the constraint limit or the constraint on the system. It could be a physical limitation like a transmission line capacity or a generation capacity constraint.

$$-p_g^{\max}(i) + p_g(i) \leq 0 \quad \forall i \tag{4}$$

$$p_g^{\min}(i) - p_g(i) \leq 0 \quad \forall i \tag{5}$$

$$-p_{dr}^{\max}(j, ds) + p_{dr}(j) \leq 0 \quad \forall j \tag{6}$$

LMP_i is the locational marginal price at node i, λ_i is the base energy price at node i, GSF_{ij} is the generational shift factor between nodes i and j, PG_j is the generation at node j, PD_j is the load (power demand) at node j and N is the total number of nodes in the power system. claims that the system congestion can be reduced using LMP and bidding based DR selection.

$$LMP_i = \lambda_i + \sum_{j=1}^{N} GSF_{ij} \times \left(PG_j - PD_j\right) \tag{7}$$

Equation 8 represents the DC Locational Marginal Price (DC-LMP) model, which accounts for congestion within a power system. In this equation, represents the component of marginal energy.

$$P_k = \lambda + \sum_{k_l=1}^{N_{lines}} \mu_{Line_{kl}} \times \frac{\partial P_{kl}}{\partial P} \tag{8}$$

$$P_k = \lambda + \lambda_{c,k} \tag{9}$$

$\lambda_{c,k}$ in Eq. (9) represents the component of congestion i.e. the locational marginal price at node "k" in the presence of congestion "C". The LMP is a measure of the marginal cost of supplying an additional unit of energy at a specific location, $\mu_{Line_{kl}}$ represents the Lagrange multiplier associated with the constraint on transmission line "k_l". Lagrange multipliers are used in optimization to incorporate constraints into the objective function. In this context, "$Line_{kl}$" represents a transmission line, connecting nodes "k" and "l".

$$\lambda_{c,k} = \sum_{k_l=1}^{N_{lines}} \mu_{Line_{kl}} \times \frac{\partial P_{kl}}{\partial P} \tag{10}$$

Here $\frac{\partial P_{kl}}{\partial P}$ represents the partial derivative of the power flow "P_{kl}" between nodes "k" and "l" concerning the overall power generation "P". In simpler terms, it measures how the changes in power flow on a specific transmission line affect the overall power generation.

By summing over all the transmission lines in the power system, the equation considers the cumulative effect of congestion on all the transmission lines. Putting it all together, the equation essentially stating that the LMP at a specific node $\lambda_{c,k}$ in the presence of congestion is influenced by the Lagrange multipliers associated with transmission line constraints and the impact of changes in power flow on those lines for the overall power generation. The summation captures the cumulative impact of congestion on all transmission lines in the system. Similar to 9 for bus "k", 11 provides the scenario for line "l".

$$p_l = \lambda + \lambda_{C,l} \tag{11}$$

The formulation for the LMP at buses "k" and "l" Is represented by Eq. (8). Here, "λ" provides the normal price, while "$\lambda_{C,l}$" and "$\lambda_{C,k}$" Represent the additional price charged because of congestion near buses "k" and "l", respectively. This additional price compensates for the damage caused during the operation of congested branches. At each bus, individual spot prices are calculated, and the difference in LMP at the buses connected to a congested branch directly reflects the congestion location. When power forecasts rise at a specific node, both spot prices and congestion flow on the constrained transmission line increase simultaneously. So, the difference between LMPs of buses "k" and "l" yields

$$\Delta p_{kl} = \left(\lambda_{C,k} - \lambda_{C,l}\right) \tag{12}$$

Here in Eq. 12, Δp_{kl} is the spot price calculated by subtracting the LMPs of buses "k" and "l". Given that the marginal energy remains consistent across all nodes in the system, it is excluded from the fluctuation in nodal prices. Consequently, the TCR can be calculated as follows:

$$TCR = \sum_{k_l=1}^{N_{line}} |\Delta p_{kl} P_{kl}| \tag{13}$$

Here in Eq. 14 TCR represents the total congestion rent, and △pkl denotes the calculated spot price, while Pkl represents the line power flow from "k" and "l". Typically, it is not sensible to install DG at buses where the power generation exceeds the demand because the LMPs are already low. To simplify the calculation process, we will initially assess the buses using the condition presented in Eq. (14)

$$P_g(j) \leq P_d(j) \quad j = 1 \ldots N_{bus} \tag{14}$$

Equation 14 represents the condition where DG placement is applicable. Pg(j) represents the power generated at bus j, while Pd(j) represents the power demand at bus j, where the value of j ranges from 1 to the number of buses. In summary, the equation indicates that it is never optimal to place a DG on any bus where the generation is already greater than the load. The placement of DG is only applicable on buses with power demand higher than the power generation at that bus. The method described for determining the placement of DRs can be applied to all designs, allowing the calculation of benefits for each point in the system. In [26] the Algorithm is explained in the following steps:

(1) Start.
(2) Import the input data from an Excel file for loading generator and load, and branch data.
(3) Run DC optimal power flow.
(4) From the results extract only those buses that keep the highest LMP.
(5) Enter the system parameters for time and locations for example for all years and buses.
(6) Calculate DRs cost for all planned designs.
(7) Provide Unit Commitment parameters (UCPs) and run OPF again for all load intervals.
(8) Calculate DRs benefit for all intervals along with BCR for DRs of each design at each year.
(9) Extract the data for optimal sizing and DR location.
(10) Stop

3 Case Studies

This paper shows two case studies are conducted. In Case 1—Sequential DG placement is considered to relieve congestion on the IEEE-3 Bus system. In this case first system is simulated without any DR and the system is analyzed to identify any congestion in the given system. The Line and Generation data of the IEEE-3 Bus system which are used in this paper are given in Tables 1 and 2 respectively. Later, DC-OPF is used to determine LMPs, and DG placement is conducted on a standard IEEE-3 Bus System as explained in the previous section and results are compared with [27].

Table 1 Generation data for each bus

Numbers for the Bus	Generation (MW)	a (Rs./MW^2)h	b (Rs./MWh)	c (Rs.)
1	100	4.8	2334	3360
3	100	0.6	660	8700

Table 2 Line data for each line

Line numbers	From Bus	To Bus	Reactance (X) pu	Amperage Rating (A)
1	1	2	0.20	430
2	1	3	0.40	525
3	2	3	0.25	365

Table 3 Configurations of methods

Configuration 1	A solar power source is added to the bus near the congested transmission line
Configuration 2	A wind source is added to the bus near the congested transmission line
Configuration 3	Both solar and wind sources are added to the bus near the congested line

Since, the system is very small, only sequential placement of different types of DGs (i.e. Solar and Wind) is considered. In Case 1—Three different configurations of DRs are considered which is summarized in following Table 3.

In Case 2—the Hybrid Sequential Placement of DG and FACT placement is considered to relieve congestion on the IEEE-30 Bus system, as explained in the algorithm of the previous section.

4 Results for the Simulation

4.1 Three Bus System—Case 1

Three Bus system is simulated both in PWS and in ETAP. DC Optimal Power Flow (DC-OPF) is executed, considering only real power. As explained in the previous section, DG units are added to alleviate congestion. The first three bus systems are simulated in both ETAP and PWS and results as shown in Fig. 1a, b respectively (Table 4).

1, 2, and IV 3 show that line 3, between bus 2 and bus 3, is congested, operating at 97.8% of its 367.4 A limit, risking physical damage with any additional load. To remove this congestion various configurations are simulated.

5 Conclusions

This paper employs a sequential approach to identify the optimal locations and sizes of distributed resources to mitigate congestion. The paper utilizes DC Optimal Power Flow (DC-OPF) methodology to compute Locational Marginal Pricing (LMP). Later, LMPs are utilized to determine the location and size of the distributed source. Three

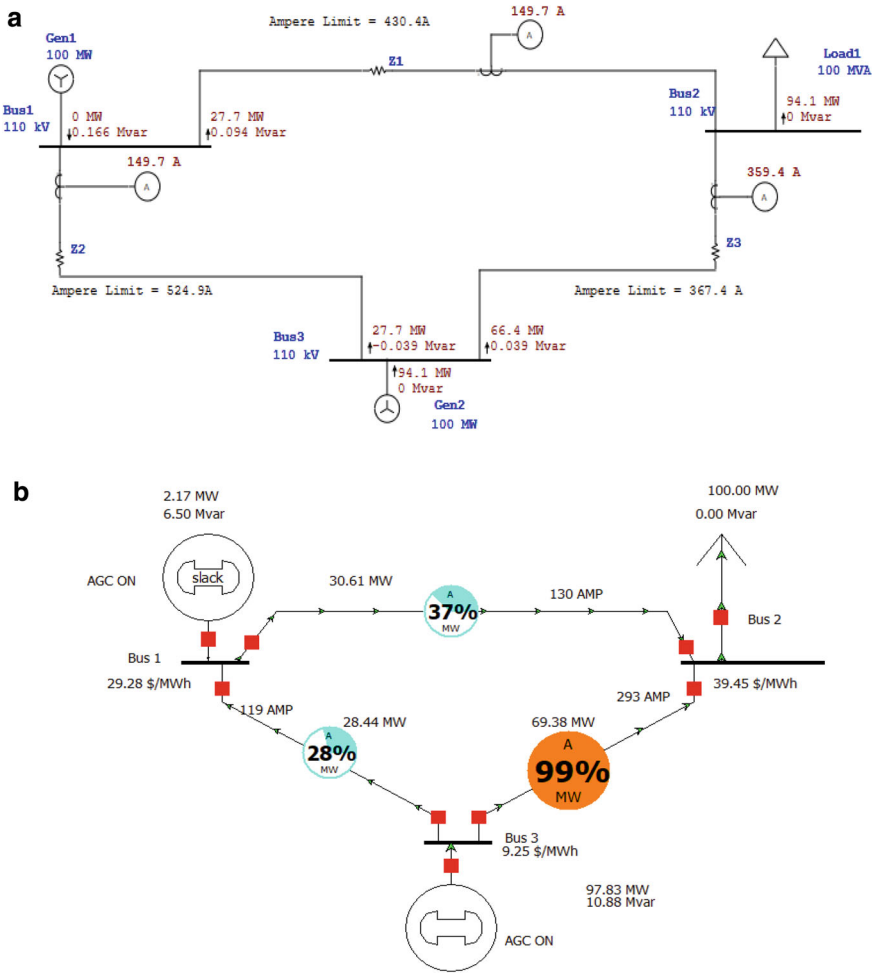


Fig. 1 **a** ETAP simulation for 3 bus system without DG, **b** PWS simulation for 3 bus system without DG

Table 4 ETAP IEEE 3 bus Simulation without DG

Branch no.	From Bus	To Bus	Amperes limit	Ammeter reading	% MW load
1	1	2	430.4	149.7	34.7
2	1	3	524.9	149.7	28.5
3	2	3	367.4	359.4	97.8

different configurations of DRs are considered in this paper. The scheme is applied on three bus systems on two different simulation platforms, i.e. on the power world simulator and ETAP. Obtained results show that third configuration the mitigates congestion best and provides more flexibility. The given scheme applies to higher bus systems.

References

1. Gautam, I., Sharma, G., Ahmer, M.F., Krishnan, N.: Methods and methodologies for congestion alleviation in the dps: a comprehensive review. Energies **16**(4), 1–28 (2023)
2. Gumpu, S., Pamulaparthy, B., Sharma, A.: Review of congestion management methods from conventional to smart grid scenario. Int. J. Emerg. Electr. Power Syst. **20**(3), 1–24 (2019)
3. Gajjala, M.M., Ahmad, A.: A survey on recent advances in transmission congestion management. Int. Rev. Appl. Sci. Eng. **13**(1), 29–41 (2021)
4. Babatunde, O.M., Munda, J.L., Hamam, Y.: Power system flexibility: a review. Energy Rep. **6**, 101–106 (2020)
5. Lai, L.L.: Power System Restructuring and Deregulation: Trading, Performance and Information Technology. John Wiley & Sons (2001)
6. Litvinov, E.: Design and operation of the locational marginal prices-based electricity markets. IET Gener. Transm. Distrib. **4**(2), 315–323 (2010)
7. Gan, M., Xie, K., Li, C.: Transmission congestion tracing technique and its application to recognize weak parts of bulk power systems. J. Modern Power Syst. Clean Energy **5**(5), 725–734 (2017)
8. Glavitsch, H., Alvarado, F.: Management of multiple congested conditions in unbundled operation of a power system. In: Proceedings of the 20th International Conference on Power Industry Computer Applications, pp. 374–380, IEEE (1997)
9. Gedra, T.W.: On transmission congestion and pricing. IEEE Trans. Power Syst. **14**(1), 241–248 (1999)
10. Yu, C.-N., Ilic, M.D.: Congestion clusters-based markets for transmission management. In: IEEE Power Engineering Society Winter Meeting (Cat. No. 99CH36233), pp. 821–832, IEEE (1999)
11. Christie, R.D., Wollenberg, B.F., Wangensteen, I.: Transmission management in the deregulated environment. In: Proceedings of the IEEE, pp. 170–195 (2000)
12. Pillay, A., Karthikeyan, S.P., Kothari, D.: Congestion management in power systems—a review. Int. J. Electr. Power Energy Syst. **70**, 83–90 (2015)
13. Chan, S., Schweppe, F.: A generation reallocation and load shedding algorithm. IEEE Trans. Power Apparatus Syst. **PAS-98**(1), 26–34 (1979)
14. Fosso, O.B., Gjelsvik, A., Haugstad, A., Mo, B., Wangensteen, I.: Generation scheduling in a deregulated system. the Norwegian case. IEEE Trans. Power Syst. **14**(1), 75–81 (1999)
15. Lie, T.T., Deng, W.: Optimal flexible ac transmission systems (facts) devices allocation. Int. J. Electr. Power Energy Syst. **19**(2), 125–134 (1997)
16. Kumar, J., Kumar, N.: Facts devices impact on congestion mitigation of power system. J. Inst. Eng. (India) Ser. B **101**, 239–254 (2020)
17. Bredesen, H.: The Nord pool market model. In: ASEAN Energy Market Integration (AEMI), Forum Paper (2016)
18. Singh, D., Misra, R.K., Singh, D.: Effect of load models in distributed generation planning. IEEE Trans. Power Syst. **22**(4), 2204–2212 (2007)
19. Gautam, D., Mithulananthan, N.: Optimal DG placement in deregulated electricity market. Electric Power Syst. Res. **77**(12), 1627–1636 (2007)

20. Sarwar, M., Siddiqui, A.S.: Zonal congestion management based on locational marginal price in deregulated electricity market. In: Annual IEEE India Conference (INDICON), pp. 1–4. IEEE (2015)
21. Singh, K., Yadav, V.K., Padhy, N.P., Sharma, J.: Congestion management considering optimal placement of distributed generator in deregulated power system networks. Electric Power Components Syst. **42**(1), 13–22 (2014)
22. Masaud, T.M., Deepak Mistry, R., Sen, P.: Placement of large-scale utility-owned wind distributed generation based on probabilistic forecasting of line congestion. IET Renew. Power Gener. **11**(7), 979–986 (2017)
23. Hemmati, R., Saboori, H., Jirdehi, M.A.: Stochastic planning and scheduling of energy storage systems for congestion management in electric power systems including renewable energy resources. Energy **133**, 380–387 (2017)
24. Hu, J., Si, C., Lind, M., Yu, R.: Preventing distribution grid congestion by integrating indirect control in a hierarchical electric vehicles' management system. IEEE Trans. Transp. Electrification **2**(3), 290–299 (2016)
25. Hu, J., Saleem, A., You, S., Nordstrom, L., Lind, M., Østergaard, J.: A multi-agent system for distribution grid congestion management with electric vehicles. Eng. Appl. Artif. Intell. **38**, 45–58 (2015)
26. Lopez, M., Martin, S., Aguado, J., de la Torre, S.: V2g strategies for congestion management in microgrids with high penetration of electric vehicles. Electric Power Syst. Res. **104**, 28–34 (2013)
27. Sreenivasulu, J., Varma, P.S., Sankar, V.: Selection of renewable power generation to reduce transmission congestion. In: International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 2048–2053. IEEE (2017)
28. Liang, Z., Chen, H., Chen, S., Wang, Y., Zhang, C., Kang, C.: Robust transmission expansion planning based on adaptive uncertainty set optimization under high-penetration wind power generation. IEEE Trans. Power Syst. **36**(4), 2798–2814 (2020)

Understanding the Role of Cyber Physical Technology in Supporting Supply Chain Flexibility: A Systematic Literature Review



Angel Singh, Bharti R. Soni, and Sachin Ghai

Abstract The purpose of this research paper is to undertake an in-depth literature review to evaluate the evolution of Supply Chain Flexibility (SCF). The study contributes to our understanding of SCF, its dimensions, and the technology used or integrated to support it. The study focuses on how technology affects SCF and the dimensions used to measure SCF. It also discusses the theories used to investigate SCF, as well as studies conducted in the Indian setting during the last decade to assist us understand previous work. The process entails developing research questions, identifying and evaluating studies, analyzing and synthesizing them, and presenting the results. The findings show that significant effort has been done in the manufacturing and automotive sectors. RBV and Dynamic capacity theory have been widely used to explore SCF, but little work has been done in India on technology integration (TI) to support SCF. "Systematic literature review" (SLR) is a reliable method supported by a large number of scientific papers. However, it is dependent on the availability & accessibility of studies, as well as the researcher's specific needs. Inclusion and exclusion of research based on specified criteria may result in publication bias, a limitation of SLR. The future scope covers possible study fields in terms of technological integration, sustainability, and SCF.

Keywords Supply Chain Flexibility · Technology · AI, · Analytics · Innovation · Digital · Sustainability

A. Singh (✉) · S. Ghai
Department of Management Studies, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India
e-mail: angelsingh1703@gmail.com

S. Ghai
e-mail: profsg.mba@geu.ac.in

B. R. Soni
Department of Business Management, Poornima University, Jaipur, India
e-mail: bharti.mnit2022@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_14

1 Introduction

The changing business environment, characterized by complexity, ambiguity, susceptibility, and uncertainty, has made it extremely difficult for firms to sustain and manage their operations [1]. Therefore, it is essential for the firms to orient its entire supply chain towards flexibility to attain competitive edge [2]. Through flexible supply chain, the ecosystem can adapt to a volatile business environment [3]. Globalized, innovative, and customer-focused companies, more responsive supplier networks, and more efficient firm-level operations are all possible outcomes of FSCs [4]. From a conceptual perspective, SCF describes the speed and scope of changes made in response to challenges from both inside and outside the company [1]. As a result, SCF is a company's capacity to absorb & reapply changes in the market setting [5]. It's described as combined capacity of all the SC participants and operations to adapt to changing customer demands while maintaining business performance & results. Shorter life cycles & tailored product & service design are the results [1, 6]. Flexibility can be the key to increasing and strengthening competitiveness in a market that is constantly evolving [7, 8]. Many scholars have been developing concepts, theories, and applications to improve flexibility in SC management and company operations in recent years [9–11].

The definition and conceptualization of SCF as a distinct term are points of discussion among SC scholars. The ways in which various organizational flexibility elements can impact perceptions of SC in general are still largely unknown. But SCF is at the top of every corporate leader's agenda. This is due to the fact that companies have to deal with a complex, dynamic, and unpredictable environment that is caused by the following factors: (1) globalization; (2) technological innovation and change; (3) SC disruptions; and (4) rapid changes in customer needs and expectations [6, 12–14]. In such a situation, achieving SCF is essential. Flexibility is essential for production and SC. The research recognizes its importance, and more studies identifying the causes and consequences of SCF have been conducted recently [12]. Many organizations have evolved as a result of customer-driven marketplaces, rapid innovation cycles, and faster technological improvements. Furthermore, because of fluctuating demand, many companies in cyclical industries must swiftly adjust their manufacturing output without incurring large expenses. Producers must therefore adapt more rapidly and effectively to changes in output volume and product diversity. Additionally, because entire supply chains—rather than just individual companies—are subject to competition, they need to become more adept at responding to changing consumer demands and delivering above-average added value, which makes production and supply chain flexibility crucial. Adaptable SC and increased production flexibility are thought to be the newest tools for gaining a clear competitive edge since they enable quicker and more affordable responses to changing client demands. Despite the fact that manufacturers and the industrial SC as a whole have only lately started to increase their flexibility, manufacturing and SCF have long been the focus of intense research in the scientific community [15].

The SC segment experiences a considerable number of advancements each year. Modern technologies are changing how business is done and upending the SC industry. Inefficiencies in the current systems highlight the enormous number of new technologies. To increase process efficiency and maintain their competitiveness or survival, businesses employ innovative technologies. A corporation may have a wide range of technologies at its disposal, making it challenging to select one without prior experience or understanding. Businesses that are reluctant to embrace new technologies fear that they will lag their rivals. Given that "innovation is seen as a means of gaining a competitive edge and stimulating economic growth," companies need to reconsider how they plan for these technological developments and start paying more attention to the process of adopting new ideas [16]. Businesses usually engage in new technology to strengthen their internal operations and relationships with business partners. This is done to accommodate growing competition, changing customer needs, regulatory changes, and technological breakthroughs. Coordinating closely and effectively is necessary to meet a range of customer requirements across the SC. Maintaining a high standard of quality and affordability while ensuring that the products are delivered on time and at the right location is also crucial. Effective cooperation and coordination of activities, however, are not enough for firms to thrive. To succeed in this highly uncertain business environment, companies need to be adaptable both inside and outside of their SC. It is believed that one of the most important success factors that contributes to a durable competitive advantage is flexibility [17].

In order to ascertain the impact of technology on SCF, the theories employed to analyze SCF, the sectors in which studies have been conducted, and the dimensions of SCF in the literature, the current study aims to evaluate the prior research on the relationship between SCF and technology. By conducting a comprehensive literature review, which will extensively examine the body of research on SCF and its connection to technology and business success, the study will address the research questions. The review will look at studies that look at how different technologies enable SCF in order to help with this understanding. The second research question will be addressed by identifying and categorizing SCF dimensions—such as volume flexibility, delivery flexibility, sourcing flexibility, and so forth—according to their significance in the literature. Additionally, investigate the pertinent theoretical frameworks used to evaluate SCF, such as the "Resource-Based View and Dynamic Capability theory," to gain understanding of how well the various theories explain how the mechanisms in SC foster flexibility. The paper is organized into four sections to accomplish this goal, beginning with this introduction in Sect. 1. The process is described in Sect. 2, the results are compiled in Sect. 3, and the conclusions and limitations are shown in Sect. 4.

2 Methodology

2.1 Literature Review

Systematic literature reviews are seen as important by researchers since they combine previous research on a particular area to provide scientific knowledge to support practice. Its primary advantage is that it follows a number of well-established phases, which makes it easy for other researchers to recreate or validate [18]. This study used this methodology to gather data from other articles that discuss SCF and technology at the same time. The Main Topics and Research Lines that offer the least amount of evidence for the simultaneous link between SCF and technology have been identified after an examination and classification of the relevant data from these studies. Important aspects in performing a systematic literature review include identifying the subject area and creating a process for locating, assessing, and synthesizing the literature. The five steps listed below are used in this paper's systematic literature review [18]:

1. Research questions formulation
2. Studies identification
3. Selection and the evaluation of studies
4. Conduct analysis and consolidation of studies
5. Presentation of the outcomes

2.1.1 Formulating Research Questions (RQ)

The 1st stage in doing a SLR is to determine the RQ that the research must address. Thus, the subsequent RQs were determined for the purpose of this investigation [18].

1. What impact does technology have on SCF and, in turn, firm performance?
2. Which SCF dimensions have been documented in the literature?
3. Which theories have been applied in the research of SCF?

2.1.2 Studies Identification

This step is looking for and locating pertinent research to address the research question. In this instance, Google Scholar and Scopus were the primary databases that were searched. SLR on SC management should use a range of bibliographic databases to remove publication bias [18]. Keyword preference based on the concepts of SCF, technology, and sustainability. The research keywords were supposed to appear in the "titles, abstracts, or keywords" of the publications, thus those filters were used for the document identification approach. Search keywords were generated based on previous bibliometric analyses to indicate prospective areas of research in the SCF domain. "Title" OR "Abstract" OR "Keyword" String 1—("Supply Chain Flexibility" & "Technology" OR "AI" OR "Analytics" OR "Innovation" OR "Digital" OR

"Sustainability") String 2—("SCF"). The major journals in the fields of SC& operations management provided pertinent information. The dates of publishing were not restricted, & searches were carried out from July 9, 2024, until July 31, 2024. Process found 449 papers in total.

2.1.3 Selection and Evaluation of Studies

After the first search process, the abstracts, methods, main findings, and conclusions of the chosen publications were examined to make sure they were pertinent to the RQs. When choosing the studies, the following inclusion criteria were used: peer-reviewed, English-language scientific publications that included news articles and literature reviews. Since impact-indexed journals guarantee that the selection process adheres to a set of standards that ensure the quality of primary research, papers published in these journals were taken into consideration as a factor in SLR. Additionally, papers that satisfied the study's requirements were selected using the following exclusion criteria: duplicate items and publications that have nothing to do with the study's topic (exclusions based on titles: 112, abstracts: 77) 260 papers were obtained through this procedure for in-depth examination. After considering the research to which we had full access for further analysis, 204 publications were considered (Fig. 1).

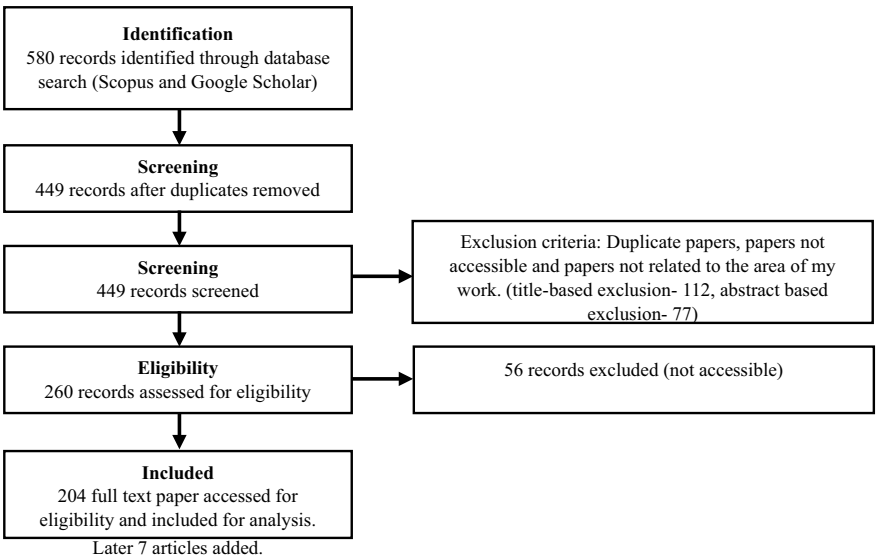


Fig. 1 Identification and selection of the papers

2.1.4 Analysis and Synthesis

The analysis began with 260 papers that were selected & evaluated on the basis of inclusion and exclusion criteria. 204 paper were selected for further analysis. The content of paper were coded to identify and synthesize what has been done in the literature so far with context to SCF and TI and what impact does technology has on SCF along with the concepts of sustainability in supply chains. The coding process involved categorizing the papers based on what work has been done in SCF, technology adopted, sector in which the work has been done, what theory has been used, the variables studied, the type of study, findings, limitations of this study and its future scope. An excel database was constructed to organize the coded information systematically which facilitated a clear overview of the research landscape concerning SCF.

2.1.5 Result Presentations

The required data was extracted from the literature to answer the RQs of the study. Figure 2 summarizes the technique for searching, selecting, and evaluating papers. The results are presented in tabular and graphical form. In this section, the study's results are presented and discussed.

3 Results and Discussion

The identified papers are first subjected to a descriptive analysis. Following this, the impact of technologies on SCF, dimensions reported in literature, theories studies, synthesis of studies conducted in India, and future study lines identified are discussed under the literature classification and analysis section.

3.1 Descriptive Results

Figure 3 depicts the time evolution of SCF research, demonstrating an evident increase in the number of articles throughout different years, indicating a considerable growth pattern. Early Years (1997–2006)—During the first ten years, just one paper was published per year: 1997, 1998, 1999, and 2003. From 2004 to 2006, the number of papers increased steadily but somewhat, with the greatest number of papers being 5 in 2006.

Moderate Growth (2007–2012)—An upward and uneven trend is noticed between 2007 and 2012, with a range of 9–10 papers per year. This period is characterized by moderate scientific activity, with a relatively consistent quantity of works.

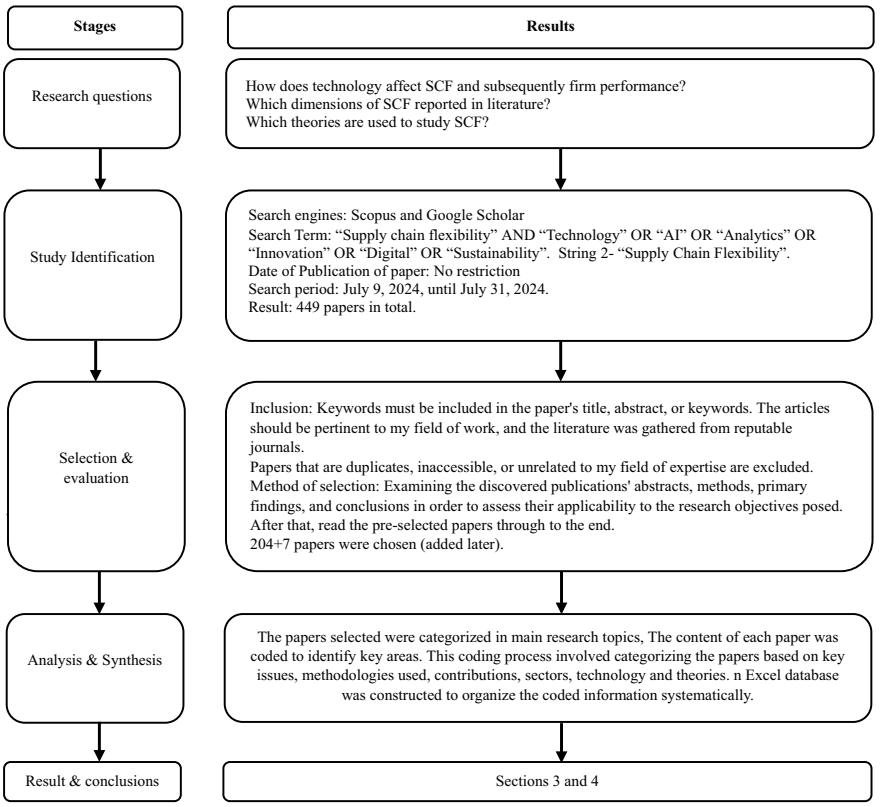


Fig. 2 Summary of research methodology

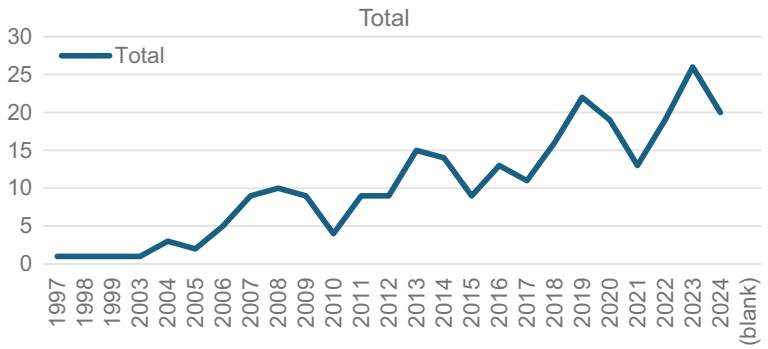


Fig. 3 Time evolution of research on SCF

Significant Growth (2013–2020): Thus, in 2013, the number of papers increased to 15, followed by 14 papers in 2014. Furthermore, the indicators stay strong from 2015 to 2020, with a maximum of 22 papers published in 2019. This time demonstrates that research production is increasing rapidly from year to year.

High Volume Period (2021–2024)—Beginning in 2019, the annual number of papers is rather high, ranging between 13 and 26 papers per year. Out of all the papers, 2023 has the most papers (26), while 2024 has 20. In conclusion. The data show that a steady stream of publications began in the late 1990s and early 2000s, and gradually increased in the mid-2000s. The trend analysis shows that there are no articles from 2009 to 2012, and that the number of papers increases from 2013 to the most current one, with most papers emerging between 2020 and 2023.

The distribution of different sectors and industries is represented in Fig. 4, in which manufacturing is the most common sector listed, with 51 instances indicating that this sector is the most represented in the data. Followed by the automotive sector with 12 instances, indicating its prominence in this dataset. Industries for example Automobile Industry with 6 instances, Pharmaceutical Industry with 4 instances, and SME again with 4 instances also have multiple entries, though significantly fewer than manufacturing sector. Sectors such as FMCG, Fashion Industry, and Textile Industry have 2 occurrences each as we can see in the graph. There are many sectors that have just 1 instance, such as Agriculture companies, Cement, Construction industry, event management, and wineries, indicating they are less represented in the corpus.

3.2 Country Specific Work

A summary of the research conducted in India over the past ten years is provided in Table 1, which also includes information on the technologies covered, the work done and the industry in which it was completed.

3.3 Impact of Technologies on SCF

Technology plays a crucial role in improving SCF, organizations can enhance their upstream SCF by simplifying integrated supplier process. This simplification often involves leveraging technology to remove redundancies and streamlining operations, that leads to improved responsiveness to market changes. Technology plays a key role in making SC processes simpler. Advanced technologies can cut out extra steps, making things less complex, and reduced costs. This simplification is key to make supply chains more flexible and gaining a competitive position in the market [20]. Table 2 presents a list of technologies that have been discussed in the literature and what impact these technologies have on SCF. This provides us a comprehensive view of how firms leverage different technologies to support or enhance SCF.

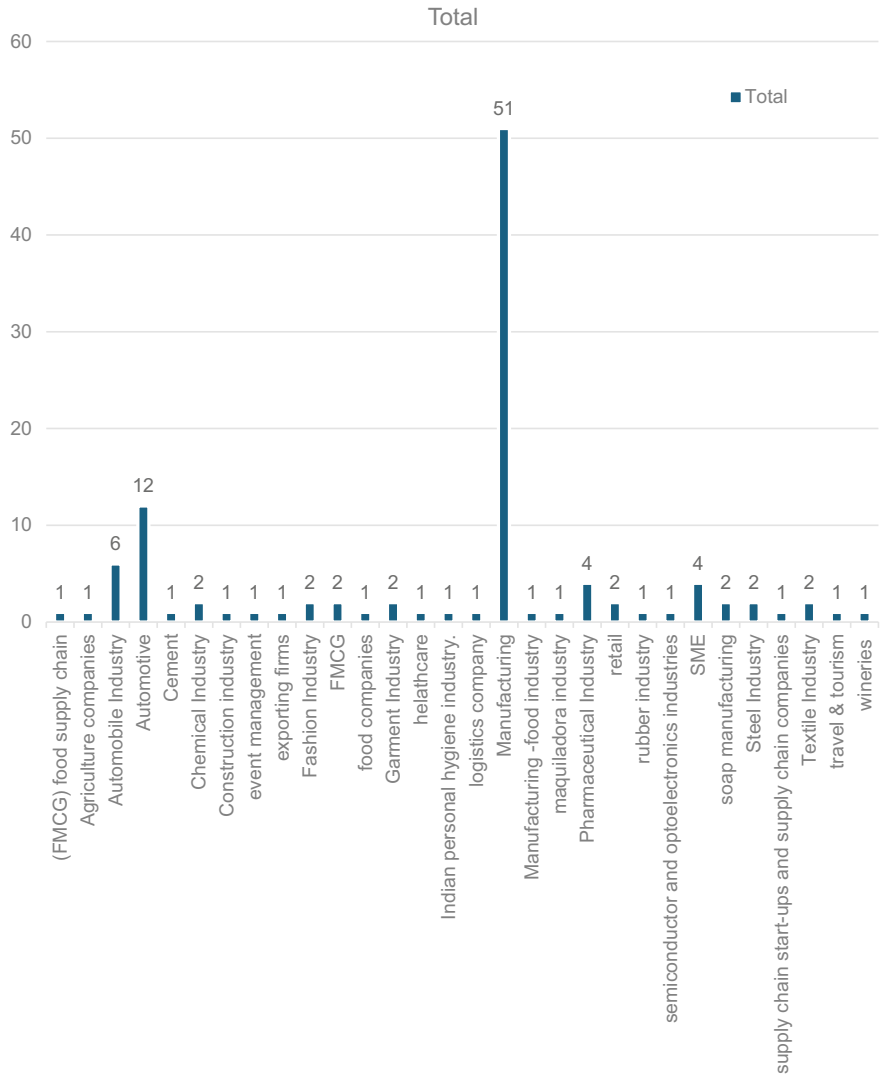


Fig. 4 The distribution of different sectors

3.4 SCF: Dimensions

Multiple dimensions of SCF have been studied in the literature; they are described in Table 3.

Table 1 Country specific work

Sl. No.	Work done and technology referred	Sector	References
1.	The study identifies, assesses, and prioritizes SC flexibilities while also investigating the relationship between them. The findings show that market flexibility has the strongest link, while new product development flexibility has the greatest influence	FMCG	[19]
2.	This study investigates how process simplification affects the relationship between upstream SCF and process integration in Indian automotive industry. The data show that top management commitment has a minor mediation effect on upstream SCF, but process simplification has a complete mediation influence on upstream SCF and supplier relationships. According to the study, in order to improve the flexibility of the upstream supply chain for Indian automotive management, firms should develop a strategy to streamline supply chain procedures through technological enhancements and supplier training	Automotive	[20]
3.	In order to address demand unpredictability, this research examines how flexible an automotive supply chain may be under sales promotional schemes (SPS)		[21]
4.	The study verifies the system dynamics model in the context of soap manufacture in India and uses a novel methodology to assess SCF. The results of the study indicate that the demand and trend for both packed and unpacked soaps vary with the season	Soap manufacturing	[22]

(continued)

Table 1 (continued)

Sl. No.	Work done and technology referred	Sector	References
5.	The purpose of this study is to evaluate the performance of an Indian soap manufacturer and calculate its SCF. In order to create a flexible SC environment, the authors attempted to determine the most crucial operational and strategic decision-making variables. This study offers a perspective on how to track performance by creating templates based on actual data and identifies crucial elements that support the successful use of flexibility principles in Indian soap manufacturing companies	Soap manufacturing	[23]
6.	Finding the causal links between several SCF elements that are significant to the Indian personal hygiene market is the goal of this study. Three main SCF aspects were considered, and the results show that product and physical distribution flexibility have a favorable impact on the SCF, however management flexibility has a negative effect overall	Indian personal hygiene industry	[24]
7.	The paper examines the barriers to flexibility in the Indian automobile SC from the standpoint of sustainability, which is an essential element of contemporary supply chains. Twenty impediments have been identified based on the results of the descriptive survey and the literature review	Automobile Industry	[25]
8.	This study examines how two key SC management practices—SC strategy (SCS) and SCF—affect three SC performance (SCP) variables and five success factors (SFS) for SCF, respectively, in order to properly apply SCM approaches in the automobile industry	Automobile Industry	[26]

(continued)

Table 1 (continued)

Sl. No.	Work done and technology referred	Sector	References
9.	The purpose of this article is to identify the essential components that facilitate the use of SCF and assess how well each component fits within the Indian personal hygiene market. The findings indicate that the SCF is influenced by seven dimensions: "process, production, sourcing, volume, market, new product development, and coordination."	FMCG	[1]
10.	Using case study methods, the paper creates a "fuzzy causal model" of business flexibility traits in the Indian pharmaceutical sector. A survey of the literature was used to identify the eight flexibility dimensions. Both technical and information technology flexibility have been mentioned in this study. It has been found that the flexibility of marketing, technology, supply chain, and information technology both contributes to the overall flexibility of the company and causes and influences other types of flexibility	Pharmaceutical Industry	[27]

3.5 Theories that Are Used to Study SCF

Several theories have been discussed in literature to study SCF, they are described in Table 4.

3.6 Supply Chain Sustainability (SCS)

A SC is sustainable when all stakeholders' "economic, environmental, and social" benefits are recognized throughout time. Flexibility can boost sustainability in turbulent environments. Improving supply chain flexibility is beneficial in SC management because it increases the supply chain's anti-risk capability, promoting continuous operation. A study investigates how agility and adaptation affect the "economic, social, and environmental" sustainability of Australia's industrial supply chain. The findings show that SC flexibility has a considerable positive effect on social and environmental sustainability. Another study used lean and elastic approaches in SC to examine their effects on the 3 dimensions of sustainability performance. All impact outcomes were favorable [51].

Table 2 Technologies impacting SCF & dimensions of SCF

Sl. No.	Technology	How Technology Impacts SCF	SCF Dimensions	References
1.	Information Technology	The survey indicates that the use of IT applications that allow for rapid modifications in supply, production, and distribution processes significantly promotes adaptability. It suggests that enhancing SCF requires a company's own capacity to develop and produce new technologies; this capacity allows businesses to gather, integrate, and implement IT resources that support responsive business operations. Businesses can swiftly modify their production levels and logistics thanks to technology, which is essential in a market that is dynamic and unstable	Ability of the firm to modify the quantity of supplier's order, delivery time span of supplier's order, production volume capacity and to reduce development cycle times	[28]
2.	Big data analytics	The paper emphasizes how using data analytics will increase flexibility by enabling companies to react to shifts in supply and demand more quickly. Digital SC platform integration enhances coordination and communication amongst SC partners, which is essential for adapting to changes in the market and boosting the overall performance of the company	Process flexibility allows for a reduction in the expense/ price of green products Flexible sourcing allows for the introduction of new alternative recycling resources The capability to reduce waste produced by the SC through volume flexibility Capability to speed the acquisition of environmental knowledge and adapt to market changes	[29]

(continued)

Table 2 (continued)

Sl. No.	Technology	How Technology Impacts SCF	SCF Dimensions	References
3.	Blockchain, AI & IOT	In order to help a business manage its SC more effectively and efficiently, the study examines how digital technology enhances its information system capabilities. It discusses how blockchain, artificial intelligence, and the Internet of Things might enhance openness, accessibility, real-time data sharing, effectiveness, and flexibility	An information data platform system can be used by enterprises to monitor and notify them to SC operational concerns Using digital technology, enterprises can adapt their SC structures in response to changes in client demand and market trends Enterprises might modify production and manufacturing processes in response to market-wide technology advancements Digital technology enables businesses to rapidly change their daily manufacturing processes, output and inventory levels, and distribution channels SC's member firms can develop important and creative knowledge through information exchange	[30]

(continued)

Table 2 (continued)

Sl. No.	Technology	How Technology Impacts SCF	SCF Dimensions	References
4.	Blockchain	The article explains how blockchain enables companies to take charge of their operations while protecting sensitive data, which improves coordination and responsiveness. It also explains how integrating partners into an organization's information system increases responsiveness and flexibility. Competitiveness is increased when a company's supply chain, which includes distributors, retailers, suppliers, transporters, and IT service providers, is flexible. This makes it possible for businesses to handle tasks like order fulfillment, distribution, and payment while identifying hidden pricing concerns and optimizing the value of the chain's operations. Scalability and adaptability are greatly impacted by blockchain, giving businesses a competitive edge	BCT enhances the capability to cooperatively manage intra- and inter-organizational processes in order to create effective and efficient flows of products, services, information, money, and decisions BCT's deep integration ability transforms the SC into a team rather than fragmented parties, improving performance in terms of cost, time, and quality	[16]
5.	SC data analytics with artificial intelligence (SCDA-AI)	The paper explains how to successfully handle uncertainties by adjusting the degree of SCF using SCD-AI. Additionally, it can enhance SCF, leading to improved performance. The development of SCD-AI's dynamic capabilities is necessary to satisfy the supply chain's demands for responsiveness and flexibility	SC can adjust production facilities, procedures, and operations SC may rationalize transportation and distribution management using information systems SC has the ability to alter its delivery lead times SC can change the size of orders	[31]

Table 3 Dimensions of SCF

Sl. No.	Dimension	Explanation	References
1.	Product flexibility	It's been described as the firm's capability to develop numerous items in order to meet diverse client wants	[24, 32]
2.	Volume flexibility	It's been described as the firm's ability to adjust the output quantity depending on market demand	[33, 22]
3.	New product development flexibility	It has been described as the firm's capacity to offer new products with contemporary characteristics	[33, 34]
4.	Sourcing flexibility	It has been described as the firm's capability to retain different suppliers for the same product	[35, 36]
5.	Physical distribution flexibility	It has been described as the firm's capability to hold numerous distribution channels to increase responsiveness	[37, 38]
6.	Demand management flexibility	It's been described as the firm's ability to come up with specific client requirements and manufacture and deliver to boost SC responsiveness	[37]
7.	Coordination flexibility	It has been described as the firm's capability to create supplier/vendor/customer relationship management procedures for an effective and well-integrated supply chain structure	[37]
8.	Manufacturing flexibility	Capacity to adjust production and manufacturing system responsiveness to meet market demands	[39]
9.	Respond to market flexibility	It has been described as the firm's ability to increase SC reactivity to cater to the shifting market needs	[40]
10.	Process flexibility	It's been described as the firm's capacity to manufacture a variety of product kinds utilizing an identical technique	[33, 35]
11.	Expansion flexibility	It has been described as the firm's ability to magnify capacity through horizontal or vertical integration	[3]
12.	Strategic flexibility	Ability of a corporation to act, respond, reposition swiftly, or adapt to highly variable market environmental situations, facilitated by its resources and competencies, in order to sustain its competitive edge	[41]
13.	Financial flexibility	Capability to decide and modify finance, assets, and financial instruments	[27]
14.	Operational flexibility	It is the capability of the firm that allows for effective production, a timely response to shifting customer needs, and the ability to quickly adjust to market fluctuations in order to maintain the company's flexibility and presence	Sharma (2014)

(continued)

Table 3 (continued)

Sl. No.	Dimension	Explanation	References
15.	Technical flexibility	The capability of businesses to choose and incorporate multiple technological systems to aid business operations and achieve organizational objectives	Sharma (2014)
16.	Organizational flexibility	Capability of learning quickly, preparation for change, and swift adaptation that helps firms thrive in volatile environments Ability to match labor force abilities to supply chain demands in order to meet customer service/demand criteria	[27, 42]
17.	Logistics Flexibility	Ability to cost-effectively accept and deliver products as suppliers and consumers change (customer location changes, globalization, delay)	[42]
18.	Resource Flexibility	The capacity to purchase goods and services in accordance to shifting needs and the availability of these resources	[43]
19.	Information Systems Flexibility	The process of modifying an organization's information systems to reflect shifting market dynamics	[43]
20.	Offering flexibility	Offering flexibility is the capacity of a supply chain connection to accommodate modifications in the offerings of goods or services created in collaboration with present partners in reaction to shifts in the commercial climate	[44]
21.	Labor Flexibility	The flexibility to adjust labor force size, job duties, and other worker obligations is known as labor flexibility	[45]
22.	Machine Flexibility	The ease with which a machine may switch between processing multiple jobs is known as machine flexibility	[45]
23.	Routing Flexibility	The ability of production cells to manage machine breakdown and complete assigned tasks is known as routing flexibility	[45]
24.	Supplier Flexibility	Supplier flexibility refers to the capacity of the business to react to and plan for changes that take place	[46]
25.	Responsive flexibility	The ability of supply chain partners to adapt swiftly and affordably to changes in the market in order to meet client demand	[40]
26.	Lead-time Flexibility	It is the capacity of a production or SC system to modify the predetermined lead time of associated tasks, procedures, or services that they offer	[47]

Table 4 Theories used to study SCF

Sl. No.	Theory	What work is done in context of SCF dimensions	References
1.	Contingency theory	An integrative framework that links different SCF aspects, examines the influence of SC complexity, and offers empirical support for the suggested linkages has been developed using contingency theory	[48]
2.	Coordination theory	In order to determine how SC partners can efficiently coordinate their actions, coordination theory has been applied. It offers a thorough grasp of how improved flexibility can result from efficient coordination among SC partners, allowing businesses to react quickly to changes and interruptions in the business environment	[44]
3.	Dynamic capability view	The dynamic capability view has been applied extensively in scholarly works and in a variety of settings. It has been used to examine how businesses handle product diversity and improve the performance of their supply chains in the ever-changing business landscape, boosting supply chain agility in service-oriented companies and exploring the connections between big data analytics, circular economy (CE) practices, sustainable supply chain management, and sustainable performance in manufacturing companies. Additionally, the theory has been used to examine the components of SC dynamic capabilities with the goal of identifying linkages between them and suggesting a method for formation. In a different context, the theory has also been used to examine the use of SC analytics and its implications for achieving a CE	[49, 49–52]

(continued)

Table 4 (continued)

Sl. No.	Theory	What work is done in context of SCF dimensions	References
4.	Resource based view theory (RBV)	Academic literature has extensively examined RBV utility in relation to the SCF. It has been used to examine the connections between SC capabilities and environmental management strategies in the automotive industry, the relationship between external SCF and SME product innovation performance, and the connections between blockchain technology, information sharing, SCF, and financial institutions' competitive advantage. It provides a fundamental framework for comprehending how financial institutions might use their assets, such information sharing and blockchain technology, to obtain a competitive edge. Furthermore, the study of BCT on SC capabilities and flexibility, how incorporating blockchain technology into information systems can improve SC performance and efficiency, and how IT can be a crucial resource that helps SC management become more competitive have all been examined. This emphasizes the significance of using IT as a strategic asset for operational success	[53, 54, 53–55]
5.	Resource-based and strategic behavior theory	Using RBV and strategic behavior theory, the study investigates the relationship between the flexibility of the SC and the SMS alliance orientation, as well as the flexibility and entrepreneurial orientation of SME entrepreneurs	[56]
6.	Resource orchestration theory	The theory of resource orchestration has been used to investigate how SMEs might improve their performance by managing their IT resources efficiently. According to the notion, a firm's capacity to successfully acquire and deploy these strategic resources has a significant impact on its performance. It has also been used to examine how the ISO 9001 standard, SCF, and ambidextrous SC approach interact	[28, 57]

(continued)

Table 4 (continued)

Sl. No.	Theory	What work is done in context of SCF dimensions	References
7.	Network theory	Building strong, trusting relationships with network partners is crucial, and network theory is utilized to analyze the relationships and interactions among enterprises inside the SC. According to the report, blockchain technology may support these connections by advancing information security and transparency, which is actually a key element in fostering confidence in commercial dealings. Furthermore, RBV theory has been used in the context of blockchain technology, which is seen as a digital resource that can give the company a competitive advantage	[58]
8.	Information processing theory (IPT)	As a critical lens through which the authors examine the connection between information processing skills, commerce, sourcing flexibility, and performance outcomes, this theory has been used as a foundational framework to investigate sourcing flexibility within supply chains	[59]

4 Conclusion

The systematic literature review of articles underlines the leading contribution of various geographical regions along with the sectors in which the work has been done with respect to SCF. The study reveals the evolving focus on TI within SCF that includes information technology and various other industry 4.0 technologies that supports SCF and multiple other operations. The study highlights various SCF dimensions that have been addressed in the literature so far, along with how they have been used to measure SCF when supported by technology. Furthermore, the study reveals various theories that have been employed to study SCF and how they have been combined with other theories to study the related aspects. A detailed description can be seen in Table 2 of the work been done in India along with the technologies been integrated in the process. Our findings highlight current advancements in SCF, identifying technologies adopted and SCF dimensions used to measure flexibility highlighting the key sectors in which the work has been done.

The study's managerial implications indicate that companies should use cutting-edge technology to concentrate more on SCF. The most important element in fostering flexibility along supply chains and enabling managers to react more quickly to changes and disturbances in the market is technology. Managers will be able to optimize multiple decision-making procedures, the real-time viewability of data streams, and SC resilience with IoT, blockchain, and BDA. In order to create supply chains that are more responsive and flexible, the study also highlights the significance of implementing some key theoretical frameworks, such as Resource-Based View (RBV) and Dynamic Capability theory. For companies in India, where there is less emphasis on SCF-related technology integration, this might be seen as a crucial time to close gaps, spur innovation, and gain a competitive edge in quickly evolving international markets.

Systematic literature review is a credible process that is supported by a vast number of research papers. However, it is dependent on the availability and accessibility of research studies, as well as the researcher's requirements. Inclusion and exclusion of studies based on specific criteria might lead to publication bias, which is a limitation of SLR. This study provides valuable insights for future exploration in terms of TI, SCS, SCF and firm performance. Future research should examine additional flexibility dimensions, objective performance measures, and other integration constructs like relationship, measurement, and technological integration, according to (Lee et al. 2013). Examine how digital technology may enhance SC performance and give businesses a competitive edge by enhancing SC skills in the Indian setting, as has been done in Bangladesh [30]. Future studies can examine how SCF and sustainability relate to one another within a specific industry, facilitating the concurrent incorporation of both elements into supply chains [25]. As suggested by [51], Future research can look at how digital capacity and flexibility impact SCS. We can infer from the gaps mentioned above that measuring the effect of TI on company performance in a specific sector when mediated by SCF can be a precise field of investigation. The influence of TI on business performance and the moderating effect of environmental

dynamism (ED) can be measured by adding another variable, such as SCS, or by studying environmental uncertainty (EU) in the model. Since little work has been done in the pharmaceutical industry regarding SCF, this industry can be the focus of future research efforts.

References

1. Singh, R.K., Joshi, S., Sharma, M.: Modelling supply chain flexibility in the Indian personal hygiene industry: an ISM-Fuzzy MICMAC approach. Glob. Bus. Rev. **24**(5), 1058–1075 (2023). https://doi.org/10.1177/0972150920923075
2. Singh, R.K., Acharya, P.: Supply chain flexibility: a frame work of research dimensions. Glob. J. Flex. Syst. Manag. **14**(3), 157–166 (2013). https://doi.org/10.1007/s40171-013-0039-4
3. Stevenson, M., Spring, M.: Flexibility from a supply chain perspective: definition and review. Int. J. Oper. Prod. Manage. **27**(7), 685–713 (2007). https://doi.org/10.1108/01443570710756956
4. Roh, J., Hong, P., Min, H.: Implementation of a responsive supply chain strategy in global complexity: the case of manufacturing firms. Int. J. Prod. Econ. **147**(Part B), 198–210 (2014). https://doi.org/10.1016/j.ijpe.2013.04.013
5. Rojo, A., Stevenson, M., Lloréns Montes, F.J., Perez-Arostegui, M.N.: Supply chain flexibility in dynamic environments: the enabling role of operational absorptive capacity and organisational learning. Int. J. Oper. Prod. Manag. **38**(3), 636–666 (2018). https://doi.org/10.1108/IJOPM-08-2016-0450
6. Duclos, L.K., Vokurka, R.J., Lummus, R.R.: A conceptual model of supply chain flexibility. Ind. Manage. Data Syst. **103**(5–6), 446–456 (2003) (Emerald Group Publishing Ltd). https://doi.org/10.1108/02635570310480015
7. Slack, N.: The flexibility of manufacturing systems. Int. J. Oper. Prod. Manag. **25**(12), 1190–1200 (2005). https://doi.org/10.1108/01443570510633594
8. Yang, J., Zhang, F., Jiang, X., Sun, W.: Strategic flexibility, green management, and firm competitiveness in an emerging economy. Technol. Forecast. Soc. Chang. **101**, 347–356 (2015). https://doi.org/10.1016/j.techfore.2015.09.016
9. Huo, B., Gu, M., Wang, Z.: Supply chain flexibility concepts, dimensions and outcomes: an organisational capability perspective. Int. J. Prod. Res. **56**(17), 5883–5903 (2018). https://doi.org/10.1080/00207543.2018.1456694
10. Manders, J.H.M., Caniëls, M.C.J., Ghijsen, P.W.T.: Exploring supply chain flexibility in a FMCG food supply chain. J. Purch. Supply Manage. **22**(3), 181–195 (2016). https://doi.org/10.1016/j.pursup.2016.06.001
11. Sreedevi, R., Saranga, H.: Uncertainty and supply chain risk: the moderating role of supply chain flexibility in risk mitigation. Int. J. Prod. Econ. **193**, 332–342 (2017). https://doi.org/10.1016/j.ijpe.2017.07.024
12. Manders, J.H.M., Caniëls, M.C.J., Ghijsen, P.W.T.: Supply chain flexibility: a systematic literature review & identification of directions for future research. Int. J. Logist. Manag. **28**(4), 964–1026 (2017). https://doi.org/10.1108/IJLM-07-2016-0176
13. Pujawan, I.N.: Assessing supply chain flexibility: a conceptual framework and case study. Int. J. Integrated Supply Manage. **1**(1) (2004)
14. Tachizawa, E.M., Thomsen, C.G.: Drivers and sources of supply flexibility: an exploratory study. Int. J. Oper. Prod. Manag. **27**(10), 1115–1136 (2007). https://doi.org/10.1108/01443570710820657
15. Seebacher, G., Winkler, H.: A citation analysis of the research on manufacturing and supply chain flexibility. Int. J. Prod. Res. **51**(11), 3415–3427 (2013). https://doi.org/10.1080/00207543.2013.774483

16. Potnis, T., Lau, Y.Y., Yip, T.L.: Roles of blockchain technology in supply chain capability and flexibility. Sustainability (Switzerland) **15**(9) (2023). https://doi.org/10.3390/su15097460
17. Kopanaki, E., Karvela, P., Georgopoulos, N.: From traditional interorganisational systems to cloud-based solutions: the impact on supply chain flexibility. J. Organ. Comput. Electron. Commer. **28**(4), 334–353 (2018). https://doi.org/10.1080/10919392.2018.1517480
18. Novais, L.R., Maqueira, J.M., Bruque, S.: Supply chain flexibility and mass personalization: a systematic literature review. J. Bus. Ind. Mark. **34**(8), 1791–1812 (2019) (Emerald Group Holdings Ltd). https://doi.org/10.1108/JBIM-03-2019-0105
19. Singh, R.K., Acharya, P.: Identification and evaluation of supply chain flexibilities in Indian FMCG sector using DEMATEL. Glob. J. Flex. Syst. Manag. **15**(2), 91–100 (2014). https://doi.org/10.1007/s40171-013-0050-9
20. Goyal, G., Samalia, H.V., Verma, P.: Mediating role of process simplification in process integration and upstream supply chain flexibility. Int. J. Product. Perform. Manage. **67**(5), 825–844 (2018). https://doi.org/10.1108/IJPPM-08-2016-0159
21. Chirra, S., Kumar, D.: Analysis of SCF under sales promotional schemes: an application of interpretive structural modelling approach. Int. J. Prod. Res. **56**(18), 6015–6033 (2018). https://doi.org/10.1080/00207543.2018.1463474
22. Singh, R.K., Modgil, S., Acharya, P.: Assessment of supply chain flexibility using system dynamics modeling. Glob. J. Flex. Syst. Manag. **20**, 39–63 (2019). https://doi.org/10.1007/s40171-019-00224-7
23. Singh, R.K., Acharya, P., Modgil, S.: A template-based approach to measure supply chain flexibility: a case study of Indian soap manufacturing firm. Meas. Bus. Excell. **24**(2), 161–181 (2020). https://doi.org/10.1108/MBE-10-2018-0080
24. Singh, R.K., Modgil, S., Acharya, P.: Identification and causal assessment of supply chain flexibility. Benchmarking **27**(2), 517–549 (2020). https://doi.org/10.1108/BIJ-01-2019-0003
25. Chirra, S., Raut, R.D., Kumar, D.: Barriers to sustainable supply chain flexibility during sales promotions. Int. J. Prod. Res. **59**(22), 6975–6993 (2021). https://doi.org/10.1080/00207543.2020.1832272
26. Chandak, A., Chandak, S., Dalpati, A.: Evaluation of impact of success factors of supply chain strategy and flexibility on supply chain performance. Pakistan J. Statistics Oper. Res. **17**(1), 185–194 (2021). https://doi.org/10.18187/pjsor.v17i1.2874
27. Akhtar, M., Asim, M.: Fuzzy causal model of enterprise flexibility dimensions using Fermatean Fuzzy DEMATEL: a case study of Indian pharmaceutical industry. Benchmarking (2024). https://doi.org/10.1108/BIJ-07-2023-0462
28. Benzidia, S., Makaoui, N.: Improving SMEs performance through supply chain flexibility and market agility: IT orchestration perspective. Supply Chain Forum **21**(3), 173–184 (2020). https://doi.org/10.1080/16258312.2020.1801108
29. Shafique, M.N., Rashid, A., Yeo, S.F., Adeel, U.: Transforming supply chains: powering circular economy with analytics, integration and flexibility using dual theory and deep learning with PLS-SEM-ANN analysis. Sustainability (Switzerland) **15**(15) (2023). https://doi.org/10.3390/su151511979
30. Shahadat, M.M.H., Chowdhury, A.H.M.Y., Nathan, R.J., Fekete-Farkas, M.: Digital technologies for firms' competitive advantage and improved supply chain performance. J. Risk Financial Manage. **16**(2) (2023). https://doi.org/10.3390/jrfm16020094
31. Barhmi, A., Laghzaoui, S., Slamti, F., Rouijel, M.R.: Digital learning, big data analytics and mechanisms for stabilizing and improving supply chain performance. Int. J. Inf. Syst. Project Manage. **12**(2), 30–47 (2024). https://doi.org/10.12821/ijispm120202
32. Um, J.: Improving supply chain flexibility and agility through variety management. Int. J. Logist. Manage. **28**(2), 464–487 (2017). https://doi.org/10.1108/IJLM-07-2015-0113
33. Fredriksson, A., Wänström, C.: Manufacturing and supply chain flexibility—towards a tool to analyse production network coordination at operational level. Strategic Outsourcing **7**(2), 173–194 (2014). https://doi.org/10.1108/SO-04-2014-0003
34. Vickery, S., Calantone, R.: Supply chain flexibility—an empirical study. J Supply Chain Manage. (1999)https://doi.org/10.1111/j.1745-493X.1999.tb00058.x

35. Sánchez, A.M., Pérez, M.P.: Supply chain flexibility and firm performance: a conceptual model and empirical study in the automotive industry. Int. J. Oper. Prod. Manag. **25**(7), 681–700 (2005). https://doi.org/10.1108/01443570510605090
36. Yi, C.Y., Ngai, E.W.T., Moon, K.L.: Supply chain flexibility in an uncertain environment: exploratory findings from five case studies. Supply Chain Manag. **16**(4), 271–283 (2011). https://doi.org/10.1108/13598541111139080
37. Yu, K., Cadeaux, J., Song, H.: Alternative forms of fit in distribution flexibility strategies. Int. J. Oper. Prod. Manag. **32**(10), 1199–1227 (2012). https://doi.org/10.1108/01443571211274521
38. Zhang, Q., Vonderembse, M.A., Lim, J.S.: Logistics flexibility and its impact on customer satisfaction. Int J Logistics Manage. **16**(1), 71–95 (2005). https://doi.org/10.1108/09574090510617367
39. Kumar, P., Shankar, R., Yadav, S.S.: Flexibility in global supply chain: a review of perspectives. In: Proceedings of Flexibility in Global Supply Chain: A Review of Perspectives (2007). https://www.researchgate.net/publication/283060677
40. Kumar, V., Fantazy, K.A., Kumar, U., Boyle, T.A.: Implementation and management framework for supply chain flexibility. J. Enterp. Inf. Manag. **19**(3), 303–319 (2006). https://doi.org/10.1108/17410390610658487
41. Singh, D., Oberoi, J.S., Ahuja, I.S.: An empirical investigation of dynamic capabilities in managing strategic flexibility in manufacturing organizations. Manag. Decis. **51**(7), 1442–1461 (2013). https://doi.org/10.1108/MD-05-2012-0332
42. Coronado, M.A.E., Lyons, A.C.: Evaluating operations flexibility in industrial supply chains to support build-to-order initiatives. Bus. Process Manage. J. **13**(4), 572–587 (2007). https://doi.org/10.1108/14637150710763586
43. Khalayleh, M.A., Bader, D., Aityassine, F.L.Y., Mohammad, A., Al-Azzam, M.K.A., Al-Awamleh, H.K., Mohammad, A.A.S.: The effect of digitalism on supply chain flexibility of food industry in Jordan. Uncertain Supply Chain Manage. **10**(4), 1549–1560 (2022). https://doi.org/10.5267/j.uscm.2022.6.001
44. Gosain, S., Malhotra, A., El Sawy, O.A.: Coordinating for flexibility in e-business supply chains. J. Manag. Inf. Syst. **21**(3), 7–45 (2004). https://doi.org/10.1080/07421222.2004.11045816
45. Gong, Z.: An economic evaluation model of supply chain flexibility. Eur. J. Oper. Res. **184**(2), 745–758 (2008). https://doi.org/10.1016/j.ejor.2006.11.013
46. Putri, R.L.S., Tarigan, Z.J.H., Siagian, H.: The effect of integrated information technology on competitive advantage through supply chain integration and supply chain flexibility. Uncertain Supply Chain Manage. **12**(3), 1841–1854 (2024). https://doi.org/10.5267/j.uscm.2024.2.018
47. Subash Babu, A.: Analysis of the dynamics between supply chain flexibility and key management ratios Dileep More. Int. J. Bus. Innov. Res. **3**(2) (2009)
48. Liao, Y.: An integrative framework of supply chain flexibility. Int. J. Product. Perform. Manage. **69**(6), 1321–1342 (2020). https://doi.org/10.1108/IJPPM-07-2019-0359
49. Ahmed, W., Najmi, A., Mustafa, Y., Khan, A.: Developing model to analyze factors affecting firms' agility and competitive capability: a case of a volatile market. J. Model. Manag. **14**(2), 476–491 (2019). https://doi.org/10.1108/JM2-07-2018-0092
50. Um, J., Lyons, A., Lam, H.K.S., Cheng, T.C.E., Dominguez-Pery, C.: Product variety management and supply chain performance: a capability perspective on their relationships and competitiveness implications. Int. J. Prod. Econ. **187**, 15–26 (2017). https://doi.org/10.1016/j.ijpe.2017.02.005
51. Zhou, Q., Wang, S.: Study on the relations of supply chain digitization, flexibility and sustainable development—a moderated multiple mediation model. Sustainability (Switzerland) **13**(18) (2021). https://doi.org/10.3390/su131810043
52. Zhu, C.Y., Zhang, D.L.: An empirical study on the mechanism of dynamic capacity formation in the supply chain. Sustainability (Switzerland) **14**(22) (2022). https://doi.org/10.3390/su142215044
53. Elrefae, G., Nuseir, M.T.: Blockchain in global finance make-over: exploring the mediating role of supply chain flexibility. Uncertain Supply Chain Manage. **10**(3), 983–992 (2022). https://doi.org/10.5267/j.uscm.2022.2.015

54. Ko, W.W.J., Liu, G., Ngugi, I.K., Chapleo, C.: External supply chain flexibility and product innovation performance: a study of small- and medium-sized UK-based manufacturers. Eur. J. Mark. **52**(9–10), 1981–2004 (2018). https://doi.org/10.1108/EJM-07-2017-0466
55. Liu, Y., Srai, J.S., Evans, S.: Environmental management: the role of supply chain capabilities in the auto sector. Supply Chain Manag. **21**(1), 1–19 (2016). https://doi.org/10.1108/SCM-01-2015-0026
56. Tokman, M., Richey, R.G., Morgan, T.R., Marino, L., Dickson, P.H.: SME supply chain portfolios: firm satisfaction and organization resources. Int. J. Logist. Manag. **24**(2), 271–300 (2013). https://doi.org/10.1108/IJLM-09-2012-0099
57. Rojo-Gallego-Burin, A., Llorens-Montes, F.J., Perez-Arostegui, M.N., Stevenson, M.: Ambidextrous supply chain strategy and supply chain flexibility: the contingent effect of ISO 9001. Ind. Manag. Data Syst. **120**(9), 1691–1714 (2020). https://doi.org/10.1108/IMDS-01-2020-0038
58. Wang, M., Yang, Y.: An empirical analysis of the supply chain flexibility using blockchain technology. Frontiers Psychol **13** (2022). https://doi.org/10.3389/fpsyg.2022.1004007
59. Wagner, S.M., Grosse-Ruyken, P.T., Erhun, F.: Determinants of sourcing flexibility and its impact on performance. Int. J. Prod. Econ. **205**, 329–341 (2018). https://doi.org/10.1016/j.ijpe.2018.08.006
60. Stohr, E.A.: (n.d.). The Flexible Enterprise. http://www.springer.com/series/10780

A Conceptual Framework to Evaluate Customers' Response to Use Product Recommendation Systems



Aman Mathan, Deepak Verma, and Divesh Kumar

Abstract E-retail platforms have been increasingly moving towards the integration of product recommendations. About one-third of business on these retailers comes from these recommendations. Meanwhile, the effectiveness of these recommendations is peculiarly measured by the accuracy of the recommended product in transactions. This research tried to reach the behaviour aspect of users through an in-depth review of the literature pertaining to the Product Recommendations System (PRS). This research found that PRS significantly influences Customer Perceived Value (CPV) and Customer Engagement (CE). The relevant and personalised product recommendations from PRS make the customer perceive greater value and increase satisfaction and loyalty. In contrast, the customers feel more engaged with the personalised recommendations, which leads to the willingness to engage with the website. Also, customers who feel engaged and valued are more likely to provide feedback and information that can be used to refine further and improve the recommendation algorithms. The influence of product recommendations on CPV and CE underscores the importance of personalised marketing and customer-centric approaches to business. The researchers failed to find any existing research connecting these variables; hence, this research proposes a conceptual model measuring the influence of PRS on CPV and CE. Additionally, this research found that the literature is ambiguous concerning the nature of the relationship between CPV and CE.

Keywords Product recommendation systems · Information systems · Relationship marketing · Task technology fit · Customer engagement · Customer perceived value

A. Mathan (✉) · D. Verma · D. Kumar
Malaviya National Institute of Technology Jaipur, Jaipur, India
e-mail: 2019rbm9501@mnit.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_15

1 Introduction

In the past few decades, a major shift in markets from brick-and-mortar models to online models is evident. About 8% of total Indian retail is expected from digital mediums by 2025. The Internet has changed the scenarios of business operations and, hence, the consumers' shopping behaviour due to its wide reachability and cost-effectiveness, making it the most convenient avenue for buying and selling products and services through e-commerce [1]. E-commerce has increased the number of options available for a customer to choose from, which makes the customers more empowered. However, this increased number of options and information may also be hectic and confusing and reduce the consumers' satisfaction level. In "cost of thinking" a concept describes the cost involved with the decision-making process whenever an individual chooses an alternative from the available options. This cost includes the time an individual may spend in identifying and evaluating the alternative options and their attributes, the limitation of the consumer's knowledge, and the overall efforts exerted by the consumer in the selection process [2].

In decision-making, individuals collect and evaluate information from both internal as well as external sources before they make the final purchase decision [3]. To ease the decision-making process, individuals may seek inputs, i.e. Word of Mouth (from friends and relatives) and expert advice. This ensures confidence and satisfaction about their choices.

Recommendations have existed for a long time, not only in the retailing of products and services but also in various other disciplines. Recommendations are simply the external source of information that eases the individual's decision-making. Traditionally, personal sources (WOM, friends, and relatives) and impersonal sources (expert advice) have been very popular recommendation sources [4]. Recommendations have been a commonly used terminology and aid in the lives of individuals. One comes across various choices in day-to-day activities while purchasing or using a product or service.

In the digital environment, one cannot ignore the internet as one of the sources of these impersonal recommendations, which could be both personalised and non-personalized. These recommendations are generated by automated users' data-backed systems, popularly known as Product Recommendation Systems (PRS). Products are featured on designated pages, and these are based on embedded recommendation systems, which explicitly showcase the user with relevant products that match the user's attributes [5]. E-commerce, compared to traditional commerce, differs in terms of experiences; the human touch is absent in online platforms. Hence, the human–machine and human–human interactions create a huge difference in the user experience [6]. Recommendation systems improve the overall experience of individuals in the way they interact with online platforms [7]. But obviously, not all individuals may have similar attributes. Hence, it's not an easy task to manage and create a unique experience.

By definition, PRS are software programs designed to gather customers' information regarding their online behaviour, preferences and interests, using customers'

demographics, search and purchase behaviours to make personalised recommendations to the customers. PRS have a variety of uses suggesting products, movies, music, books, online courses, food items, locations to travel, restaurants, hotels, nearby shops, financial investments, real estate properties, doctors, jobs, education institutions, etc.

PRS are aids to customers in decision-making processes. These systems help customers to find the desired product or service with less effort, and sometimes even better than the individual's perceived products [8, 9].

Prior literature in the area has been focused on the technical aspect of the PRS; they concentrated specifically on generating recommendations. Algorithms are an important part of any PRS. However, there is a scarcity of research on the psychological aspects of the consumer's interaction with the system and their behavioural response [10]. While several authors have conducted researches on the adoption of RS [10–17] but compared to information systems, RS are specific to the individual user. Hence, the existing information system approaches alone cannot be most appropriate for studying behaviour because they only focus on cognitive behaviour. However, human behaviour is not only restricted to the cognitive character; it also has an emotional aspect [14].

Research concerning the behavioural aspect of customers regarding PRS is in a very nascent stage in the literature. PRS are more of a marketing tool today; hence, understanding the marketing aspect of PRS is necessary [18, 19]. PRS adds value to the customer [10], firstly by reducing the search time of the customer. Secondly, it helps them to compare the alternatives; thirdly it makes the customers feel more confident and satisfied about their choices. Similarly, Customer Perceived Value (CPV) also impacts the customer's buying behaviour [20]. PRS can be considered a tool that adds value to customers on existing online shopping websites.

Customer Engagement (CE) provides evidence of beyond-purchase psychological and behavioural relationships [21]. Engagement may not necessarily be a transactional behaviour; the customer is not restricted to the consumer but also contributes to the brand, business product, etc. [22]. Recommendation systems work on the customers' data; the more the customer is engaged with the system, the better would be the recommendation. Hence, without engagement, recommendation systems lack the prediction power.

Considering the marketing aspect of PRS, just selling isn't sufficient; the elementary task of PRS is to help customers find the desired product or service. It not only adds value to the customer, but PRS is a win–win tool, which should favour both customers as well as business owners. The present research attempts to introduce a framework to measure the influence of PRS on CE and CPV. In Information Systems literature, various models for technology acceptance are available; for the present research, Task Technology Fit (TTF) [23] is the theoretical background for our study. The other technology models are focused on individual acceptance of technology, TTF, while taking care of the implementation and the performance [24]. In the present research, the customers' response to the PRS is intended to be studied. Other IT models don't hold.

Present research would significantly contribute to the lack of literature considering marketing and behavioural aspects of the PRS by providing a model framework to measure the CE and CPV.

2 Conceptual Framework and Hypothesis

2.1 Product Recommendation Systems

PRS can identify and provide several alternatives that may fulfil the needs of the customer and can provide advice on the customer's selections. PRS aims to enhance the overall customer experience, which in turn increases the customer's satisfaction in decision-making as well as engagement with the platform [25]. PRS are important tools for generating revenue for any E-commerce business; according to McKinsey & Company [19], 35% of business on Amazon.com is due to the PRS, and 75% of content played on Netflix by the users comes due to these recommendations.

Anderson [18] has discussed some very important and interesting facts regarding recommendation systems. He explained the long tail approach; which involves making popular items available both online and offline, whereas the less popular items cannot be available offline. Extending discussion, he quoted an example about two books, "Touching the Void" and "Into Thin Air" where initially, Touching the Void wasn't very successful but after a decade, Into Thin Air came up in the market, having a similar story of the author. The recommendation systems picked it up and recommended the readers of "Into Thin Air "Touching the Void", and as a result, the recommendations of Touching the Void raised and became even more popular than Into Thin Air. This explains how powerful recommendations can be; it's not always necessary that not many popular items are useful. This is the crux behind the long tail method.

PRS come under the domain of information systems. While, PRS also possess characteristics of decision support systems (DSS) and knowledge-based systems (KBS). RS serves as decision aids for the customers, while DSS are support systems for the managers and analysts [26]. Additionally, PRS provide reasoning to support the recommendations to create trust in customers [9].

2.2 Customer Perceived Value

CPV, as defined by Zeithaml [17], is "*consumer's overall assessment of the utility of a product based on the perceptions of what is received and what is given*". In relationship marketing, everything revolves around value. Value of a product may differ for two customers, as the value is assessed based on the individual's perception.

The customer may perceive greater value from a product or service when the advantages outweigh the associated pre-purchase costs; the term "*costs*" in this context comprises not only monetary value but also the cost of searching, cost of time, and mental and emotional cost involved in pre-purchase [27].

CPV has been described as a twin-fold concept in the literature, it is the ratio between the perceived takeaways and perceived sacrifices of the customer in searching, evaluating, purchasing, and using the product/service. Subsequently, scholars established CPV as a multidimensional domain. Sheth [28], in particular, proposed CPV as a five-dimensional concept based on social value, epistemic value, conditional value, emotional value and functional value. Whereas Sweeney [29] defined it as a three-dimensional concept that involves emotional, social and functional value.

CPV has been studied at large, particularly with the prospect of online shopping satisfaction and loyalty; CPV is closely associated with PRS. CPV has a significant influence on the buying behaviour of the customer in the pre-purchase stage [20], it is the main objective of the PRS. PRS are nothing but a value addition to the existing online shopping website. This facilitates the customer's task to search for a product, find new products alternatives and help in developing demand. Hence, it is essential to include CPV in PRS literature. To assess the effectiveness of PRS, it is necessary to evaluate the value that is added to the customer rather than solely relying on click rates and purchases.

2.3 Customer Engagement

There are three different perspectives on CE in literature: psychological process [30], focal object motivation perspective [21, 31] and non-transactional perspective [32]. The main focus of Marketing Science Institute (MSI) in priorities (2014–2016) includes the question of "*How should engagement be conceptualised, defined, and measured*?" The existing CE literature is inadequate since it is unable to clarify the antecedents and consequences of CE [33].

PRS are interactive platforms that utilise extensive long-term customer data. More amount of data leads to better recommendations, thereby emphasising the significance of engagement in PRS. The focal object concept does not hold well in PRS, as customer may possibly already be engaged with the product/brand. PRS cannot be seen solely as the medium of selling, its focus must be encouraging customer's return purchase. Most of the PRS are designed to recommend products that would complement past purchases, thus increasing the CE and probability of future transactions. The level of CE directly correlates with the potential for future business opportunities. This is because the success of the business depends on the interaction [34]. CE in the present research does not hold on the focal object as brand or firm, but rather on the customer intention. CE is an inherent necessity for the PRS, it can be defined as the behaviour through which customers make contributions either

explicitly or implicitly on the e-commerce platform, which may or may not result to actual monetary transactions [21, 30–32, 35, 36].

CE with PRS does not always necessitate an actual sales transaction; customers can just engage by using PRS to search for desired products. Consequently, increasing the exposure and knowledge of the customer, leading to a higher sense of satisfaction when they research rather than simply making a purchase [2]. Higher satisfaction level leads to more engagement [32, 37]. Thus, it is imperative to incorporate CE in the existing literature of PRS. While various studies have examined CE into the current corpus of PRS research in e-commerce, but have been neglected in PRS, as recommendations are the key factors for engaging customers in shopping online.

2.4 Conceptualisation of the Framework

There are various models for technology adoption that are present and have been widely used in practice. That can be classified into two different categories: First, *Individual Acceptance Model* that focuses on individual acceptance of technology by using intention or usage as a dependent variable. It includes TRA, TPB, TIB, TAM, UTAUT etc. [24]. Second, *Implementation Success Model* that focuses on the implementation and performance outcome of the technology (example: IS, TTF etc.).

The present research aims to derive a model for evaluating the customer's behavioural perspective to determine the influence of PRS on CE and CPV. The framework utilised in this research (Fig. 1), is derived from the Task Technology Fit (TTF) model. The TTF model, proposed in 1995 by Goodhue and Thompson, aligns with the DeLone and McLean IS Model of 1992 [23]. TTF examines the effects of technology use on task performance. TTF was employed earlier to elucidate the usage of knowledge management systems, blogs, and m-commerce. TTF has also been extended with the TAM and UTAUT in order to study the adoption behaviour of the customer [38, 39]. TTF has been employed to investigate the benefits gained by the users, with an adequate balance between technology and user requirements for a particular task [40]. In comparison to the other available models, TTF is suitable in the present context because it shows a significant and positive relationship when measuring the customer response to the use of technology [40]. While other theories discuss about adoption, customers using PRS have possibly already adopted this technology implicitly while shopping online. Therefore, the TTF model has been utilised as the primary constituent of the conceptual model. The distinct components of the conceptual model: Antecedents, Task Technology Fit, and Consequences are detailed in detail below:

Antecedents

Antecedents of TTF are the interactions among the Task Characteristics, Technology Characteristics and Individual Characteristics [23, 41, 42]

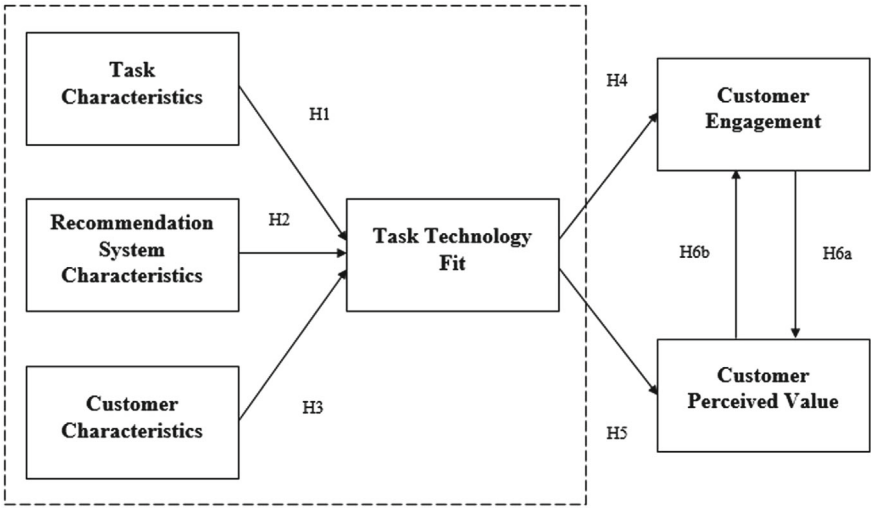


Fig. 1 Conceptual framework

Task Characteristics (TC)—The term "*task*" refers to the customer's use of an e-commerce website to find the desired product they intend to buy or explore. Therefore, TC measures the characteristics of the task.

H1: Task Characteristics have a positive influence on Task Technology Fit.

Technology Characteristics (TECHC)—In the present context of this research, TECHC refers to the characteristics of PRS. TECHC quantifies the functionality of PRS by evaluating how efficiently it responds to user queries based on user experience.

H2: Recommendation Systems Characteristics have a positive influence on Task Technology Fit.

Individual Characteristics (IDVC)—IDVC refers to the user or customer characteristics. IDVC measures the customer's assessment of their abilities while using online shopping websites to search and evaluate products they intend to investigate or buy. It also includes their expertise regarding the usage of the internet and online shopping frequency.

H3: Customer Characteristics have a positive influence on Task Technology Fit.

Consequences

The present research aims to measure the influence of customer response to the use of PRS on CE and CPV while shopping online.

PRS works on customer data to enhance its functionality. A data-rich system will be able to provide more efficient recommendations. Therefore, the greater the customer's engagement with the uses of PRS, the higher their satisfaction would be [32, 37]. Thus, PRS is intended to engage customers for a longer duration rather than

solely focusing on sales. This demonstrates that CE is a dependent variable that has a positive correlation with PRS.

H4: PRS positively influence CE in e-retail.

PRS efficiently improves the process of searching, comparing and finding related products, saving the customers valuable time. Additionally, customers can compare with similar products, which may differ in terms of price and benefits they offer [27]. It results in a more aware pre-purchase decision; these club up in providing the customer with more value. [2]. Therefore, PRS leads to an increase in CPV, which is a dependent variable that demonstrates a positive correlation with PRS.

H5: PRS positively influence CPV in e-retail.

The literature confirms the existence of the relationship between CE and CPV. An engaged customer would have a higher level of satisfaction [21, 43], resulting in better derived value [2, 25, 44]. However, clarification is lacking on the precise nature of the relationship between CE and CPV. Existing literature supports that CPV influences CE [45–48] and that CE can also have an influence on CPV [49, 50]. Therefore, it is equitable to expect that the relationship between CE and CPV is reciprocal in nature, hence forming a non-recursive model.

H6: The relationship between CE and CPV is non-recursive.

H6a: The influence of PRS on CE is influenced by CPV.

H6b: The influence of PRS on CPV is influenced by CE.

3 Conclusion

The framework proposed in this research aims to develop a stronger relationship between the business and the customers. This research is intended to find the influence of PRS on CE and CPV, which are key to increasing business and revenue. This would, in turn, help identify the necessary steps to increase engagement and value.

Customer relationship is the most important aspect; as most of the customers are just being pushed with products that are solely based on algorithms, the present framework will find out what lags in terms of the usefulness of these PRS. Furthermore, it would aid in suggesting better solutions to align with customer psychology, thereby enhancing the relationship with the e-commerce provider. Additionally, this would fill the void of scarce literature to measure the influence of customers' responses to the use of PRS on CE and CPV.

While PRS has been extensively researched in the past, this corpus of research is still highly active, with a predominant focus on the development of recommendation algorithms and models. Therefore, this study would belong to the club that showcases the customer psychological and behavioural aspects of PRS. The literature presents an unclear relationship between CE and CPV; while researches indicate a reciprocal relationship between these variables. The Lucidness of the relationship between CPV and CE is absent. Hence, this framework will fill this using the TTF model, which would significantly leverage future studies related to CE and CV.

References

1. Hamed, S., El-Deeb, S.: Cash on delivery as a determinant of E-commerce growth in emerging markets. J. Glob. Mark. **33**, 242–265 (2020). https://doi.org/10.1080/08911762.2020.1738002
2. Shugan, S.M.: The cost of thinking. J. Consumer Res. **7**, 99 (1980). https://doi.org/10.1086/208799
3. Chiu, Y.P., Lo, S.K., Hsieh, A.Y., Hwang, Y.: Exploring why people spend more time shopping online than in offline stores. Comput. Human Behav. **95**, 24–30 (2019). https://doi.org/10.1016/j.chb.2019.01.029
4. Martínez-López, F.J., Rodríguez-Ardura, I., Carlos Gázquez-Abad, J., Sánchez-Franco, M.J., Cabal, C.C.: Psychological elements explaining the consumer's adoption and use of a website recommendation system. Internet Res. **20**, 316–341 (2010). https://doi.org/10.1108/10662241011050731
5. Lin, Z.: An empirical investigation of user and system recommendations in e-commerce. Decis. Support. Syst. **68**, 111–124 (2014). https://doi.org/10.1016/j.dss.2014.10.003
6. Lin, A.J., Hsu, C.L., Li, E.Y.: Improving the effectiveness of experiential decisions by recommendation systems. Expert Syst. Appl. **41**, 4904–4914 (2014). https://doi.org/10.1016/j.eswa.2014.01.035
7. Senecal, S., Nantel, J.: The influence of online product recommendations on consumers' online choices. J. Retail. **80**, 159–169 (2004). https://doi.org/10.1016/j.jretai.2004.04.001
8. Ansari, A., Essegaier, S., Kohli, R.: Internet Recommendation Systems (2000). https://doi.org/10.1509/jmkr.37.3.363.18779
9. Xiao, B., Benbasat, I.: E-commerce product recommendation agents: use, characteristics, and impact. MIS Q. **31**, 137–209 (2007). https://doi.org/10.1088/1751-8113/44/8/085201
10. Martínez-López, F.J., Esteban-Millat, I., Argila, A., Rejón-Guardia, F.: Consumers' psychological outcomes linked to the use of an online store's recommendation system. Internet Res. **25**, 562–588 (2015). https://doi.org/10.1108/IntR-01-2014-0033
11. Benbasat, I., Wang, W.: Trust in and adoption of online recommendation agents. J Assoc. Inf. Syst. **6**, 72–101 (2005). https://doi.org/10.17705/1jais.00065
12. Gan, X., Jiao, Y., Liu, L., Zhang, Y.: Research on the factors influencing users' adoption intention of E-commerce recommendation system. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 563–574. Springer International Publishing (2018). https://doi.org/10.1007/978-3-319-93803-5_53
13. Jones, N., Pu, P.: User technology adoption issues in recommender systems. In: Proceedings of 22nd HCI International Conference 2020 (2020)
14. Komiak, Y.X., Benbasat, I.: The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q. **30**, 941 (2006). https://doi.org/10.2307/25148760
15. Martínez-López, F.J., Esteban-Millat, I., Cabal, C.C., Gengler, C.: Psychological factors explaining consumer adoption of an e-vendor's recommender. Ind. Manag. Data Syst. **115**, 284–310 (2015). https://doi.org/10.1108/IMDS-10-2014-0306
16. Martinez-Lopez, F.J., Cabal, C.C., Gazquez-Abad, J.C., Rodriguez-Ardura, I.: A theoretical approximation to the psychological aspects related to the consumer's adoption and use of a website recommendation system. In: NAFIPS 2009—2009 Annual Meeting of the North American Fuzzy Information Processing Society, pp. 1–6. IEEE (2009). https://doi.org/10.1109/NAFIPS.2009.5156458
17. Shi, S., Gong, Y., Gursoy, D.: Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: a heuristic—systematic model. J. Travel Res. 004728752096639 (2020). https://doi.org/10.1177/0047287520966395
18. Anderson, C.: The long tail: why the future of business is selling less of more. In: Hyperion (2006)
19. MacKenzie, I., Meyer, C., Noble, S.: How Retailers Can Keep up with Consumers. McKinsey & Company. https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers. Last accessed: 2023/06/15 (2013)

20. Parasuraman, A., Grewal, D.: The impact of technology on the quality-value-loyalty chain: a research agenda. J. Acad. Mark. Sci. **28**, 168–174 (2000). https://doi.org/10.1177/0092070300281015
21. Vivek, S.D., Beatty, S.E., Morgan, R.M.: Customer engagement: exploring customer relationships beyond purchase. J. Mark. Theory Pract. **20**, 122–146 (2012). https://doi.org/10.2753/MTP1069-6679200201
22. So, K.K.F., King, C., Sparks, B.A., Wang, Y.: Enhancing customer relationships with retail service brands: the role of customer engagement. J. Serv. Manag. **27**, 112–116 (2016). https://doi.org/10.1108/JOSM-05-2015-0176
23. Goodhue, D.L., Thompson, R.L.: Task-technology fit and individual performance. MIS Q. **19**, 213–233 (1995). https://doi.org/10.2307/249689
24. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**, 425 (2003). https://doi.org/10.2307/30036540
25. Fitzsimons, G.J., Lehmann, D.R.: Reactance to recommendations: when unsolicited advice yields contrary responses. Mark. Sci. **23**, 82–94 (2004). https://doi.org/10.1287/mksc.1030.0033
26. Benbasat, I., Desanctis, G., Nault, B.R.: Empirical Research in Managerial Support Systems : A Review and Assessment l, vol. 6, pp. 203–226 (1993)
27. Slater, S.F., Narver, J.C.: The positive effect of a market orientation. J. Bus. Res. 69–73 (2000)
28. Sheth, J.N., Newman, B.I., Gross, B.L.: Why we buy what we buy: a theory of consumption values: discovery service for Air Force Institute of Technology. J. Bus. Res. **22**, 159–170 (1991)
29. Sweeney, J.C., Soutar, G.N.: Consumer perceived value: the development of a multiple item scale. J. Retail. **77**, 203–220 (2001). https://doi.org/10.1016/S0022-4359(01)00041-0
30. Brodie, R.J., Hollebeek, L.D., Jurić, B., Ilić, A.: Customer engagement. J. Serv. Res. **14**, 252–271 (2011). https://doi.org/10.1177/1094670511411703
31. Jaakkola, E., Alexander, M.: The role of customer engagement behavior in value co-creation: a service system perspective. J. Serv. Res. **17**, 247–261 (2014). https://doi.org/10.1177/1094670514529187
32. Verhoef, P.C., Reinartz, W.J., Krafft, M.: Customer engagement as a new perspective in customer management. J. Serv. Res. **13**, 247–252 (2010). https://doi.org/10.1177/1094670510375461
33. Maslowska, E., Malthouse, E.C., Collinger, T.: The customer engagement ecosystem. J. Mark. Manag. **32**, 469–501 (2016). https://doi.org/10.1080/0267257X.2015.1134628
34. Tsagkias, M., King, T.H., Murdock, V.: Challenges and research opportunities in eCommerce search and recommendations. SIGIR Forum. **54**, 1–23 (2020)
35. Bowden, J.: The process of customer engagement: a conceptual framework. J. Mark. Theory Pract. **17**, 63–74 (2009). https://doi.org/10.2753/MTP1069-6679170105
36. Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., Tillmanns, S.: Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value, pp. 297–310 (2010). https://doi.org/10.1177/1094670510375602
37. Monferrer, D., Moliner, M.A., Estrada, M.: Increasing customer loyalty through customer engagement in the retail banking industry. Spanish J. Mark. ESIC **23**, 461–484 (2019). https://doi.org/10.1108/SJME-07-2019-0042
38. Tam, C., Oliveira, T.: Performance impact of mobile banking: using the task-technology fit (TTF) approach. Int. J. Bank Mark. **34**, 434–457 (2016). https://doi.org/10.1108/IJBM-11-2014-0169
39. D'Ambra, J., Wilson, C.S.: Explaining perceived performance of the world wide web: uncertainty and the task-technology fit model. Internet Res. **14**, 294–310 (2004). https://doi.org/10.1108/10662240410555315
40. Lin, H.C., Han, X., Lyu, T., Ho, W.H., Xu, Y., Hsieh, T.C., Zhu, L., Zhang, L.: Task-technology fit analysis of social media use for marketing in the tourism and hospitality industry: a systematic literature review. Int. J. Contemp. Hosp. Manag. **32**, 2677–2715 (2020). https://doi.org/10.1108/IJCHM-12-2019-1031

41. Goodhue, D.L.: Development and measurement validity of a task-technology fit instrument for user evaluations of information systems. Decis. Sci. **29**, 105–138 (1998). https://doi.org/10.1111/j.1540-5915.1998.tb01346.x
42. Goodhue, D.L.: Understanding user evaluations of information systems. Manage. Sci. **41**, 1827–1844 (1995). https://doi.org/10.1287/mnsc.41.12.1827
43. Hollebeek, L.D.: The customer engagement/value interface: an exploratory investigation. Australas. Mark. J. **21**, 17–24 (2013). https://doi.org/10.1016/j.ausmj.2012.08.006
44. Vargo, S.L., Lusch, R.F.: Service-dominant logic: continuing the evolution. J. Acad. Mark. Sci. **36**, 1–10 (2008). https://doi.org/10.1007/s11747-007-0069-6
45. Itani, O.S., Kassar, A., Maria, S., Loureiro, C.: Value get, value give: the relationships among perceived value, relationship quality, customer engagement, and value consciousness. Int. J. Hosp. Manage. **80**, 78–90 (2019). https://doi.org/10.1016/j.ijhm.2019.01.014
46. Hoon, Y., Kim, D.J., Wachter, K.: A study of mobile user engagement (MoEN): engagement motivations, perceived value, satisfaction, and continued engagement intention. Decis. Support. Syst. **56**, 361–370 (2013). https://doi.org/10.1016/j.dss.2013.07.002
47. Xie, H., Zhang, Y., Duan, K.: Evolutionary overview of urban expansion based on bibliometric analysis in web of science from 1990 to 2019. Habitat Int. **95**, 102100 (2020). https://doi.org/10.1016/j.habitatint.2019.102100
48. Zhang, M., Bockstedt, J.: Complements and substitutes in online product recommendations: the differential effects on consumers' willingness to pay. Inf. Manage. **57**, 103341 (2020). https://doi.org/10.1016/j.im.2020.103341
49. Marbach, J., Lages, C.R., Nunan, D.: Who are you and what do you value? Investigating the role of personality traits and customer-perceived value in online customer engagement. J. Mark. Manag. **32**, 502–525 (2016). https://doi.org/10.1080/0267257X.2015.1128472
50. Zhang, M., Guo, L., Hu, M., Liu, W.: Influence of customer engagement with company social networks on stickiness: mediating effect of customer value creation. Int. J. Inf. Manage. **37**, 229–240 (2017). https://doi.org/10.1016/j.ijinfomgt.2016.04.010

Probabilistic Assessment on Natural Frequencies of Functionally Graded Plates Using Artificial Neural Network Approach



R. K. Rout, P. K. Karsh, and R. R. Kumar

Abstract The research paper provides a probabilistic study of free vibration for functionally graded material plates through an artificial neural network technique. The Latin hypercube sampling technique has been applied to train the ANN model. Monte Carlo simulation and the finite element methods is used to authenticate and verify the intended methodology for probabilistic natural frequency study of FGM plates. The Probabilistic variability of input factors namely density, modulus of elasticity, modulus of shear, Poisson ratio has been considered. The material parameters are calculated by utilizing the power law. The impact of different characteristics, such as temperature, power law exponent, twist angle and degree of variability on the initial three natural frequencies is evaluated. In contrast to the conventional Monte Carlo simulation, the suggested ANN model is highly computationally efficient. This ANN built finite element (FE) methodology can also be applied to other structural analyses involving complex geometries.

Keywords Functional graded material · Finite element method · Artificial neural network · Free vibration study · Stochastic study

1 Introduction

A functional graded material is a material that exhibits variations in its properties, density, composition, and microstructure throughout its volume [1]. FGMs are engineered to perform specific functions and are used in different engineering fields, involving, aerospace, biomedical, automotive, marine and civil structures [2]. In

R. K. Rout · R. R. Kumar (✉)
Department of Mechanical Engineering, National Institute of Technology Arunachal Pradesh, Arunachal Pradesh, India
e-mail: ravi@nitap.ac.in

P. K. Karsh
Department of Mechanical Engineering, Parul Institute of Engineering & Technology, Parul University, Vadodara, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_16

biomedical application, implants are manufactured from FGMs because many parts of human body consist of FGM as structures. The FGM's material properties differ on the specific position of the constituent materials within the material, resulting in varying properties throughout the volume. Mechanical, electrical, thermal and chemical characteristics vary with depth in FGMs to carry out particular tasks. Depending on the intended use, the design of FGMs incorporates both continuous and stepwise gradation of material qualities [3]. Compared to conventional composite materials, FGM is an advancement. The majority of composite structures feature a sharp interface between the two materials, which causes layer breakdown under high loads and temperatures. FGMs were created to overcome this restriction by substituting a continuous interface for the sharp one, which greatly lowers the likelihood of delamination. The FGMs are classified according to their chemical composition, microstructure, and porosity [4]. FGMs are a type of advanced material composed of two distinct materials: metal and ceramic. Figure 1 shows the dimensional view of an FGM plate with cantilever support, where the upper side is rich in ceramic material, and the lower side is rich in metallic component. Composites have certain limitations, such as failure under extreme working conditions due to delamination between layers. However, FGMs do not have distinct layers, making them more suitable for long-term use under harsh conditions [5]. The Random Factor Method (RFM), which accounts for random material properties, was used to dynamically assess FGM beams. The material variation in the FGM is modeled by power-law distribution. The results were compared to those obtained using Monte Carlo simulation (MCS) and first-order perturbation method. Several studies on the stochastic investigation of composites have been conducted in the past, exposing various approaches and methods [6–10]. Njim et al. investigates the natural frequency of FG porous plates using classical plate theory, validated by finite element analysis and experiments. It concludes that porosity and material gradients significantly influence natural frequencies, with strong agreement between methods [11]. Sharma et al. generalized differential quadrature (GDQ) method and Timoshenko beam theory to model the free vibration behavior of axially functionally graded (AFG) beams with power-law material gradients. It concludes that natural frequencies depend on volume fraction indices and beam length, showing consistent trends across parametric studies [12]. Burlayenko et al. develops an FE model using ABAQUS shell elements for analyzing free vibrations of FGM sandwich panels with varying material gradation laws. Employing the UMAT subroutine for property gradations, the model predicts natural frequencies with high accuracy and minimal computational cost. The findings highlight the influence of gradation types and panel configurations, providing valuable benchmark data [13]. Hadji et al. provides an analytical solution for FG nanoplates using a refined four-variable plate theory based on non-local elasticity. The model efficiently reduces computational complexity by simplifying the governing equations, analyzing the effects of parameters like boundary conditions, aspect ratios, and non-local factors on natural frequencies. The results demonstrate significant parameter-dependent influences on natural frequencies and validate the method as effective for non-local vibration analysis of FG nanoplates [14]. Baghbadorani et al. investigates the free vibrations of FG shells using FSDT, Donnell

kinematic relations, and the Halpin–Tsai model for elastic properties. An analytical solution is applied for simply boundary conditions to derive natural frequencies and mode shapes. Results show that increasing GPL weight fraction and optimized reinforcement patterns significantly enhance vibration performance, with FG-X pattern yielding the highest frequencies [15]. Miao et al. develops a unified method to analyse the vibration characteristics of dual-functional gradient carbon nanotube-reinforced composite (FG-CNTRC) laminated cylindrical shells. The Rayleigh–Ritz method combined with artificial spring technology was used to study the impact of CNT distribution, matrix material, and volume fraction on free vibration. Results show that CNT addition can increase the shell's natural frequency by up to 50%, with the maximum frequency occurring at specific thickness and material configurations. The study highlights the sensitivity of natural frequency to CNT volume fraction and matrix material changes [16]. Sinha et al. investigates the effect of cutout on free vibration of composite plates by using experimental analysis and finite element modelling. The FEM model accounts for factors like layer number, fibre orientation, weight fractions, aspect ratio, and cut-out size/position, with boundary conditions (clamped, simply supported, and cantilevered). Results show that natural frequencies increase with thickness and fibre content, and that cut-out size and position significantly influence frequency, providing valuable insights for design and structural health monitoring [17]. Duc et al. investigates the free vibration of cracked FG carbon nanotube reinforced composite. The study shows that increasing crack length, plate length, and thickness ratio reduces the vibration frequency, with boundary conditions playing a key role in stiffness. The vibration frequency increases with CNT weight ratio, while crack slope slightly decreases the frequency. The results offer insights for future research on cracked FG CNTRC plate vibration behaviour [18]. Arani et al. presents a study on the vibration behaviour of FG cylindrical panels. It explores how the orientation, distribution, and volume fraction of carbon nanotubes (CNTs), along with the geometrical parameters, influence the natural frequencies and dynamic responses of the panel. The results indicate that CNTs arranged in certain patterns (e.g., circumferential direction) lead to reduced vibration amplitudes, and increasing CNT volume fraction enhances the natural frequencies. The study also highlights the impact of geometrical factors like length and thickness on the panel's vibrational characteristics [19]. Verenkar et al. investigates the free vibration of composite plate with and without delamination, using FE with the ACP module. The results align well with existing literature for intact plates. The analysis shows that delamination have impact on higher vibration modes, with considerable impact on modes 4, 5, and 6. A noticeable difference in natural frequencies of intact vs. damaged plates occurs as constraint levels increase. The ACP module, tailored for composite analysis, proves superior to ANSYS APDL for composite structures due to its specialized tools [20].

The stochastic natural frequencies of FGM plates are analyzed in the current study using an artificial neural network (ANN). Input parameters include material and geometric characteristics like temperature, twist angle, and power-law index. There are six sections to the paper: The mathematical formulation is explained in Sect. 2, the power-law approach to material modelling is covered in Sect. 2, the ANN method is efficient to the stochastic dynamic investigation of FGM with distinct constraints

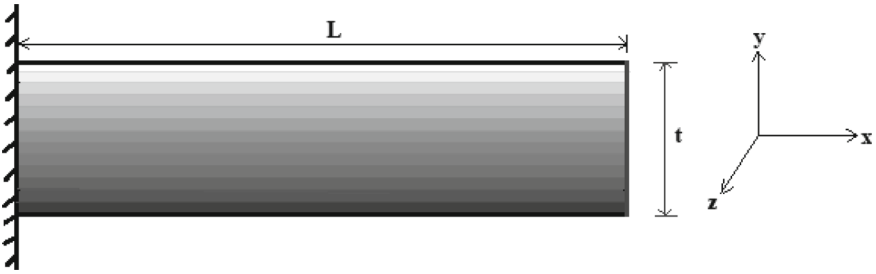


Fig. 1 Geometrical representation of FGM plate

in Sect. 3, the consequences are discussed in Sect. 4, and the study's conclusions and recommendations for further research to expand on this work are presented in Sect. 5.

2 Material Modelling Formulations

The material characteristics of FGM are influenced by ambient temperature. The function of material attributes "S" and temperature (T) are associated as,

$$M = M_0 + M_{-1}T^{-1} + 1 + M_1T + M_2T^2 + M_3T^3 \tag{1}$$

where the temperature coefficients in Kelvin are M_0, M_{-1}, M_1, M_2, and M_3 [21]. The material's effective characteristics, ensuring a smooth gradation of properties from one material to another throughout the thickness of the plate is evaluated by employing power-law distribution [11],

$$M(t) = M_m + (M_c - M_m)\left[\frac{z}{t} + \frac{1}{2}\right]^p \tag{2}$$

In the context of FGM plates:

- M_c represents the material properties of the ceramic element.
- M_m represents the material properties of the metallic component.
- t denotes the width of the plate.
- p represents power-law interpreter, which defines the gradation profile of the material properties.
- $z = t/2$ refers to the position across the thickness of the plate, with $z = 0$ at the mid-plane.

3 Stochastic Methods Utilizing ANN Model

In this work, the best intricate input factors namely, density, modulus of elasticity, modulus of rigidity and Poisson's ratio are considered for combined variation. In addition, stochastic free vibration analysis uses the material and geometrical properties of the FGM plate, along with the ambient temperature, as input parameters.

This approach allows for a comprehensive understanding of the stochastic behavior of FGM plates under various condition:

1. Combination of different material characteristics

$$g\{M(\overline{\omega})\} = \{\phi_1(M_1(1),\ M_1(2), \ldots M(i))\}$$

2. Distinct variances in the stochasticity percentage

$$P(\overline{\omega}) = \{P_{(1)}, P_{(2)}, \ldots P_{(i)}\}$$

3. Distinct temperature variance

$$T(\overline{\omega}) = \{T_{(1)}, T_{(2)}, \ldots T_{(i)}\}$$

4. Distinct variance in index of power law

$$p(\overline{\omega}) = \{p_{(1)}, p_{(2)}, \ldots p_{(i)}\}$$

5. Distinct variance in angle of twist only

$$\psi(\overline{\omega}) = \{\psi_{(1)}, \psi_{(2)}, \ldots \psi_{(i)}\}$$

The building of the ANN is illustrated in Fig. 2, which involves of three main components: weights, biases, and an activation function. An artificial neural network is modeled after the structure and function of biological neural networks, serving as a computational tool designed to reduce both time and cost. In an ANN, the cause and effect relationship is learned by training the network with input and output information set. ANN is also well-suited for complex systems that require the analysis of multiple inputs and outputs [22]. Figure 3 presents the steps involved in performing probabilistic natural frequency by employing the ANN model.

The bias (bi) is added to the sum of the inputs and their respective weights [23], shown as:

$$Z_i = \sum_{i=1}^{n} w_{ij} I_i + b_i \tag{3}$$

Z_i is transformed by using an activation function sometimes referred to as a transfer function, *R*i outcome of this modification provides a value known as unit's

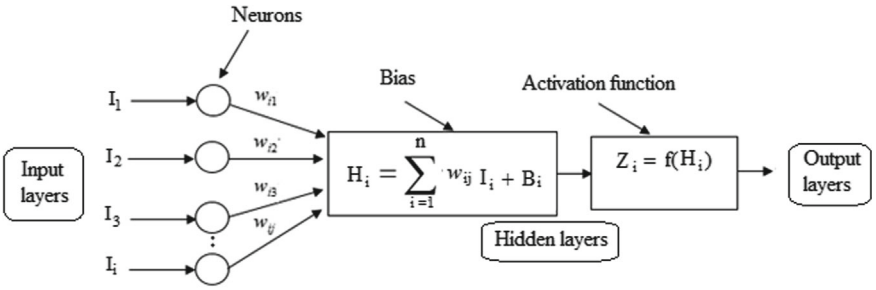


Fig. 2 Components of ANN

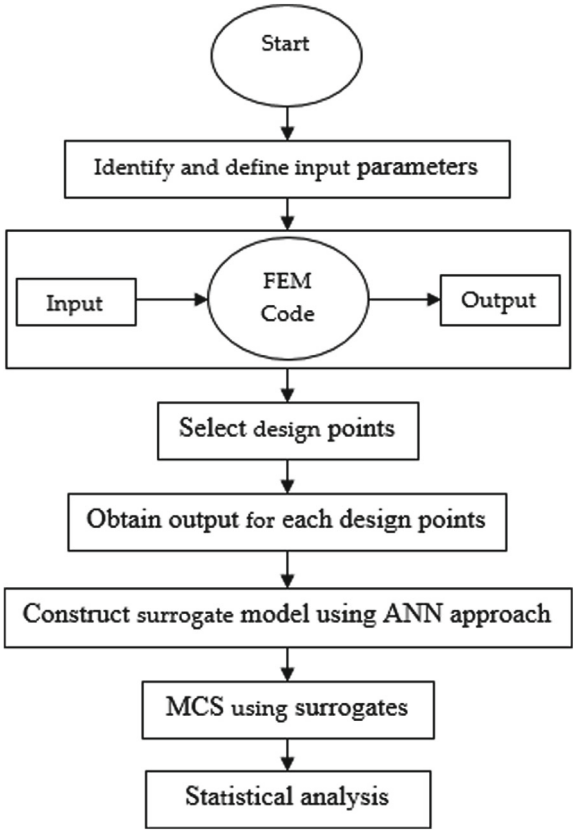


Fig. 3 Stages of probabilistic free vibration analysis

"activation"

$$R_i = f(Z_i) \tag{4}$$

4 Results and Discussion

Here we uses probability density function (PDF) plots as a conventional standard to evaluate the final outcomes. The outcomes suggested by the ANN model are compared to those derived from traditional MCS using the first three natural frequencies. A comparative analysis of the ANN model and MCS results is conducted. For both the ANN and MCS models, Fig. 4 displays the PDFs for the first three natural frequencies that come from the joint adjustment of mass density, Poisson's ratio, shear modulus, and modulus of elasticity. Larger sample sizes enhance the accuracy of the ANN model; for example, N = 1024 produces results that are almost exactly the same as MCS. The natural frequencies (rad/s) derived from the ANN model in comparison to the original FE model, taken into account a sample size of N = 1024, are shown in scattergrams in Fig. 5. With a nearly same range for the most basic natural frequency, the results reveal that the ANN model closely resembles MCS results. The impacts of temperature, twist angle, power law exponent, and degree of stochasticity are also evaluated using the ANN model. The PDFs of the first three natural frequencies (in rad/sec) for 10%, 20%, and 30% stochasticity (PP) are shown in Fig. 6. Figure 6 makes it clear that the stochasticity of all three natural frequencies grows as PP does. In contrast to P = 20% and P = 30%, the stochastic fluctuation in the output quantity of interest (QoI) is negligible for P = 10% for all natural frequencies.

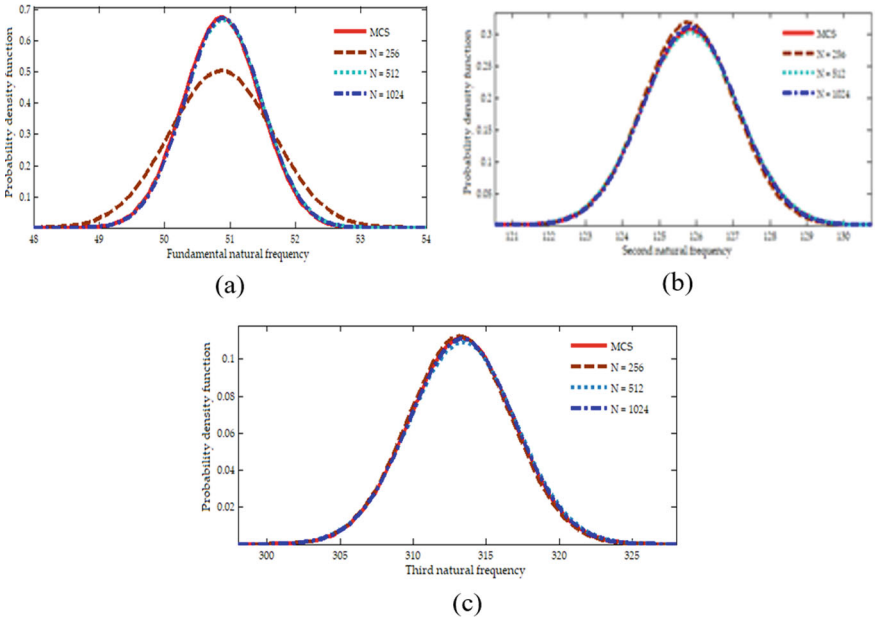


Fig. 4 Comparative study between MCS and ANN model (N)

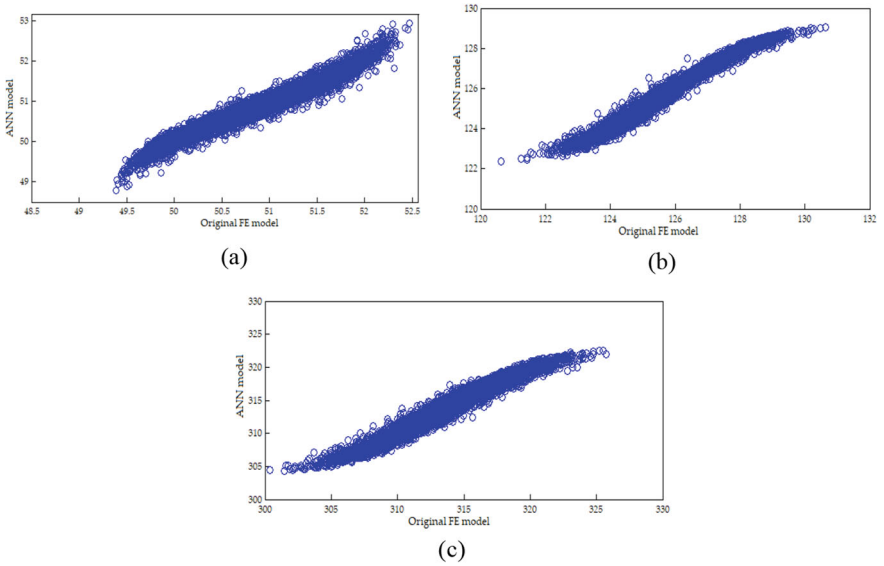


Fig. 5 Scattergrams for original FE MCS versus ANN model (N = 1024)

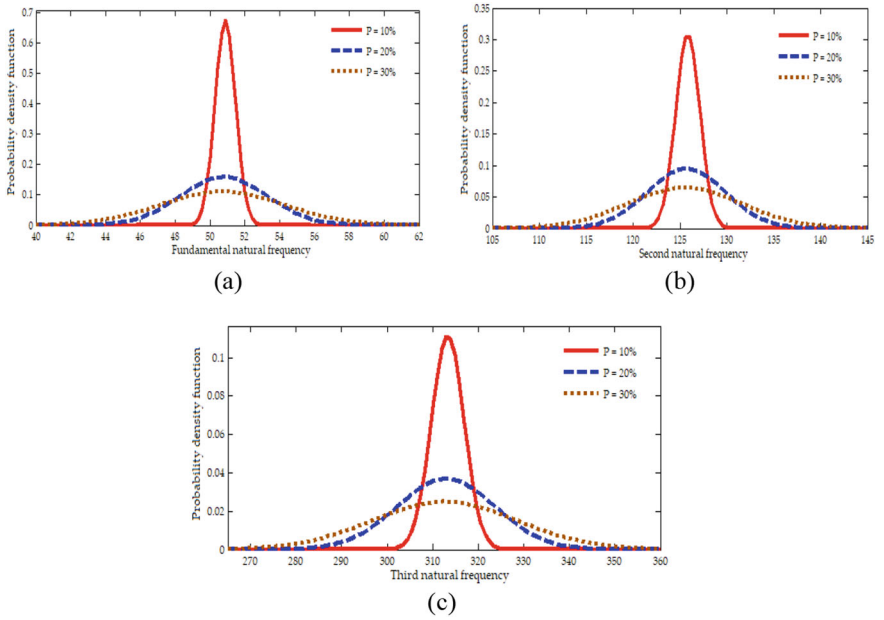


Fig. 6 Consequences of degree of stochasticity (P)

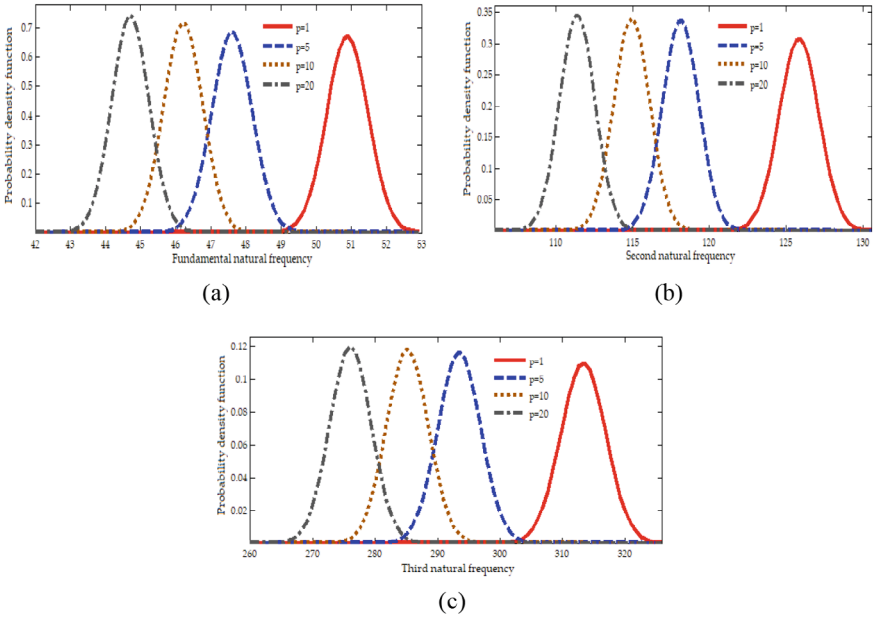


Fig. 7 Consequences of power law exponent (p)

Figure 7 shows how natural frequencies are influenced by the power-law index (p), showing that all natural frequencies decrease as the index increases. The reason for this decline is that the material's strength deteriorates with increasing p. Figure 8 shows how temperature affects the first three natural frequencies. It is found that all natural frequencies drop with increasing temperature because the FGM structure becomes less stiff at higher temperatures. Figure 9 illustrates how the plate's twist angle affects natural frequencies. The results show that when the twist angle increases, the first natural frequency gradually declines. Conversely, other frequencies first climb and then decline when the plate's twist angle grows further.

5 Conclusions

This study examines the free vibration of FGM plates using a stochastic finite element method and an artificial neural network. The joint variation of material attributes, along with the distinct change in other factors, is taken as the independent parameters. The impact of constraints such as temperature, power-law exponent and twist angle and are also investigated by employing the ANN model. By using the ANN model, the number of iterations is considerably decreased from 10,000 to 1024, resulting in reduced time and cost compared to the traditional MCS, without sacrificing correctness. In further research, other surrogate models can be expanded

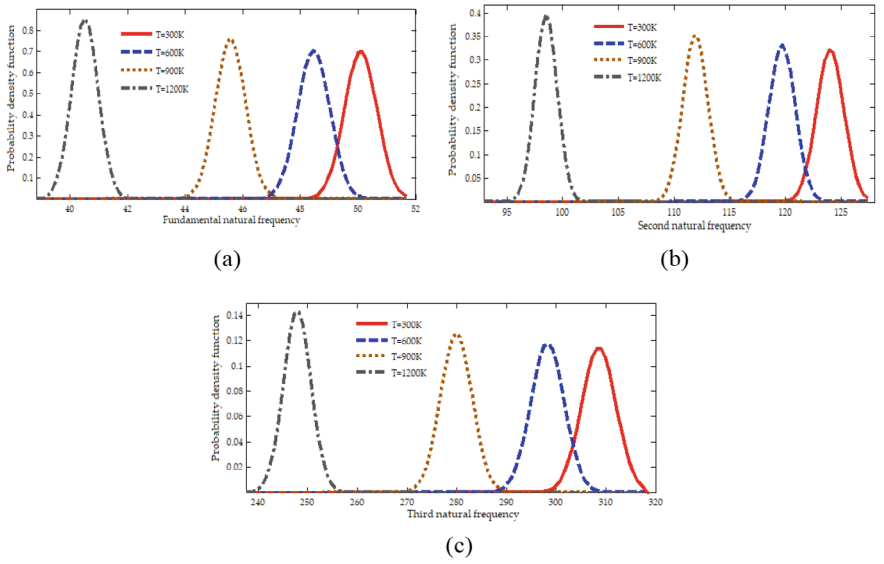


Fig. 8 Consequences of temperature

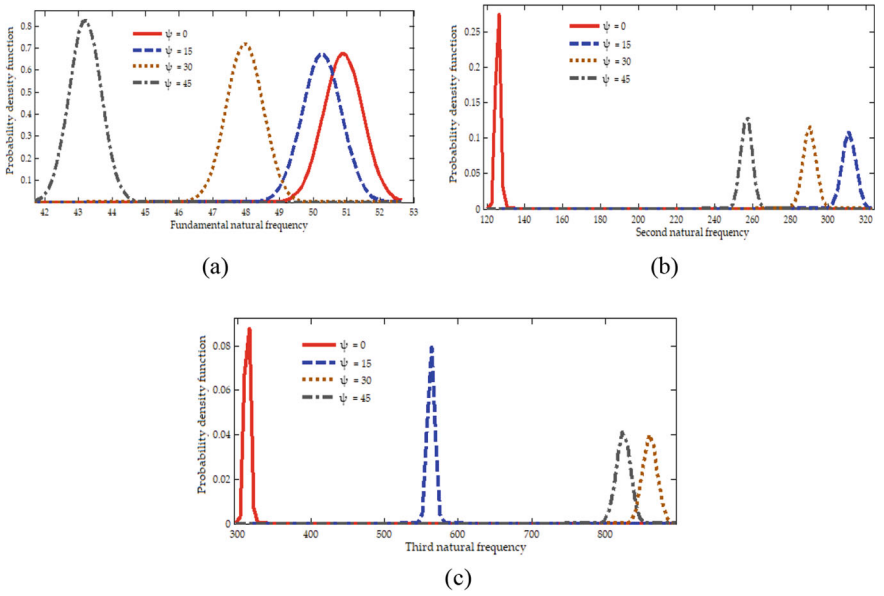


Fig. 9 Consequences of twist angle on the natural frequencies

to handle increasingly complicated structures, showing its usefulness in stochastic dynamic analysis.

References

1. Niino, M., Hirai, T., Watanabe, R.: The functionally gradient materials. J. Jpn. Soc. Compos. Mater. **13**, 257–264 (1987)
2. Report on Fundamental study on relaxation of thermal stress for high-temperature material by tailoring the graded structure. Department of Science and Technology Agency (1992)
3. Dumont, A.L., Bonnet, J.P., Chartier, T., Ferreira, J.M.F.: $MoSi_2/Al_2O_3$ FGM: elaboration by tape casting and SHS. J. Eur. Ceram. Soc. **21**, 2353–2360 (2001)
4. Nemat-Alla, M.M., Ata, M.H., Bayoumi, M.R., Khair-Eldeen, W.: Powder metallurgical fabrication and microstructural investigations of aluminum/steel functionally graded material. Mater. Sci. Appl. **2**, 1708–1718 (2011)
5. Vimal, J., Srivastava, R.K., Bhatt, A.D., Sharma, A.K.: Free vibration analysis of moderately thick functionally graded skew plates. Eng. Solid Mech. **2**, 229–238 (2014)
6. Kumar, R.R., Pandey, K.M., Dey, S.: Radial basis function based probabilistic buckling behavior of sandwich plates. In: Computational Mathematics, Nanoelectronics, and Astrophysics: CMNA 2018, Indore, India, November 1–3, pp. 3–11. Springer Singapore (2021)
7. Vaishali, Karsh, P.K., Kushari, S., Kumar, R.R., Dey, S.: Stochastic free vibration and impact responses of functionally graded plates: a support vector machine learning model approach. J. Vibration Eng. Technol. **11**(7), 2927–2943 (2023)
8. Karsh, P.K., Thakkar, B., Kumar, R.R., Dey, S.: Probabilistic oblique impact analysis of functionally graded plates—a multivariate adaptive regression splines approach. Eur. J. Comput. Mech. 223–254 (2021)
9. Kumar, R.R., Karsh, P.K., Pandey, K.M., Dey, S.: Stochastic natural frequency analysis of skewed sandwich plates. Eng. Comput. **36**(7), 2179–2199 (2019)
10. Kumar, R.R., Mukhopadhya, T., Pandey, K.M., Dey, S.: Prediction capability of polynomial neural network for uncertain buckling behavior of sandwich plates. In: Handbook of Probabilistic Models, pp. 131–140. Butterworth-Heinemann (2020)
11. Njim, E.K., Bakhy, S.H., Al-Waily, M.: Free vibration analysis of imperfect functionally graded sandwich plates: analytical and experimental investigation. Arch. Mater. Sci. Eng. **111**(2), 49–65 (2021)
12. Sharma, P., Singh, R.: A numerical study on free vibration analysis of axial FGM beam. Mater. Today Proc. **44**, 1664–1668 (2021)
13. Burlayenko, V.N., Sadowski, T., Altenbach, H.: Efficient free vibration analysis of FGM sandwich flat panels with conventional shell elements. Mech. Adv. Mater. Struct. **29**(25), 3709–3726 (2022)
14. Hadji, L., Avcar, M., Civalek, Ö.: An analytical solution for the free vibration of FG nanoplates. J. Braz. Soc. Mech. Sci. Eng. **43**(9), 418 (2021)
15. Baghbadorani, A.A., Kiani, Y.J.C.S.: Free vibration analysis of functionally graded cylindrical shells reinforced with graphene platelets. Compos. Struct. **276**, 114546 (2021)
16. Miao, X., Li, C., Jiang, Y.: Free vibration analysis of metal-ceramic matrix composite laminated cylindrical shell reinforced by CNTs. Compos. Struct. **260**, 113262 (2021)
17. Sinha, L., Das, D., Nayak, A.N., Sahu, S.K.: Experimental and numerical study on free vibration characteristics of laminated composite plate with/without cut-out. Compos. Struct. **256**, 113051 (2021)
18. Duc, N.D., Minh, P.P.: Free vibration analysis of cracked FG CNTRC plates using phase field theory. Aerosp. Sci. Technol. **112**, 106654 (2021)
19. Arani, A.G., Kiani, F., Afshari, H.: Free and forced vibration analysis of laminated functionally graded CNT-reinforced composite cylindrical panels. J. Sandwich Struct. Mater. **23**(1), 255–278 (2021)

20. Verenkar, S., Sridhar, I., Gouda, P.S.S.: Numerical investigations on free vibration analysis of delaminated composite plate. Indian J. Sci. Technol. **16**(42), 3714–3720 (2023)
21. Touloukian, Y.S.: Thermophysical Properties of High Temperature Solid Materials. McMillan, New York (1967)
22. Pareek, V.K., Brungs, M.P., Adesina, A.A., Sharma, R.: Artificial neural network modeling of a multiphase photodegradation system. J. Photochem. Photobiol. A **149**, 139–146 (2002)
23. Manohar, B., Divakar, S.: An artificial neural network analysis of porcine pancreas lipase catalysed esterification of anthranilic acid with methanol. Process Biochem. **40**(10), 3372–3376 (2005)

Damage Detection of Box Girder Based on Impact Load and the Improved Wavelet Packet-Permutation Entropy



Yutao Zhou and Jyoti K. Sinha

Abstract Damage identification, both size and location, in any orthotropic steel bridge girders is important for timely maintenance and avoiding any catastrophic failure. Previous experimental studies have shown that the measured vibration acceleration data can be used to detect the damage of orthotropic plates of steel box girders when subjected to the sinusoidal excitation. In this paper, wavelet packet decomposition (WPD) is used to deal with the measured vibration acceleration response, and then the permutation entropy (PE) of the reconstructed signal is estimated. Then the damage index is defined according to the permutation entropy difference (PED) between the damage state and the health state, which is used to determine the actual location and relative size of the damage. Through finite element model of the case structure and vibration acceleration responses under different hammer loads are considered. This study demonstrates that the vibration response under different impact loads can also identify the severity and location of damage.

Keywords Steel box girder · Orthotropic plate · Wavelet packet decomposition (WPD) · Damage identification · Permutation entropy (PE) · Permutation entropy difference (PED) · Impact load

1 Introduction

The importance of damage diagnosis of steel box girder cannot be ignored. As a component of bridge structure, steel box girder bears the weight of vehicles and provides indispensable support for the whole bridge system. Over time, these beams are placed in a state of complex stress under the influence of many factors, such as traffic loads, environmental factors (such as temperature changes and exposure

Y. Zhou (✉) · J. K. Sinha
Dynamics Laboratory, School of Engineering, University of Manchester, Manchester, UK
e-mail: yutao.zhou@postgrad.manchester.ac.uk

J. K. Sinha
e-mail: jyoti.sinha@manchester.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_18

to corrosive substances), and material aging. As a result, various forms of failure can occur, from surface cracks and corrosion to more severe structural deformation. Timely and accurate diagnosis of this injury is critical for the following reasons. First, it enables engineers to assess the structural integrity of steel box girders and identify potential weaknesses or vulnerabilities that could compromise the overall stability of the bridge. Second, effective damage diagnosis facilitates proactive maintenance strategies, allowing targeted repairs or reinforcement to be implemented before minor issues escalate into more serious structural problems. This proactive approach not only helps prevent catastrophic failures and minimize the risk of bridge collapse, but it also extends the life of infrastructure and reduces the need for costly emergency repairs and replacements. Overall, the importance of steel box girder damage diagnosis underpins the safety, resilience and sustainability of bridge structures, ensuring that they provide continued functionality and reliability to the communities they serve.

The Wavelet Packet Permutation Entropy (WPPE) method has been applied in various fields, including structural health monitoring [1–3]. In this study, a time-based analysis was used to track the latest advances in WPPE applications to ensure that the literature review reflected the latest developments. In previous studies, WPPE has been used to analyze vibration signals obtained from structural responses under sinusoidal loading conditions [2]. These studies use WPPE to detect and quantify damage to engineered structures [2–4]. By analyzing the change of vibration signal entropy before and after damage, WPPE can effectively identify the structural dynamic changes related to damage, so as to accurately diagnose the location, severity and progression of damage [5]. One of the advantages of WPPE is that it can capture complex nonlinear characteristics of structural vibration signals, making it suitable for analyzing non-stationary and nonlinear signals commonly used in structural health monitoring applications [6]. In addition, WPPE has high computational efficiency and does not require a large amount of data preprocessing, which enhances its practicability for real-time monitoring and analysis of engineering structures [7]. Overall, the application of WPPE in previous studies has demonstrated its effectiveness as a tool for detecting and evaluating damage to engineered structures, helping to improve the safety and reliability of critical infrastructure assets.

Although the WPPE method has been successful in diagnosing the damage of steel box beams under sinusoidal excitation loads [2], its effectiveness under hammer loads or impact loads needs to be verified by simulation analysis. The reasons are twofold. First, hammer-load or impact load conditions represent dynamic loading scenarios—which may cause a different type and size of structural response compared to sinusoidal excitation. These dynamic loading conditions can lead to complex vibration patterns and signal characteristics that may affect the applicability and accuracy of WPPE methods in damage diagnosis. Secondly, through numerical simulation analysis, control experiments and input parameters can be precisely carried out, which is convenient to systematically study the performance of steel box girder under hammer or impact loads.

Therefore, the finite element model of high-precision steel box girder is established in this study. Input different hammer loads (impact loads) into the finite element

model and get different acceleration responses. And wavelet packet decomposition (WPD) method is used to analyze the measured vibration acceleration response, and then the permutation entropy (PE) of the reconstructed signal is estimated. Then the damage index is defined according to the permutation entropy difference (PED) between the damage state and the health state, which is used to detect the location and size of the damage. Through finite element model and vibration acceleration response under hammer load, the effectiveness of this method is further verified. By validating the WPPE method under these dynamic loading scenarios through simulation analysis, researchers can enhance confidence in its reliability and robustness for real-world applications, thus ensuring accurate and effective damage diagnosis of steel box girders across a range of loading conditions.

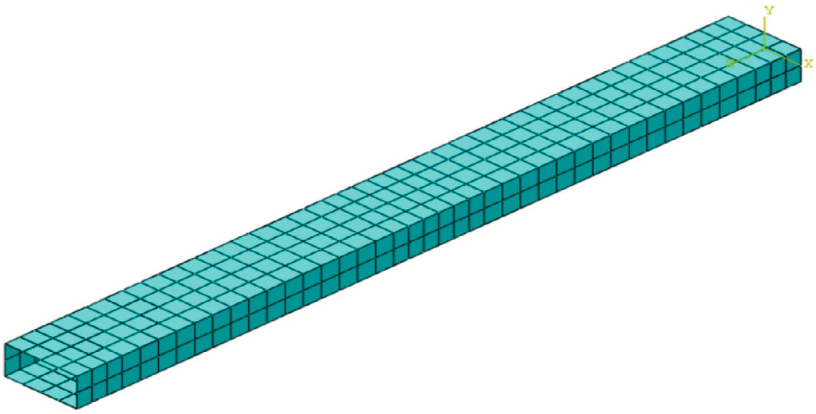
The first chapter is the introduction, which summarizes the background and significance of the study. In Sect. 2, the numerical simulation method and the loading method of the structure in the virtual experiment are described systematically. In the Sect. 3, the realization of WPPE theory is explained in detail, and in the Sect. 4, the virtual experiment is verified. Finally, in Sects. 5 and 6, the results of this study are discussed and summarized.

2 Simulation

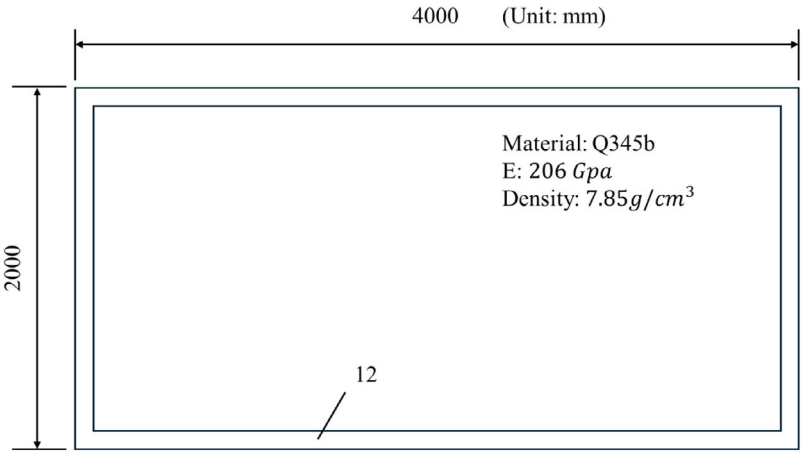
2.1 Finite Element Model

In order to verify the robustness of the WPPE theory, this paper has used a finite element (FE) model of the orthotropic plate of the steel box structure, as shown in Fig. 1a. The length of this virtual bridge is 50 m. The detail of the section is shown in Fig. 1b. The material of this model is Q345b, where the elasticity modulus is 206 Gpa, and the density is 7.85 g/cm^3. And this virtual bridge adopts simple support constraints at both ends. The first and second order frequency of this model is shown in Figs. 2.

Three conditions were considered in this paper, as shown in Table 1. The Condition 1 is healthy (no damage) and the damage sizes of 3 mm × 3 mm × 3 mm and 6 mm × 6 mm × 6 mm as the Conditions 2 and 3 respectively. The damages are simulated in the FE model at the top plate as shown in Fig. 3a. The actual location of the human-made damage is between sensors, 2–3 and 2–4, as shown in Fig. 3b where, 2 − n (n = 1, 2, …, 8) is the position of the sensors considered in the FE model. The acceleration vibration responses at these sensor locations are then estimated using the FE model for the impact load applied onto the top plate.



(a) Finite element model

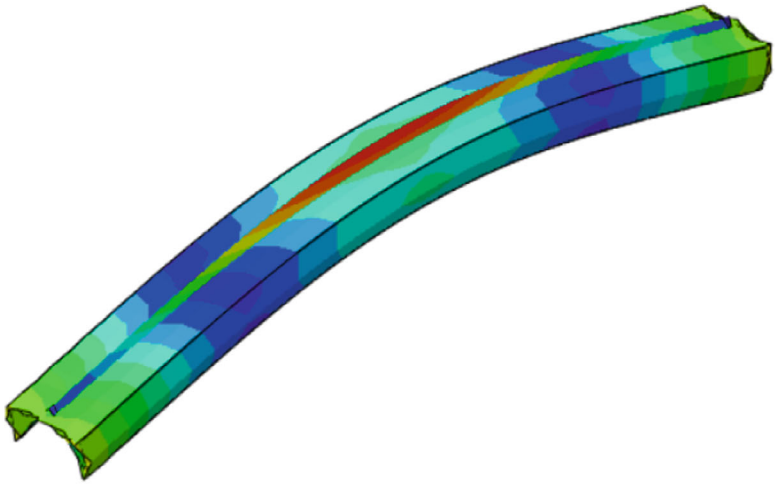


(b) Detail of the section

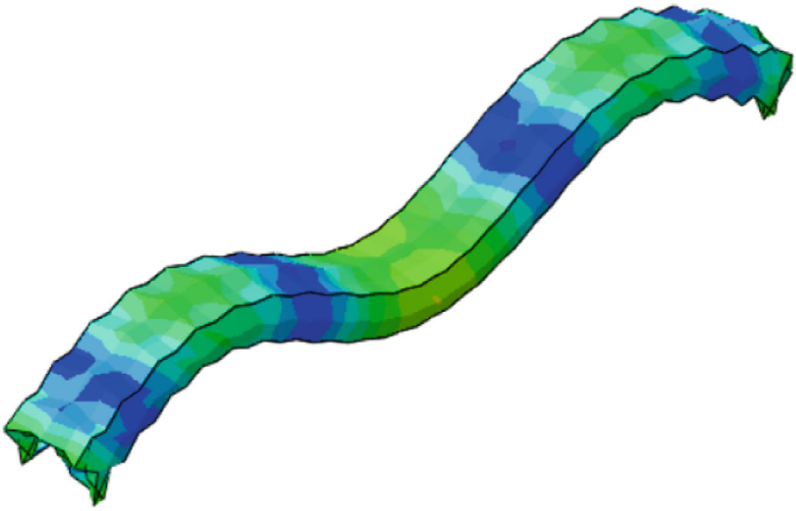
Fig. 1 Simulation of the steel girder

2.2 Loading Conditions (LC)

In order to fully verify the robustness of WPPE technique in identifying structural damage under the impact load for the excitation. The impact load by the hammer is considered as the virtual experiment and is shown in Fig. 4. The maximum value of the hammer load is 120N, and the load input time is 0.002 s.



(a) 1st frequency=8.12Hz (1st order symmetric vertical bending)

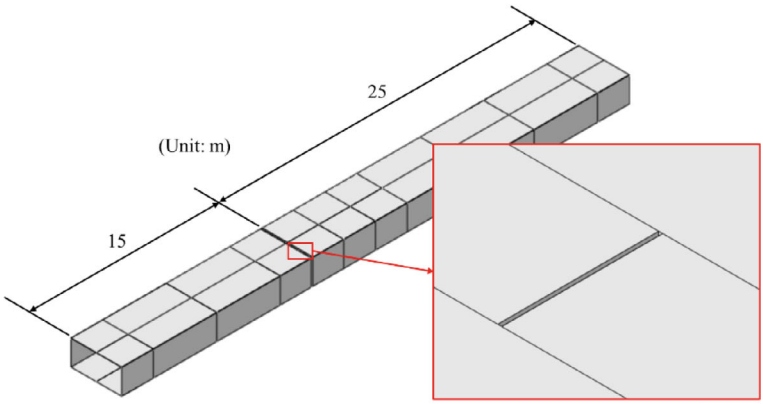


(b) 2nd frequency=36.324Hz (2nd order symmetric vertical bending)

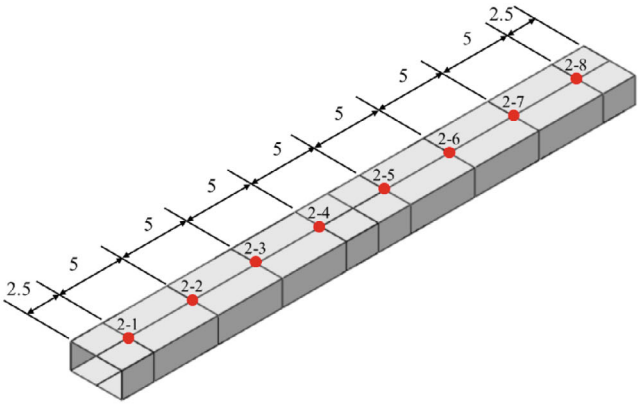
Fig. 2 Frequency and model shape of the simulation model

Table 1 Damage conditions (DC)

Condition	Length of damage (cm)	Width of damage (mm)	Depth of damage (mm)	Note
Condition 1	0	0	0	Healthy
Condition 2	3	3	3	Minor damage
Condition 3	6	6	6	Serious damage



(a) Simulation model of damage



(b) Location of sensors (Unit: m)

Fig. 3 Location of sensors and damage on bridge deck

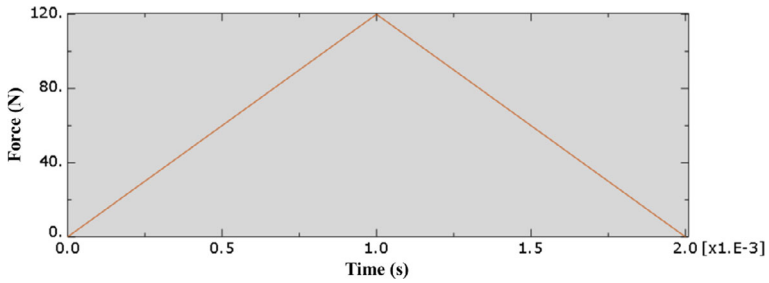


Fig. 4 Loading conditions (LC)

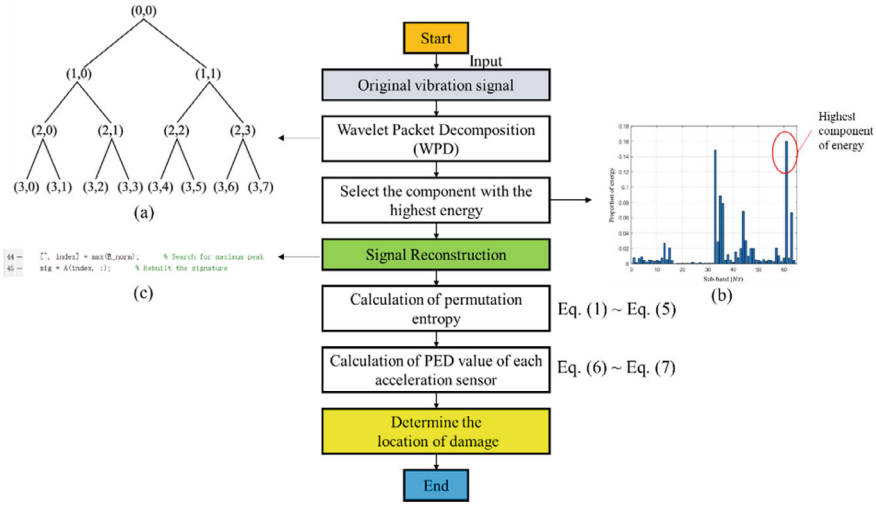


Fig. 5 Framework of WPPE

3 Wavelet Packet Permutation Entropy (WPPE)

The overall framework of the improved WPPE which is proposed in this paper as shown in Fig. 5.

4 Acceleration Responses and Data Analysis

The estimated acceleration vibration responses for Conditions 1, 2 and 3 for the input impact force for the excitation using the FE model are shown in Fig. 6 at the location of sensor #2–4. The computed acceleration spectra for the acceleration responses in Fig. 6 are shown in Fig. 7. These responses are used into the WPPE framework for

calculation, and the energy proportion within each band after WPD are obtained. These are shown in the Fig. 8.

Once the human-made structural damage is found, the energy proportion of each subband changes significantly after WPD. The subband with the highest energy ratio is selected for information remodeling. The calculated PE value and PED value of the structure before and after the damage, at all sensor locations are shown in Fig. 9. The PED value of the sensor far away from the damaged location is close to 0, indicating that the part of the structure is having no damage, whereas the PED values at the 2–4 and 2–5 sensors near the damage site showing higher values. This observation conforms the presence of the damage and the damage location. Higher PED values also indicate the bigger size of the damage for the Condition 3 compared to Condition 2.

5 Discussion

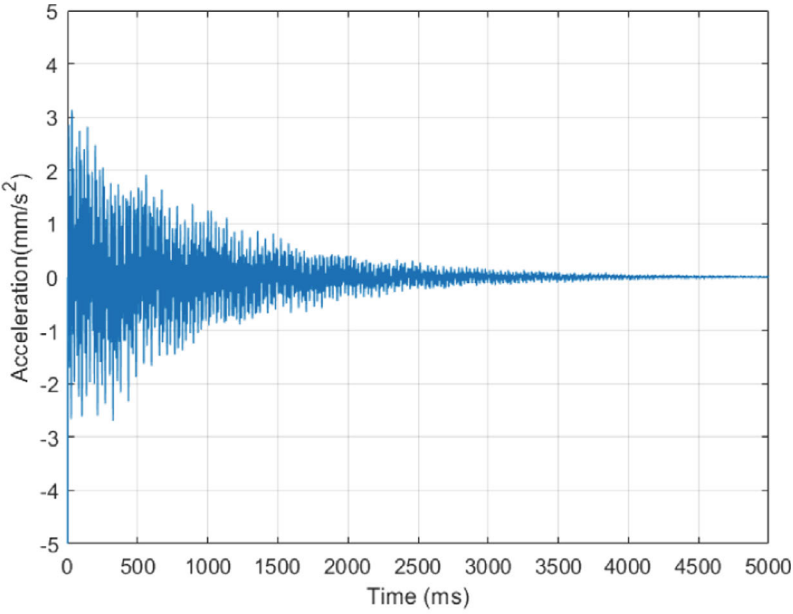
In our research, the WPPE method is used to quickly identify structural damage, which has important management significance in the field of structural health monitoring. Through the analysis of the vibration acceleration response of the structure, this method can effectively and quickly detect the health of the structure, especially in the non-destructive detection has a significant advantage.

Traditional vibration analysis methods, such as modal analysis or frequency-based damage detection methods, although capable of identifying structural damage, often require long detection times and complex equipment (such as shaking tables). In contrast, the WPPE method adopted in this study can significantly accelerate the detection speed through shock excitation, and does not rely on complex external equipment, indicating that it has higher operability and efficiency in engineering practice.

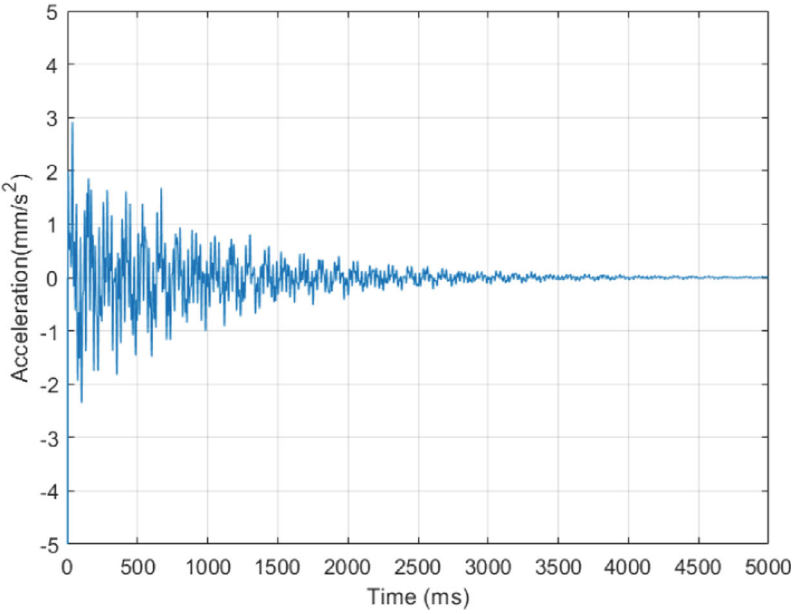
Other similar studies usually use time domain or frequency domain signal processing methods, such as wavelet transform, Fourier transform, etc., to identify the damage. These methods have limitations in dealing with non-stationary signals, but the wavelet packet decomposition (WPD) in this study can better deal with complex vibration signals, and improve the sensitivity to small damage by combining the arrangement entropy (PE), making it more suitable for early damage detection.

Some recent studies have employed machine learning methods for structural damage identification, and while these methods show good accuracy in some cases, they often require large amounts of training data and rely on complex modeling processes. In this study, WPPE method can effectively identify damage without large amount of data training, providing a simple and effective solution for small projects or new projects without large amount of historical data.

In addition, other non-destructive detection techniques, such as ultrasonic or infrared thermal imaging, can achieve structural damage detection to a certain extent, but their equipment is expensive and complex to operate. This study realizes non-destructive detection by using a simple shock excitation test, which not only has low

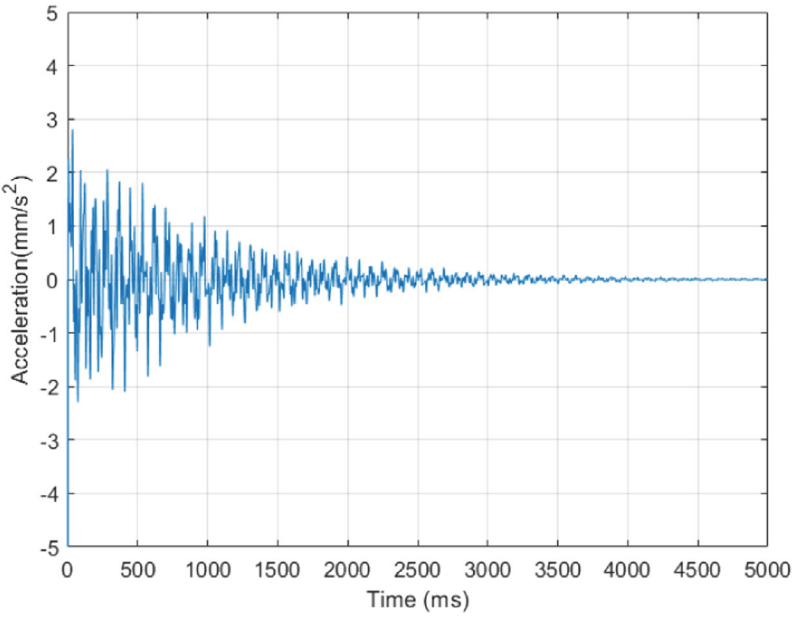


(a) DC1



(b) DC2

Fig. 6 Acceleration response in time domain (sensor 2–4)



(b) DC3

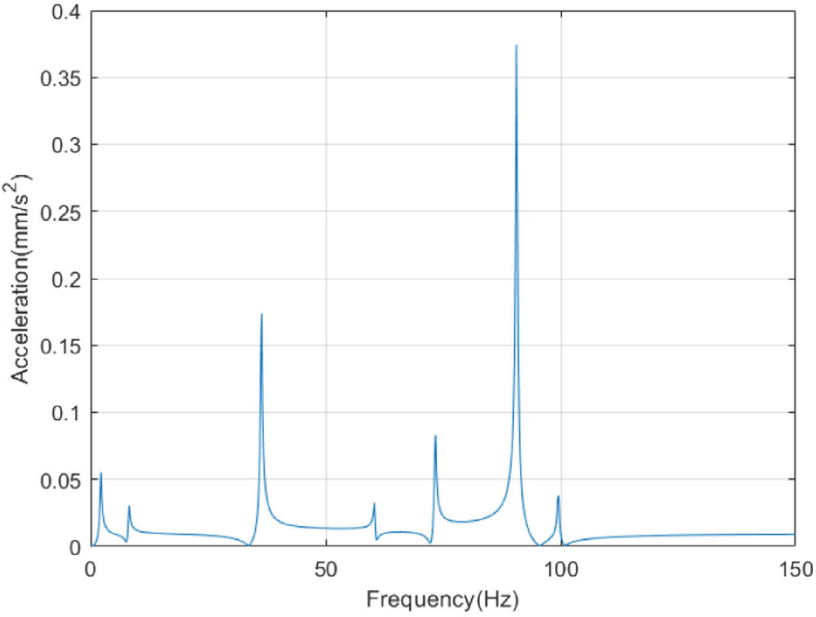
Fig. 6 (continued)

equipment requirements, but also can obtain detection results quickly, reducing cost and complexity.

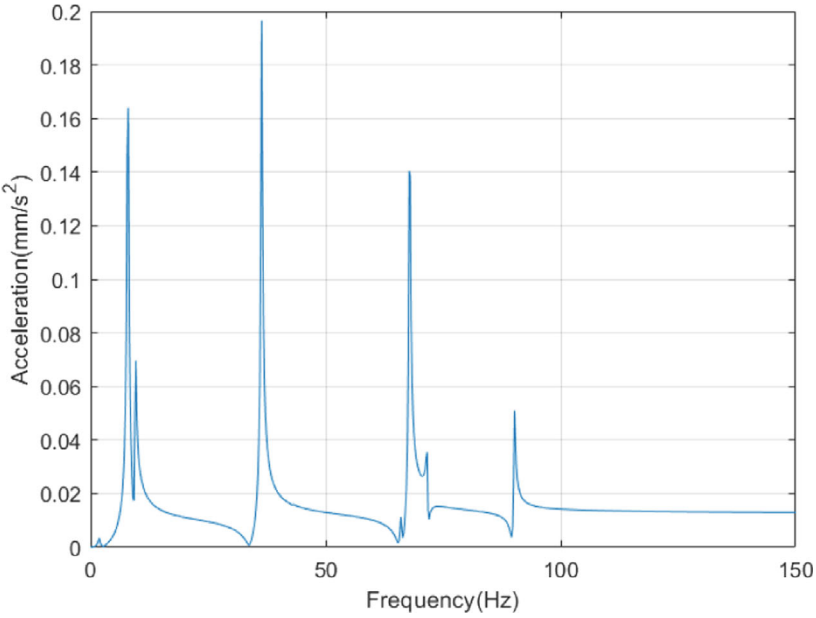
6 Conclusion

The WPPE method is used here to identify the damage in steel deck structures. The method uses WPD to analyze the data of vibration acceleration responses, and then the PE from the reconstructed signals to identify the damage. The method is further validated when the responses due to the impact input excitation through an FE model of a simple steel-deck structure. The impact excitation is generally a quick approach compared to the excitation by the shaker. Hence, the current study demonstrated that the health of any structures can be accessed quickly using the non-destructive impact tests.

In our research, the wavelet packet permutation entropy (WPPE) method has already been proved that it can be used to rapidly identify structural damage, which is of great significance in management. The method significantly reduces inspection time, improves maintenance decision efficiency, reduces operational costs, and improves safety through non-destructive impact testing. The wide applicability of the WPPE approach makes it an effective tool for managing diverse structural assets,



(a) DC1



(b) DC2

Fig. 7 Acceleration response in the frequency domain (sensor 2–4)

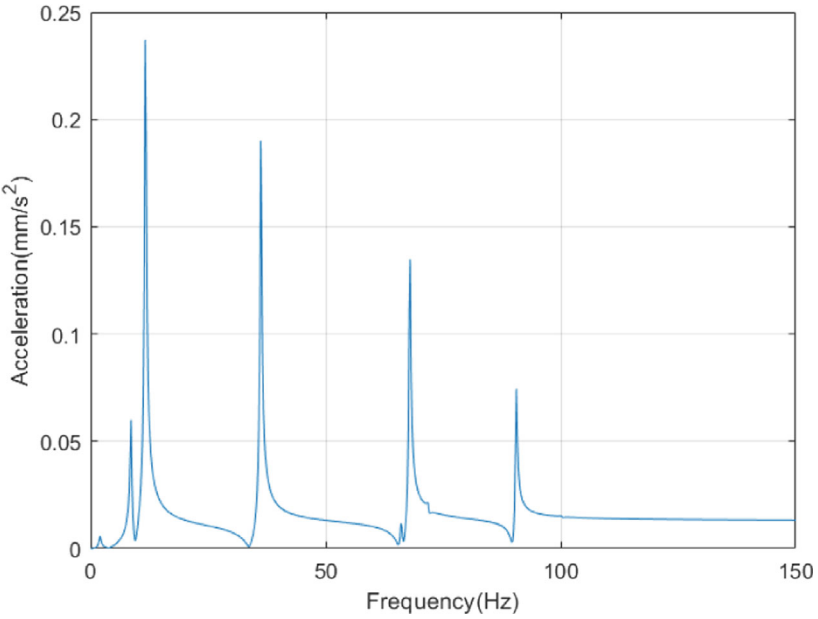
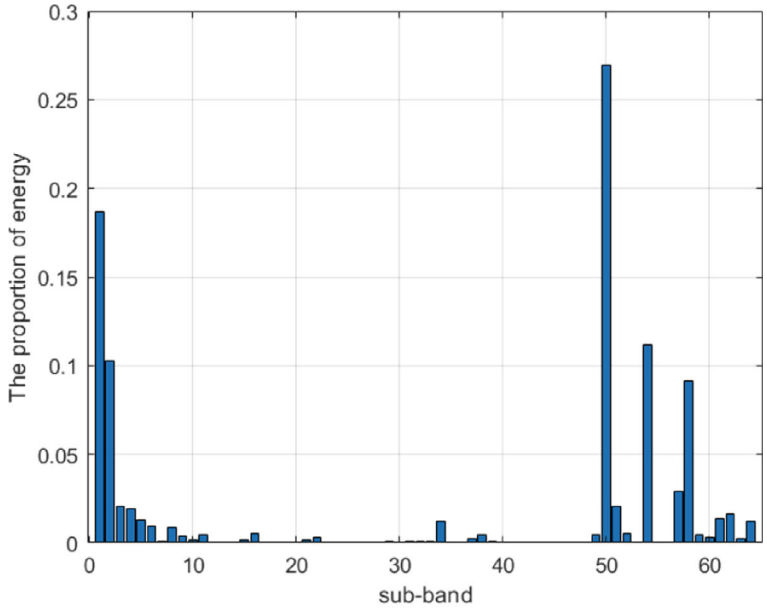
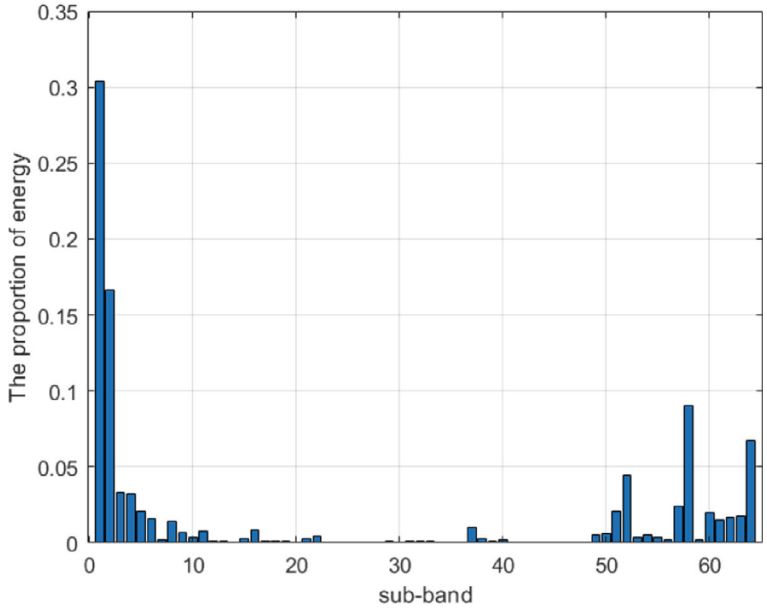


Fig. 7 (continued)

supporting the establishment of preventive maintenance systems that extend the life of structures, optimize resource allocation, and help managers ensure the safety of facilities and personnel while reducing downtime.

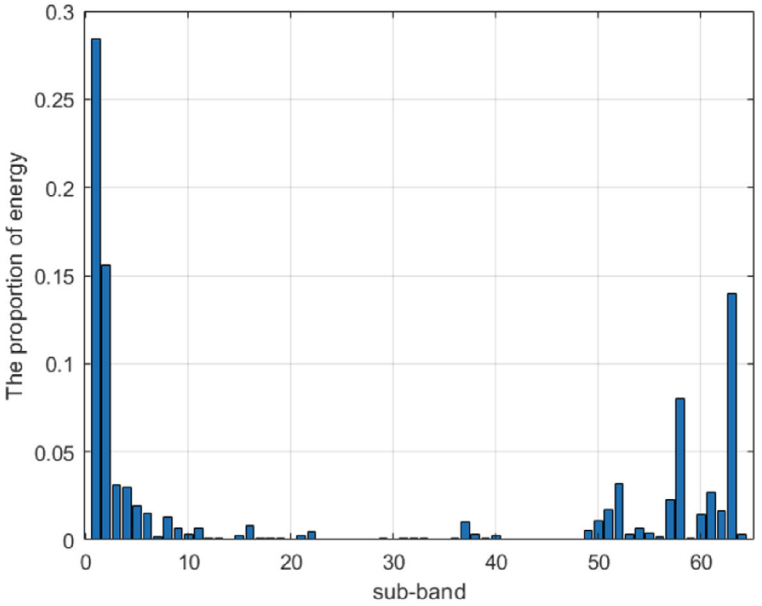


(a) DC1



(b) DC2

Fig. 8 Energy ratio in each sub-band



(c) DC3

Fig. 8 (continued)

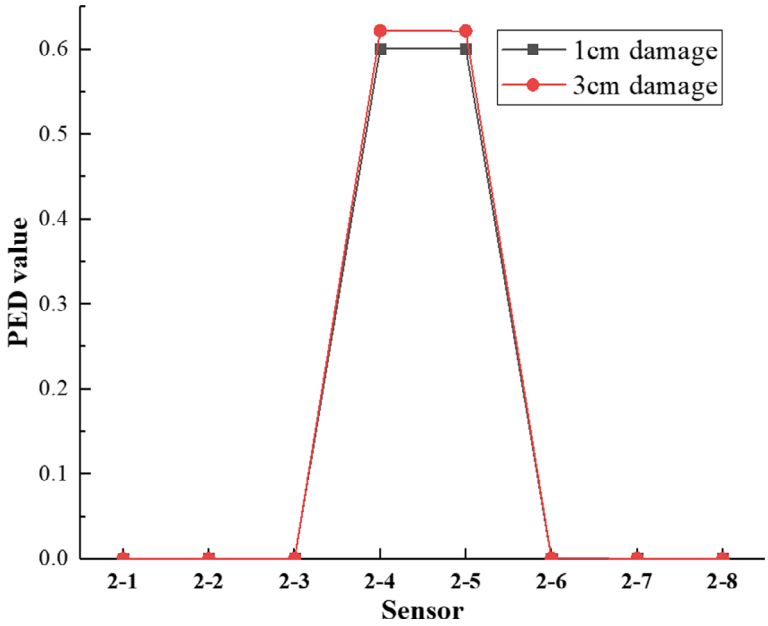


Fig. 9 PED value

References

1. Iqbal, N.V., Subramaniam, K., Shaniba Asmi, P.: Wavelet packet entropy based control of myoelectric prosthesis. Biomed. Pharmacol. J. **11**(1), 375–380 (2018). https://doi.org/10.13005/bpj/1382
2. Zhou, Y.T., Zhuang, Y.Z., Sinha, J.K.: Damage identification for orthotropic steel bridge girder based on wavelet packet permutation entropy. In: Proceedings of Unified 2023 Conference (2024)
3. Zhao, L.Y., Wang, L., Yan, R.Q.: Rolling bearing fault diagnosis based on wavelet packet decomposition and multi-scale permutation entropy. Entropy **17**(9), 6447–6461 (2015). https://doi.org/10.3390/e17096447
4. Qu, J.X., Zhang, Z.S., Wen, J.P., Guo, T., Luo, X., Sun, C., Li, B.: State recognition of the viscoelastic sandwich structure based on the adaptive redundant second-generation wavelet packet transform, permutation entropy and the wavelet support vector machine. Smart Mater. Struct. **23**, 085004 (2014). https://doi.org/10.1016/j.ymssp.2019.06.027
5. Li, D.Y., Cao, M.S., Deng, T.F., Zhang, S.X.: Wavelet packet singular entropy-based method for damage identification in curved continuous girder bridges under seismic excitations. Sensors (Basel) **19**(19), 4272 (2019). https://doi.org/10.3390/s19194272
6. Yang, Q.N., Wang, J.L.: A wavelet based multiscale weighted permutation entropy method for sensor fault feature extraction and identification. J. Sensors 9693651 (2016). https://doi.org/10.1155/2016/9693651
7. Pan, H.X., Pan, M.Z., Xu, X., Ma, B.X.: High-speed automaton fault diagnosis based on wavelet permutation entropy. In: 2018 10th International Conference on Advanced Infocomm Technology (ICAIT). IEEE, Stockholm, Sweden (2018)

Gearbox Anomaly Detection in Wind Turbines Using Classical Machine Learning



Duy Tran, Pedro Lind, and Arvind Keprate

Abstract The primary goal of this paper is to develop a data-driven model that can classify or predict the anomalies of wind turbines' gearboxes by applying machine learning techniques. Owing to the tremendous demand for wind energy, the turbines move offshore and continue growing in size and number. Consequently, this will result in an increment of operational and maintenance costs, which simultaneously poses many challenges for the wind industry. From the literature review, studies highlighted the drivetrain and gearbox as being responsible for major failures and the downtime of wind turbines. This paper examines the most typical malfunctioning modes of the wind turbine gearbox. The importance and evolution of operational condition monitoring are covered. Subsequently, a data-driven approach is considered, and the possibilities of various experimentations are discussed. A machine learning pipeline with classification algorithms is developed to distinguish between healthy and damaged scenarios. The accuracy of the methods applied in this paper is verified with the actual healthy and damaged vibratory signals from a reference wind turbine gearbox. This can open the gate to further investigations and real-time applications for condition monitoring of wind turbines' gearboxes under harsh operational environments.

Keywords Wind turbines · Vibration-based condition monitoring · Data-driven modelling

D. Tran · A. Keprate (✉)
Green Energy Lab, Department of Mechanical, Electrical and Chemical Engineering, Oslo Metropolitan University, Oslo, Norway
e-mail: arvindkeprate@gmail.com

P. Lind
AI Lab, Department of Computer Science, Oslo Metropolitan University, Oslo, Norway

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_19

1 Introduction

1.1 Background

Wind power generation can be considered as an effective form of clean and sustainable energy with various operations both onshore and offshore [1]. Over the last few decades, the wind sector has undergone remarkable expansion. By the end of 2028, the worldwide installed onshore and offshore wind power capacity is forecast to reach 1000 GW, as reported by the International Energy Agency [2]. Statistics indicates that there will be a significant increase in wind power generation in the following decades. This reflects a massive potential in the wind industry, which requires proper exploration and research direction.

Regarding this concern, the research and commercial application of wind with optimized turbines, underpinned by an impartial and trusted voice, is vital for the renewable energy industry. With that mentioned, machine learning (ML) or artificial intelligence (AI) is no longer science fiction; they are woven into daily lives to shape the world's future. In the rapidly evolving landscape, these technologies would empower humans with effective mechanisms to advance the safety and performance of wind turbines (WTs) and inspire this paperwork.

Despite the positive trend, the wind industry is being challenged by premature breakdowns in turbine components. The financial impact of these failures has been amplified by the escalation in turbine dimensions, particularly in major drive train elements, such as main shaft, gearbox, and generator. Enhancing the reliability of the drive train would be imperative for mitigating the costs associated with wind energy production [3].

Two German analyzing programs on WT performance (WMEP and LWK) have indicates that gearboxes are the primary source of downtime per failure [4]. These studies clearly shows the relatively low gearbox failure rates between the two survey methods. However, significant downtime for the gearbox has been recorded, leading to increased maintenance costs and a drop in profit [5].

The reason for the high downtime of gearboxes is due to the complicated repair procedures. In particular, offshore WT maintenance requires specialist equipment such as heavy lifting cranes and support vessels or additional issues from potentially severe weather and rough sea wave conditions [5]. The aim of CM in WTs is to increase the service lifetime expectancy of components whilst reducing O&M cost by instantly detecting defects prior to failures [6, 7]. By monitoring the status of critical rotating components, the CMSs can continuously indicate significant changes in developing failures through a combination of sensors and signal processing equipment [8]. With proper data acquisition and appropriate signal processing, operational anomalies or faults can thus be detected in the early phase. This allows corresponding measures to be planned in time, which can mitigate the failure of components. Subsequently, it would record increased reliability, availability, maintainability, and safety whilst the downtime and O&M costs would be minimized [8].

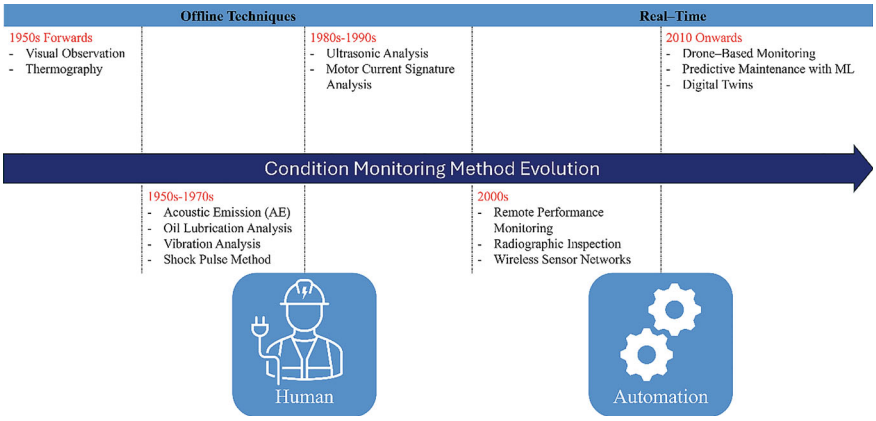


Fig. 1 Condition monitoring method evolution

Figure 1 provides a visual representation of the evolution of CM in WTs. These CM methods, with their unique approaches, have enabled the detection of anomalies or faults in WTs, thereby maintaining high availability or avoiding unplanned downtime [9]. The evolution of CM in WTs has been guided by the nature of technological advancement, a key aspect that this figure illustrates.

In the early phase of the 1950s forwards, inspection techniques such as visual observation or thermography provided informative details on machinery components. Nevertheless, applying these inspection methods during the operation can be challenging as it requires destructive procedures to evaluate individual components [10]. In the second half of the twentieth century, data about the drivetrain with advanced sensors could be transferred to a portable information delivery system and analyzed later in the office. The offline CMSs would require much lower investment but a considerable demand for manual workload from highly experienced personnel or special heavy equipment. Moreover, it can record several faults that may go unnoticed if there are dependencies on specific environmental conditions in the sea [9].

To tackle these technical obstacles for the offline approach, the studies and applications of remote and real-time performance monitoring have exploded from the beginning of the twenty-first century. Bridging the gap between damaged status and sensor signals would boost the development of an approach for failure detection and severe estimation of WTs [11]. From the last decade, reducing the cost and environmental footprint of wind energy has been the primary EU's aim of a zero-emission wind-powered future by 2050 [12]. By applying digitalization, condition-based monitoring embedded with digital twin technology or predictive maintenance with ML models would be advanced diagnostic techniques to reveal the operating machinery faults in the incipient phase before any failures might occur [13]. These intelligent monitoring procedures can be applied either offline or online and categorized into two strategies: model-based or data-driven [13].

In this paper, the primary focus is to develop a data-driven model to classify anomalies from vibration signals of the WT's gearbox. In Sect. 2, materials and methods used in this paper are discussed. Then in Sect. 3, analysis and results are presented and finally, in Sect. 4, suitable conclusions are presented.

2 Materials and Methods

2.1 Data Description and Acquisition

To obtain historical data for the paper, a pre-built set of data was fetched from the NREL [3, 14]. This data package consists of vibratory acceleration signals from eight sensors (AN3 to AN10), as shown in Fig. 2. In particular, the data samples of 40 kHz signals are circumscribed within ten one-minute intervals and documented in '.mat' format for damaged and healthy states. In total, the dataset consists of 20 mat files for eight acceleration sensors. In addition, the collected sample dimension for each sensor is identical, with 2.4 million values stored in one corresponding column. The list of sensor descriptions and locations where the accelerometers are mounted on the outside of the gearbox to sense vibratory acceleration can be found in [14].

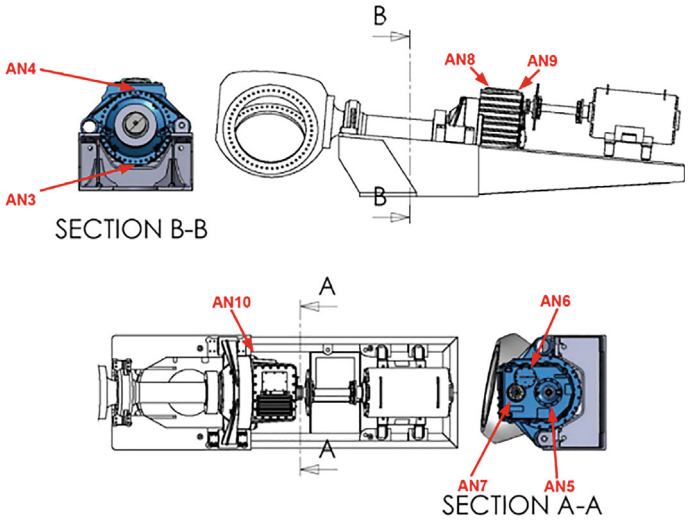


Fig. 2 Vibration sensor locations (Sheng, 2014b)

2.2 Data Preparation

In this section, a detailed review of data processing, setting up prediction rules, and creating the training/testing environments for classification ML pipeline/models will be discussed, and the systematic approach to building up the prediction rules or models will be elaborated with specific specifications.

2.2.1 Data Loading and Processing

The sensor signals are stored in '.mat' file format from the data source. Hence, the 'scipy.io' module is required to handle such a specific file format from MATLAB. Data files from damaged and healthy states will be loaded into the kernel. In total, there are 10 '.mat' files for each operation state that will be handled during the data importation. Each '.mat' file contains specific columns for all vibratory acceleration sensors. Thus, it requires procedures such as splitting columns from the original '.mat' files, categorizing them by sensor types, and concatenating historical data into corresponding data frames (df1 to df10). When loading into the scripting workplace, damaged and healthy datasets are checked for abnormal attributes such as the number of rows, columns, and missing values. Hence, this ensures the superior quality of the datasets utilized throughout this paper.

2.2.2 Exploratory Data Analysis

After the data is imported into the terminal, the correlations among acceleration sensors should be considered to identify the most critical components inside the WT's gearbox during healthy and damaged operational conditions. With the help of the heatmap plotting feature from the Seaborn library, the correlation between accelerometers can be represented in a 2-dimensional (2D) data visualization technique to identify the proper sensors for further analyses. Since, the torque column in the healthy package was missing hence, it was dropped in the damaged data package to match the plotting dimension and comparison purposes. The upper and lower parts of the heatmaps are generally symmetric around the diagonal line. The values of correlation coefficients for those two parts are identical but positioned differently due to the nature of the matrix representation. In practice, correlations close to 1 or -1 indicate a robust linear relationship, while correlations close to 0 indicate a weak or no linear relationship. The intensity of the color reflects the strength of the correlations. To keep the provided information compact and a simplistic interpretation, the upper part is concealed, and the lower triangle is displayed together with the diagonal line.

To interpret and analyze the distribution of collected datasets efficiently, the histogram is subsequently considered and used as a graphical representation. It condenses data series into comprehensive numerical data by grouping them into

logical ranges known as bins, which will be used to identify patterns or characteristics, providing insights into the underlying datasets. Acceleration data from AN3 to AN10 sensors is plotted and visualized using the 'matplotlib' module. In total, there are eight sensors, each with healthy and damaged data, which do not fit into one single plot. Therefore, the histograms are split into two separate plots, and each plot consists of eight subplots arranged in a layout of 2×4 (columns $\times$ rows). The plots will be saved as pictures in the portable network graphics (PNG) format to preserve the original quality, high precision, and details.

At the commencement, the histograms are plotted as one figure containing all subplots from AN3 to AN10 sensors. Consequently, this will separate the histograms between damaged and healthy data. This results in the discontinuation of data comparison. Hence, the Python script with histogram plotting mechanism is modified to rearrange the subplots' layout. After the modification, the first figure contains eight subplots (AN3 to AN6), and it is from AN7 to AN10 for the second one. Each row displays the damaged data on the left side and healthy data on the right side. Approaching this layout enables a direct comparison between the two operational states for each sensor, and the distribution differences can be highlighted with noticeable shifts in the data's central tendency, spread, or shape. The distribution curves are added to the histograms by utilizing the kernel density estimation (KDE) function. These smooth lines help visualize how the data points are distributed across different variable values. In particular, lower peaks or flat areas indicate lower data density, whilst the curve peak suggests regions where data points are more concentrated.

2.3 Machine Learning Pipeline

2.3.1 Dataset Setup

Training and testing datasets should be set up before initiating the ML procedure for the classification models. The code blocks with pre-built Python libraries and data loading mechanisms can be reduplicated from the ones in the data preparation section. The data frame of the healthy state containing 2.4 million samples is vertically stacked over a similar one of the damaged state. With the help of the concatenation function, it creates a column containing 4.8 million samples and enables the ability to label the operational states later in the preparation process. Thus, a function is created to split the column into sliding windows and label the data.

The 2.4 million samples are split into 1000 sliding windows, each containing 2400 samples. In total, there are 2000 sliding windows for the entire column. The first 1000 windows belong to the healthy state and are labeled with '0'. The remaining 1000 windows in the damaged state are labeled with '1'. The labelling step plays a vital role in the supervised ML process, which provides the necessary ground truth for the model/pipeline to learn and make classifications/predictions accurately. In supervised learning, the models rely on labeled samples where each data point (input) is associated with a corresponding target label or output. This labeled dataset is

utilized to train the pipeline/model to learn the underlying patterns and relationships between inputs and outputs [15].

2.3.2 Dataset Split and Scaling

To maintain the continuous flow of the pipeline development, the labeled samples should be split into training and testing sets. The primary aim of splitting data into corresponding subsets is to evaluate the performance of an ML pipeline/model on unseen data. Subsequently, this process can assist in assessing how well the models generalize to new or unseen samples. From the 'scikit-learn' module, the 'train_test_split' method randomly shuffles and splits the dataset into training and testing subsets. The splitting ratio is initially determined as 80%/20% by the 'test_size' parameter and could be tuned further in the result analysis section to obtain the optimal setting. For the ratio of 80/20, the training (X) and testing (y) datasets are then split into 1600 and 400 labeled samples, respectively. Furthermore, the data points in 'X' and corresponding labels in 'y' are shuffled together, which ensures that the training and testing subsets are representative of the overall healthy and damaged concatenated dataset.

The next task is to apply feature scaling, which involves transforming the features of the processed dataset to a similar scale. After the signal analysis section, the recorded data indicates a wide range of values for damaged and healthy datasets. Therefore, feature scaling brings all samples to a similar scale, which improves the performance and stability of various ML algorithms. Moreover, scaling can prevent features with larger scales from dominating those with more minor scales during the pipeline/model training. The scaling method for training the pipeline is selected as 'StandardScaler'.

2.3.3 Data Pipeline Construction

The first step is to define a list of various classification algorithms for comparison, where each potential one represents a different ML model for training and evaluation. In total, eight models were selected for the pipeline development. In the pipeline construction phase, there is a primary loop to iterate over each algorithm with a well-structured framework for training purposes. For the ML pipeline development, the damaged and healthy datasets are scaled inside the loop due to the nature of the 'make_pipeline' method and iteration mechanism. The purpose of this arrangement is to refresh the dataset before initiating a new training once the iteration for the previous classification algorithm is accomplished. The scaling method is selected as 'StandardScaler', and the working principle is described with details in the data split and scaling subsection.

Once the pipeline is constructed, it will then be trained by utilizing the training 'X_train' and 'y_train' datasets obtained from the splitting phase. The 'StandardScaler' scaling method processes the training 'X_train' dataset to perform two statistical

computations on the entire feature during the training procedure. In practice, the average value of the feature across all instances in the 'X_train' dataset will be calculated. Then, the variance measure of how much the feature values spread out from the mean can be determined. After that, the classification algorithm is trained on those scaled datum. The 'fit' method transforms and stores these calculated parameters (mean and variance). This transformation ensures that each feature is centered and scaled, which makes the features comparable among the classification algorithms and keeps facilitating model training.

2.3.4 Prediction with Machine Learning Algorithms

Moving to the evaluation phase, the performances from those classification models with different parameter configurations would be applied to classify/predict the gearbox 's damaged operational state. Once the training process had been completed, the pipeline would make predictions by applying the 'predict' method on the testing 'X_test' dataset, which was prepared and scaled with learned parameters from the training dataset.

As mentioned in the previous subsection, the performance metrics are computed and formulated based on the predicted labels ('y_pred') and the ground truth labels ('y_test'). These two variables will be the required inputs to plot the confusion matrices. This implementation makes the result visualization more attractive and transparent instead of listing the models' performance in numerous digits of a comparison table.

2.3.5 Validation with Arbitrary Acceleration Signals

Moving to the final phase, the ML models are tested with unseen data to verify the classification efficiency and accuracy. The validating data is prepared by randomly selecting healthy and damaged signals from the other 'mat' files. In particular, a similar procedure for data preparation will be implemented as in the ML pipeline development. As the performances of the other classification algorithms are verified by the testing dataset during the training/testing procedure, the two optimized models (Logistic Regression and SVC) will be processed and tested with the validation dataset. From the tuning phase, the hyperparameters produce noticeable effects on the classification performance, which will make it worth the effort to test the model with new ranges of data and analyze the obtained results.

3 Analysis and Discussion

3.1 Data Analysis

Firstly, the results of data analysis are displayed in this section to gain valuable insights into the structure, properties, and characteristics of the healthy and damaged datasets. By investigating these features, statistical properties, and distributions of the datasets, specific patterns can be underlying for further analyses. The data from the first 'mat' file for damaged and healthy states is selected for the detailed analysis and ML pipeline development. The data from the other 'mat' files is used to generate arbitrary data ranges in damaged or healthy states to validate the performance of ML pipelines/models.

3.1.1 EDA of Vibratory Acceleration Sensors

From the successful implementation of data loading and the initial quality check, to study the correlation of the acceleration signals from the prepared datasets, the optimized heatmap in Fig. 3 (left side) depicts that there are relatively high magnitudes between the sensors AN5, AN6, and AN7. These adjacent accelerometers represent damaged vibratory acceleration at the Low-Speed Shaft (LS-SH), Intermediate-Speed Shaft (IMS-SH), and High-Speed Shaft (HS-SH), respectively. After shifting to another rotating stage, a significantly positive correlation between AN5 and AN6 sensors is recorded. On the other hand, it indicates a negative correlation between AN6 and AN7 sensors.

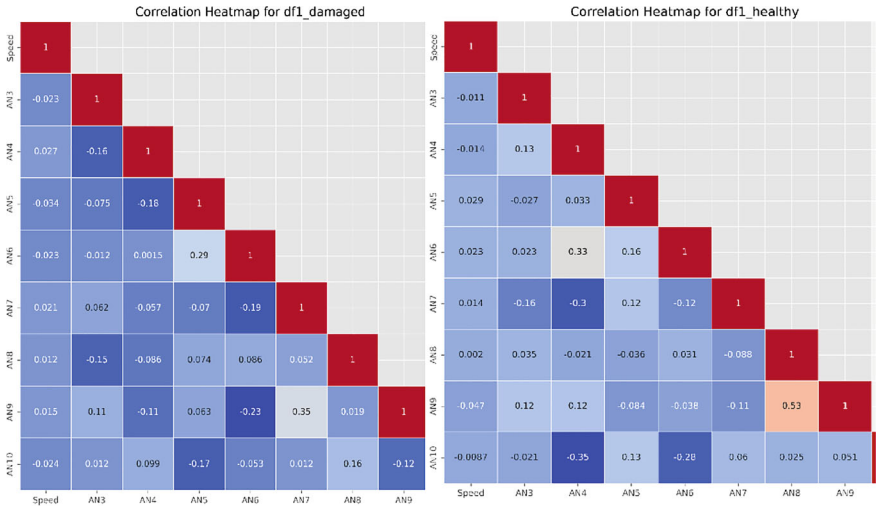


Fig. 3 Correlation of accelerometers for damaged and healthy data

1 (right side) reflects the similar characteristics with the healthy data package for the adjacent AN5, AN6, and AN7 sensors. The correlation between these sensors maintains positive and negative patterns in comparison to the damaged dataset, while it is recorded a slight drop in the magnitude of the relationship between these variables. The obtained statistical values and correlations show a consistent pattern in the interfering operational region of the WT gearbox.

Figure 3 depict the data distribution for all acceleration sensors (AN3 to AN10). The shape of histogram plots is symmetric and balanced around the center, resembling a bell curve, and reflecting a steady distribution pattern. AN5, AN6, and AN7 acceleration sensors display the most normal distribution pattern with small values of standard deviation (σ), which implies that values of vibratory signals are clustered about the mean instead of spreading out in a broader range of data. This supports the identification task of healthy and damaged scenarios. Hence, the AN5, AN6, and AN7 accelerometers are selected to proceed further in anomaly detection of the WT's gearbox.

3.1.2 Vibratory Signal Analysis

The vibratory acceleration signals from the first 'mat' file will be further processed and prepared for the ML development. The signals of the entire data frame are first plotted to get an overview on the correlation between damaged and healthy data. Subsequently, 1000 small ranges, each containing 2400 samples, are derived from the original data frames. These sliding windows are then visualized to provide a deeper analysis of signal attributes and are finally prepared as inputs for the ML pipeline/models.

Figure 4 depicts the entire length of acceleration signals containing 2.4 million samples for AN5, AN6, and AN7 sensors. The healthy signals of the AN5 sensor are relatively small while the signals for both healthy and damaged stages for the AN6 sensor increase. This also matches with the positive correlation between the AN5 and AN6 sensors. Meanwhile, the signals from the AN7 sensor drop slightly in comparison to those from the AN6 sensor. This reflects in the negative correlation between AN6 and AN7 sensors.

Acceleration signals are then split into 2400-sample sliding windows to get a closer look at the data. Figure 6 displays the first three sliding windows for the AN5 sensor. The vibratory signals reflect a stable manner in terms of attributes and magnitude. Hence, this indicates that these 1000 sliding windows can be utilized as the inputs for the ML pipeline/model development. Because of the characteristic consistency, it formulates the training and testing datasets in a well-prepared manner.

The prepared data from the three selected sensors are now ready for the next stage of the ML pipeline development. The obtained training and test results of the ML models will be displayed and discussed in the following section.

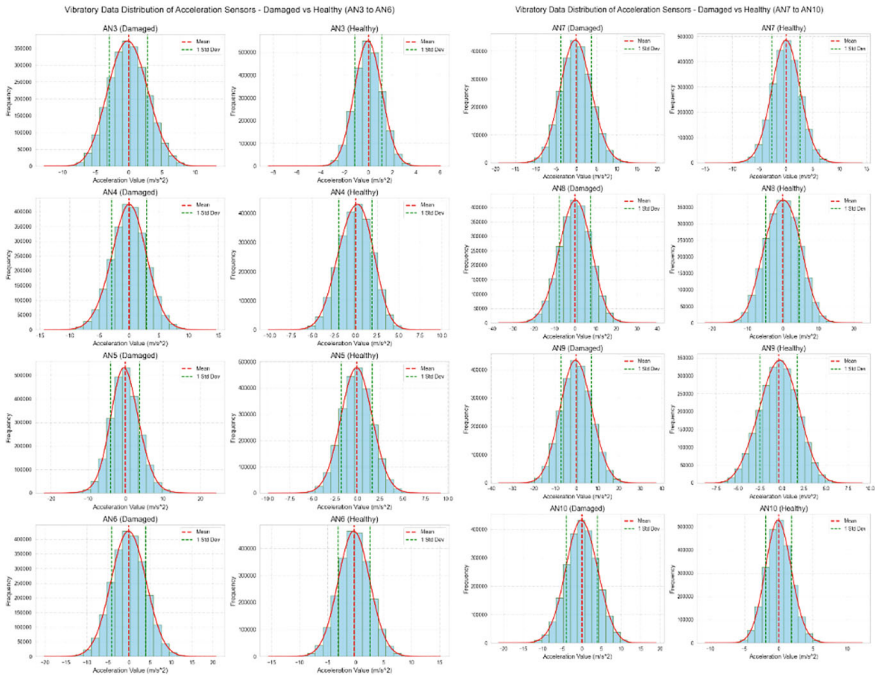


Fig. 4 Vibratory data distribution of acceleration sensors

3.1.3 Machine Learning Pipeline—Classification Algorithm Performance

Once the signal analysis has been successfully conducted, prior to the ML procedure, the healthy and damaged datasets must be prepared for the training and testing tasks. There are in total 2000 sliding windows with 2400 data points for each. To serve the supervised learning models, the dataset is labeled with '0' (the first 1000 windows) and '1' (the rest 1000 windows) for healthy and damaged operational stages respectively. After that, the splitting mechanism arranges the training and testing sets with a ratio of 80/20. The datasets are then scaled with the 'StandardScaler' method to bring all data points to a similar scale, improving the ML pipeline's performance and stability. The ML pipeline then trains on eight classification algorithms in a sequence with a loop mechanism of training and testing datasets. The obtained results for the three acceleration sensors, AN5, AN6, and AN7, are documented as performance metrics, and the model efficiencies are visualized with confusion matrices (Fig. 5).

Figure 6 demonstrates the results achieved by the AN5 sensor's ML pipeline. From this original ML process, classification algorithms such as Random Forest, XGB, LGBM, and SVC_rbf reached the optimal performance. On the contrary, other algorithms, such as Logistic Regression, Decision Tree, SVC_linear, KNeighbors and Dummy, have room for further optimization. The precision and recall scores

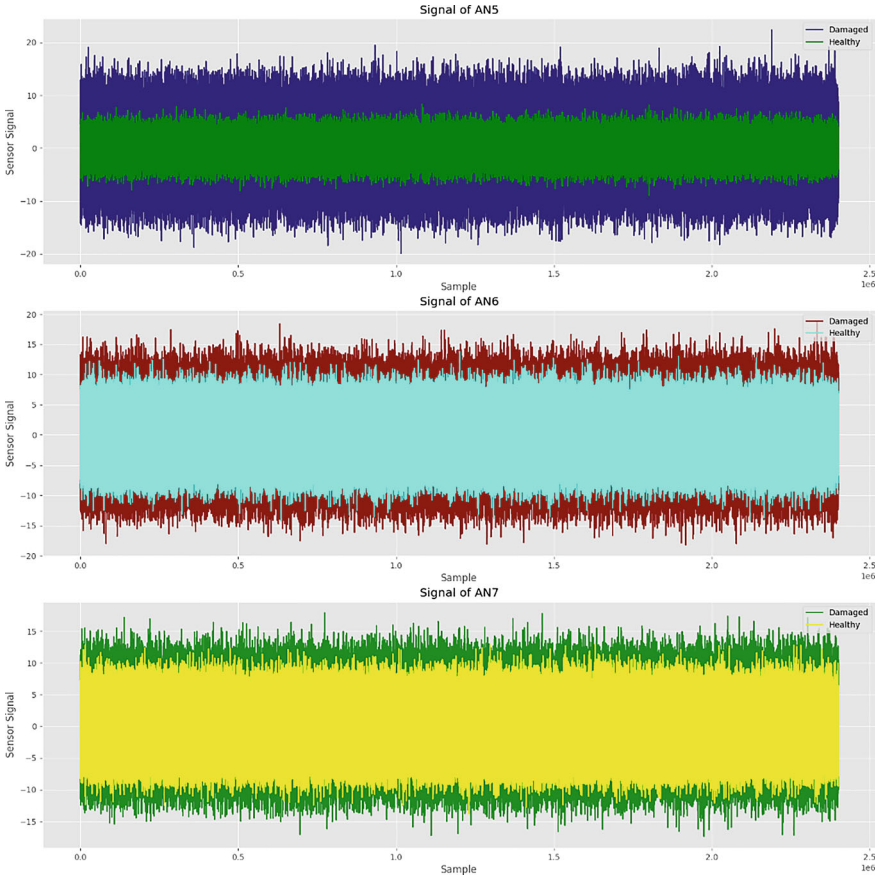


Fig. 5 All sample comparisons of AN5–AN6–AN7 sensors

should ideally approach the value of 1, which indicates the minimization of FP and FN components in the denominator of the Eqs. (4) and (5). The obtained results from Fig. 7 validated this theoretical concept. The algorithms with high values of precision and recall scores can capture all positive instances (FN + TP) in the most accurate manner. In the current analyzing scenario, positive instances refer to the genuinely correct classification/prediction of damaged operational stages. Figure 8 shows several algorithms possessing relatively high values for FN and FP components such as KNeighbors and Dummy.

Moving to the ML pipeline of the AN6 sensor, the obtained results reflect similarities in comparison with the ones from the AN5 sensor. In Fig. 6, the Logistic Regression and SVC_linear models display a significant increase in comparison to the equivalent ones from the AN5 sensor. On the contrary, the KNeighbors model

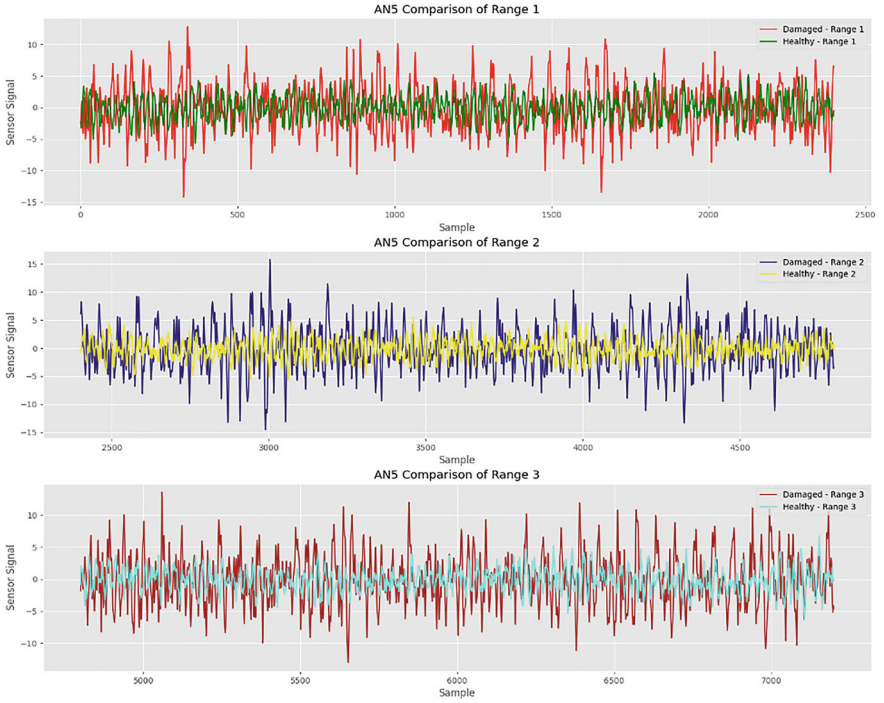


Fig. 6 Filtered range comparisons of AN5 sensor

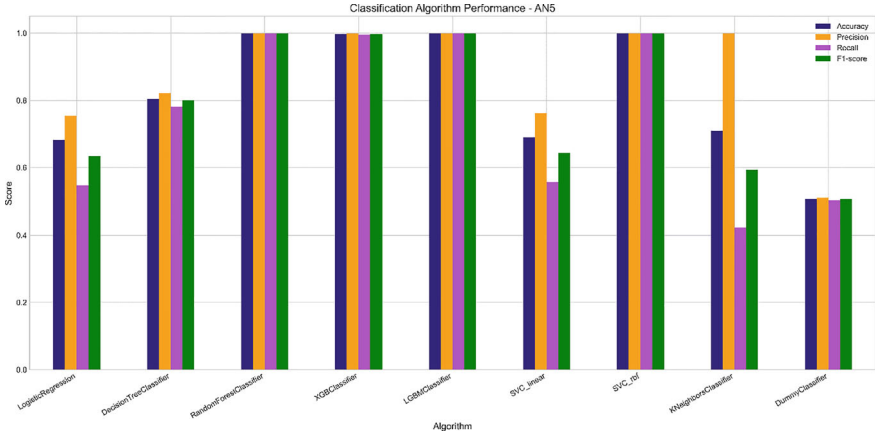


Fig. 7 Classification performance of the ML pipeline for AN5 sensor

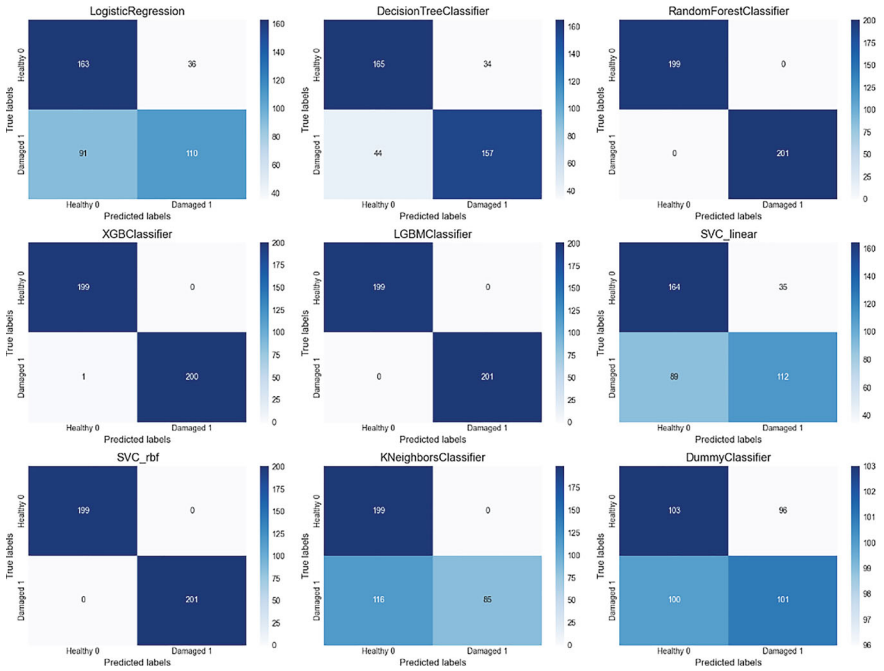


Fig. 8 Confusion matrix of the ML pipeline for AN5 sensor

has a dramatic drop in recall and F1 scores, which leads to the instability of classification/prediction performance for damaged cases. Making appropriate adjustments during the tuning process requires a large amount of effort.

Owing to the low recall score value, in Fig. 8, the misclassification of FN for the KNeighbors model is extremely high with respect to the characteristic of the Eq. (5). Moreover, the classification of FP for the Dummy model also outweighs the TN rate. These obtained results are not preferable for the final aim of the paperwork, which is to detect the actual damaged operational stages (anomalies).

In pursuit of continuity, Fig. 9 indicates that the two classification algorithms (one from Logistic Regression and one from SVC) are the most potential candidates for further tasks. The KNeighbors and Dummny classification algorithms display instability and poor performance in classifying/predicting the damaged cases in Fig. 9. This also represents as undesirable classifications/predictions of FP and FN in Fig. 12 (Figs. 10 and 11).

From the obtained training results of the ML pipeline for the three selected sensors, the KNeighbors and Dummy models return a low level of classification/prediction for actual damaged operation cases (TP). Furthermore, instability also occurs between the recall scores of the KNeighbors algorithm, and the pretty low values of the F1 scores pose challenges in finding the final classification balance between healthy and damaged cases. Among the rest of the algorithms, Logistic Regression and SVC models possess relatively high performance throughout the three selected sensors. As

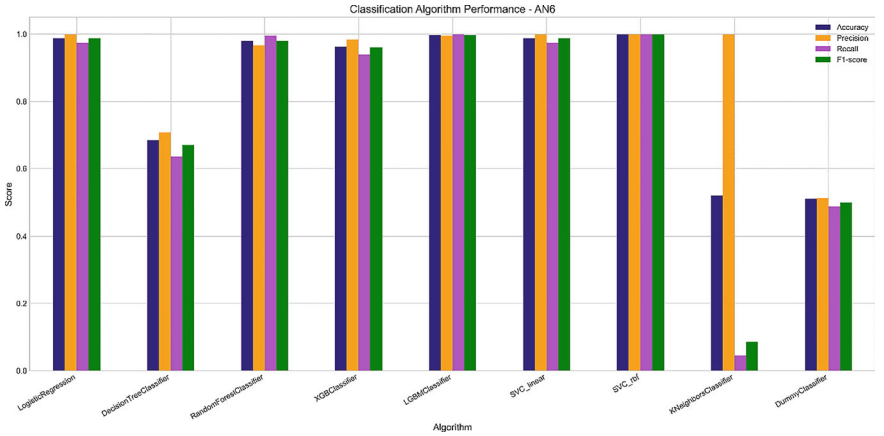


Fig. 9 Classification performance of the ML pipeline for AN6 sensor

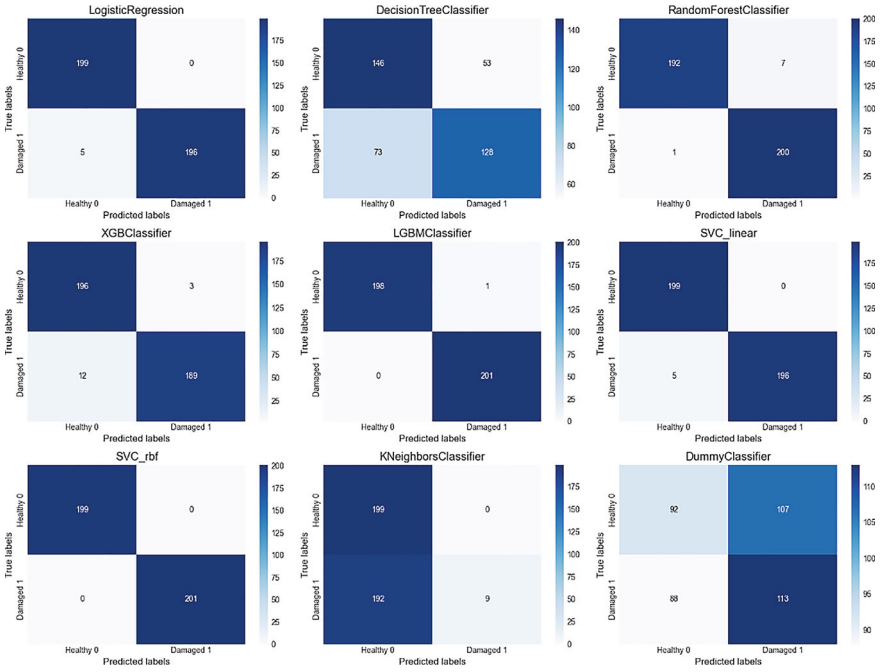


Fig. 10 Confusion matrix of the ML pipeline for AN6 sensor

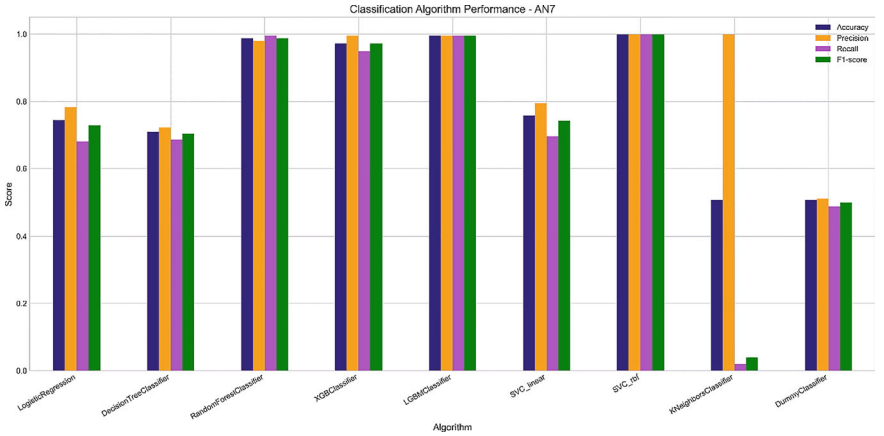


Fig. 11 Classification performance of the ML pipeline for AN7 sensor

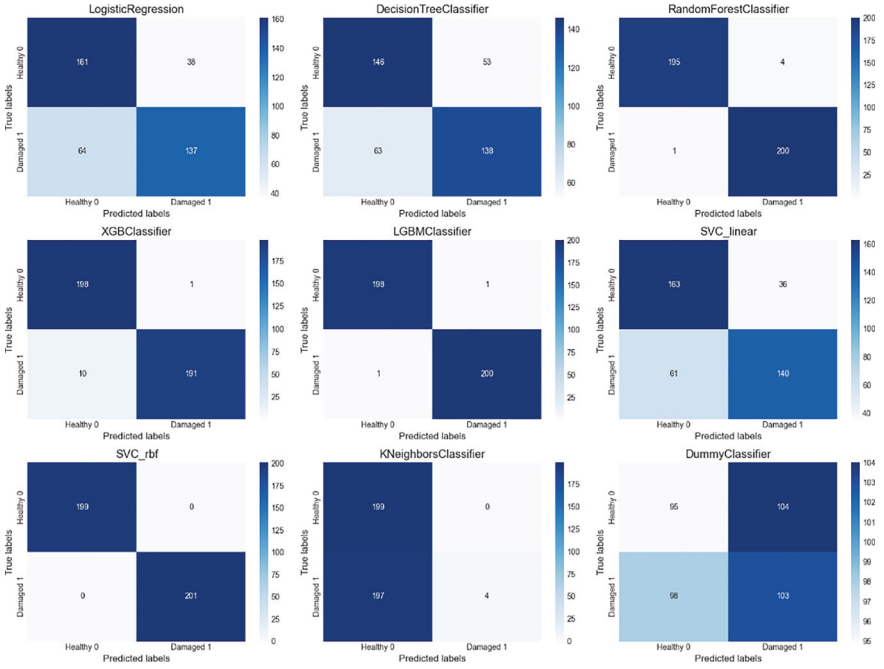


Fig. 12 Confusion matrix of the ML pipeline for AN7 sensor

a result, these two models will be validated further for the classification/prediction of anomalies for the WT gearbox.

3.1.4 Model Validation

In this section, the two classifications (Logistic Regression and SVC) are tested with a prepared validation dataset. Thus, this allows the ability to assess the models' performance on unseen data. The validation dataset is created by accessing the healthy and damaged data from the second 'mat' file. In particular, 400 arbitrary signal ranges, each containing 2400 samples, are formulated. This validation dataset is then utilized to verify the ML models and generate the following confusion matrices in Figs. 13 and 14.

Figure 13 demonstrates the performance of Logistic Regression models for AN5, AN6, and AN7 sensors under exposure to the unseen dataset. From the appropriate tuning conditions and correlations between hyperparameters, the correct classification and prediction of the three Logistic Regression models remain the same efficiency and precision in comparison to the testing datasets during the ML development phase. Furthermore, it is recorded that there are marginal improvements in the correct number of TP (damaged cases). Sharing reactions similar to those of the Logistic Regression models, the outcomes for the optimized SVC models also reflect a high classification/prediction efficiency level with the untrained dataset, as shown

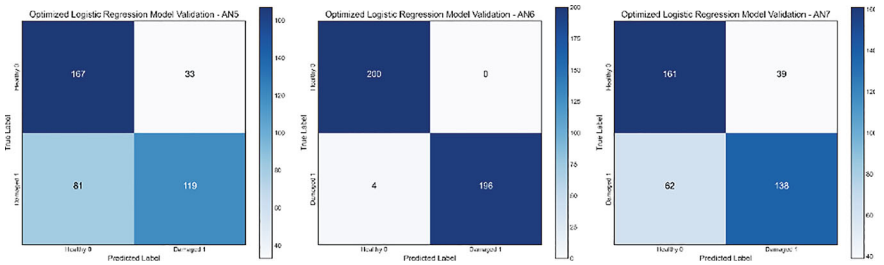


Fig. 13 Logistic Regression model validation for AN5, AN6 and AN7 sensors

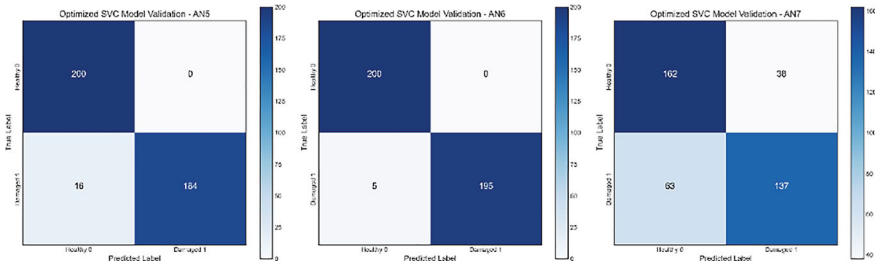


Fig. 14 SVC model validation for AN5, AN6 and AN7 sensors

in Fig. 14. This promises considerable potential for further development of the ML models and the real-time applications on offshore WTs.

4 Conclusion

This section will reevaluate the main results obtained from implementing this paper. This presented report studied the theory, algorithm constructions, training, testing, and validating methodologies for the ML classification pipeline/models. In total, 4.8 million samples of damaged and healthy datasets, collected from the 3-stage WT gearbox at the NREL, are considered to perform the EDA and processed for the development of ML pipeline/models. From the data analysis phase, there are solid and direct correlations between AN5, AN6, and AN7 accelerometers. Furthermore, due to the limited time of the paper, those three accelerometers were selected to build up the ML pipeline/models. When there are changes in rotation speed at the interfering operational region, the magnitude of acceleration signals adapts accordingly to the corresponding sensors.

Due to the realistic demand of applying the data-driven approach to CM for the WT gearbox, an ML pipeline is developed with eight classification algorithms to classify/predict or detect anomalies (damaged cases). From the original data, the samples are split into 1000 ranges for each operational state. For serving the supervised learning models, these ranges, containing 2400 data points each, are then designated as '0' and '1' for healthy and damaged cases, respectively. From the obtained results, the classification algorithms, such as Random Forest, XGB, LGBM, and SVC_rbf, achieved optimal performance. Subsequently, Logistic Regression and SVC (with different kernel types) models are selected among the other algorithms for validation purposes. By validating with the arbitrary acceleration signals, the two models confirm the classification/prediction reliability under exposure to unseen data. Hence, the data-driven models in this paper can be used for monitoring the anomalies (damaged cases) of the WT gearbox.

References

1. Joe, C., David, C., Heather, L., Brian, H., Foley Aoife, M.: Managing Future Offshore Wind Power Variability in a European Supergrid (2013). https://doi.org/10.1109/EEEIC.2013.6549582
2. International Energy Agency (IEA): Renewables 2023 Analysis and Forecasts to 2028 (2023). https://www.iea.org/reports/renewables-2023
3. Sheng, S.: Wind Turbine Gearbox Vibration Condition Monitoring Benchmarking Datasets (2014b). www.nrel.gov/wind
4. Faulstich, S.: Windenergie Report Deutschland 2008 (2008). www.iset.uni-kassel.de
5. Crabtree, C.: Survey of Commercially Available Condition Monitoring Systems for Wind Turbines (2010). https://www.researchgate.net/publication/242606016

6. Salameh, J.P., Cauet, S., Etien, E., Sakout, A., Rambault, L.: Gearbox condition monitoring in wind turbines: a review. In: Mechanical Systems and Signal Processing, vol. 111, pp. 251–264. Academic Press (2018). https://doi.org/10.1016/j.ymssp.2018.03.052
7. Nie, M., Wang, L.: Review of condition monitoring and fault diagnosis technologies for wind turbine gearbox. Proc. CIRP **11**, 287–290 (2013). https://doi.org/10.1016/j.procir.2013.07.018
8. García Márquez, F.P., Tobias, A.M., Pinar Pérez, J.M., Papaelias, M.: Condition monitoring of wind turbines: techniques and methods. Renew. Energy **46**, 169–178 (2012). https://doi.org/10.1016/j.renene.2012.03.003
9. Tomasz, B., Haddar, M., Bartelmus, W., Chaari, F., Zimroz, R.: Vibration-Based Condition Monitoring of Wind Turbines, vol. 14 (2019). http://www.springer.com/series/13418
10. Papaelias, M., Márquez, F.P.G.: Wind Turbine Inspection and Condition Monitoring. In Non-Destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets, pp. 19–29. Elsevier (2020). https://doi.org/10.1016/B978-0-08-101094-5.00002-2
11. Sørensen, B.F., Lading, L., Mcgugan, P., Debel, M., Kristensen, C.P., Hansen, G.C.: Fundamentals for Remote Structural Health Monitoring of Wind Turbine Blades—A Pre-Project (2002). https://orbit.dtu.dk/en/publications/fundamentals-for-remote-structural-health-monitoring-of-wind-turb-5
12. Stadtmann, F., Rasheed, A., Kvamsdal, T., Johannessen, K.A., San, O., Kolle, K., Tande, J.O., Barstad, I., Benhamou, A., Brathaug, T., Christiansen, T., Firle, A.L., Fjeldly, A., Froyd, L., Gleim, A., Hoiberget, A., Meissner, C., Nygard, G., Olsen, J., Skogas, J.O., et al.: Digital twins in wind energy: emerging technologies and industry-informed future directions. IEEE Access **11**, 110762–110795 (2023). https://doi.org/10.1109/ACCESS.2023.3321320
13. Surucu, O., Gadsden, S.A., Yawney, J.: Condition monitoring using machine learning: a review of theory, applications, and recent advances. In: Expert Systems with Applications, vol. 221. Elsevier Ltd (2023). https://doi.org/10.1016/j.eswa.2023.119738
14. OEDI: Wind Turbine Gearbox Condition Monitoring Vibration Analysis Benchmarking Datasets (2014). https://doi.org/10.25984/1844194
15. Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., Nenadic, G.: Machine learning methods for wind turbine condition monitoring: a review. In: Renewable Energy, vol. 133, pp. 620–635. Elsevier Ltd (2019). https://doi.org/10.1016/j.renene.2018.10.047

Enhanced Gear Fault Diagnosis Using Optimized Spectrogram-Based CNNs for Multiple and Compound Fault Detection

Priyom Goswami, Vedant Agrawal, Uday Chaudhary, and Rajiv Nandan Rai

Abstract Industrial gearboxes are critical components in various mechanical systems, but their reliability can be severely affected by faults, leading to costly downtimes and safety hazards. Traditional fault detection methods, such as manual inspections and vibration analysis, often fall short in identifying faults in their early stages and are sensitive to changes in operating conditions. This paper presents a novel approach utilizing Convolutional Neural Networks (CNNs), specifically a modified VGG16 architecture, for the detection of multiple and compound faults in industrial gearboxes. The proposed method leverages Short-Time Fourier Transform (STFT) spectrograms generated from vibration signals, allowing the model to self-learn fault features. Data augmentation through a striding technique enhances the training dataset, leading to improved fault classification accuracy. Experimental results demonstrate that the modified VGG16 model achieves a classification accuracy of 99.52%, significantly outperforming traditional methods. The VGG 16 Model is compared with other CNN models and it was found that VGG16 performs better in all scenarios. This study provides a robust, scalable solution for real-world gearbox fault diagnosis under varying operating conditions.

Keywords Gear fault detection · Spectrograms · CNN · Multiple gear faults · VGG16

P. Goswami (✉) · R. N. Rai
Subir Chowdhury School of Quality and Reliability, Indian Institute of Technology Kharagpur, Kharagpur, India
e-mail: priyom.mechengg@gmail.com

V. Agrawal · U. Chaudhary
Mechanical Engineering Department, Indian Institute of Technology Kharagpur, Kharagpur, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_20

1 Introduction

Industrial gearboxes are vital components in various mechanical applications, playing a crucial role in power transmission and torque conversion. However, their reliability can be significantly compromised by faults, leading to costly downtimes, safety hazards, and compromised operational integrity [1, 2]. Traditional fault detection methods, which often rely on manual inspections or vibration analysis [3], are time-consuming and prone to human error, and may not identify faults in their early stages. Statistical features acquired based on the collected vibration signals are not robust to changes in the operating settings of the gearbox until a major change in vibration signature is evident due to increase in severity of the faults in gears [4].

Recent advancements in sensor technologies and machine learning algorithms have paved the way for more efficient and accurate fault detection methodologies. With increase in computational performance of computers, Convolutional Neural Networks (CNNs) act as a potent tool for vision based fault detection and analysis. CNNs are widely used these days to have emerged for efficiently detecting and categorizing faults in gear systems because of they offer several advantages over conventional approaches, including automatic feature extraction, robustness to variations in operating conditions, and end-to-end learning capabilities [5].

Early fault detection methods primarily utilized time-domain analysis of vibration signals for the classification of faults [6, 7]. While computationally efficient, these methods often lack the sensitivity needed to detect incipient faults [8]. Frequency-domain analysis techniques like the Fast Fourier Transform (FFT) have been widely used to identify characteristic fault frequencies in gear systems [5, 9]. Advanced signal processing techniques such as Wavelet Transform and Empirical Mode Decomposition (EMD) provide multi-resolution analysis and adaptive decomposition of complex signals, respectively [10]. Authors in [11] used a multilayer perceptron model to classify gear chip and wear from the Shannon entropy ratio parameter extracted from the continuous wavelet transform of the time domain vibration signal. Additionally, the advent of machine learning has revolutionized gear fault diagnosis. Techniques such as K-Nearest Neighbours and Artificial Neural Networks (ANN) have been successfully applied to classify gear faults based on extracted features [12]. However, these methods often require manual feature extraction, which can be time-consuming and may not capture all relevant information [13].

Deep learning techniques, particularly CNNs, have shown significant promise in gear fault diagnosis. CNNs and machine learning algorithms are efficient in their pattern recognition thus poses the ability for multi-class classification along with fault interactions. Authors in [14] used CNNs to classify gearbox faults, and the learning features of the CNN filters are visualized to understand the physical fault diagnosis phenomena. The input to the CNN can be both either one-dimensional for just the time series data or two-dimensional for both the time as well as frequency data. The performance of CNN based fault diagnosis algorithms depends on the quality of input images and the CNN architecture. The conversion of raw vibration

data into images can be done using short time Fourier transform (STFT) or wavelet transformation (WT).

The identification of fault severity is critical in determining the maintenance, which is significant to the safety of mechanical equipments. Gai et al. [15] proposed a parameter-optimized deep belief network (DBN) using the sparrow search algorithm for gear fault severity detection. Their method achieved over 96% detection accuracy by optimizing DBN parameters, demonstrating improved sensitivity and accuracy. Moreover, when a gear or bearing malfunctions, it often triggers faults in other interconnected components, leading to the development of compound faults [16]. Hence, there is a need to concurrently identify single fault as well as detect compound faults (interacting faults) in gearbox systems. Some researchers approach different severity levels of the same fault type as distinct faults or categorize compound faults as entirely new fault types. Consequently, they apply single-task diagnosis methods even when dealing with multi-task fault diagnosis scenarios. For instance, [17] proposed a CNN model integrating multi-sensor data for diagnosing compound faults in planetary gearboxes. Similarly, [18] introduced a fault diagnosis approach based on smooth pseudo-Wigner-Ville distribution and deep learning, enabling effective identification of fault types and degradation levels in rolling bearings. Although these methods demonstrate high diagnostic accuracy in specific scenarios, they may not consistently align with real-world conditions. Utilizing a single-task diagnosis method for multi-task fault diagnosis can lead to confusion among multiple diagnostic tasks and ineffective feature extraction from tasks. Therefore, a need for a multi-task classification model which is able to correctly detect the fault type as well as fault interactions is of paramount importance for real world applications.

Building on these advancements, this paper presents a novel approach that leverages CNNs for multiple fault detection along with compound faults in industrial gearboxes. The key contributions of the paper are as follows:

- Developing a CNN model that can auto learn fault features of gearbox based on the generated spectrogram images and classify multiple faults as well as interactive faults in gears.
- We determine the optimal segmentation length, overlap level, and striding parameters for generating spectrogram images. This optimization enhances the representation of vibration signals in the time–frequency domain, crucial for effective feature extraction by the CNN model.
- To address the challenge of limited data, we implement a data augmentation technique using a striding method, which increases the number of spectrogram images by nine to ten times. This augmentation significantly improves the training process and reduces the risk of overfitting, leading to a more robust and generalized model.

We utilize STFT spectrograms generated from vibration signals as input to a modified VGG16 CNN architecture. The VGG16 model is selected for its depth and proven success in various image classification tasks. Unlike shallower architectures, VGG16's 16-layer depth allows it to capture complex spatial hierarchies, which is crucial for identifying subtle and compound faults in gearboxes. Its architecture is

effective in feature extraction with minimal manual intervention, allowing it to learn fault-related features directly from spectrograms. Additionally, modifying VGG16 enabled the integration of dropout and dense layers for more precise fault classification in noisy environments. The following sections provide a detailed overview of the research methodology used in this study. Section 2 describes the experimental setup in detail. Section 3 outlines the comprehensive methodology adopted for the investigation. Section 4 presents the results and discusses their significance. Finally, Sect. 5 concludes the paper by summarizing the key findings and their implications.

2 Experimental Setup

Figure 1 depicts the gearbox test setup that is designed and installed at Indian Institute of Technology Kharagpur. The experimental apparatus comprises a 15 kW 3-phase induction motor capable of generating a maximum output of 20 hp. The motor operates at a constant speed of 3000 rpm, so, to vary the speed, required for designed experiments, a Variable Frequency Drive (VFD) is integrated. VFD allows speed adjustments ranging from 0 to 50 Hz with a resolution of 0.1 Hz. The motor is coupled to a 2-stage, 4-speed industrial gearbox with spur gear configuration. The gears are manufactured using 40C8 steel and have undergone case hardening. Furthermore, the gearbox is connected to a dynamometer with a maximum braking torque capacity of 125 Nm, which is used to load the gears.

The accelerometers employed in this setup are tri-axial Integrated Electronic Piezoelectric (IEPE) accelerometers capable of logging acceleration range from 0 to 50 g. The accelerometers are positioned at the top of the gearbox (nearest to the gear meshing zone) and also at the inlet and outlet bearing housing of the gearbox. Data

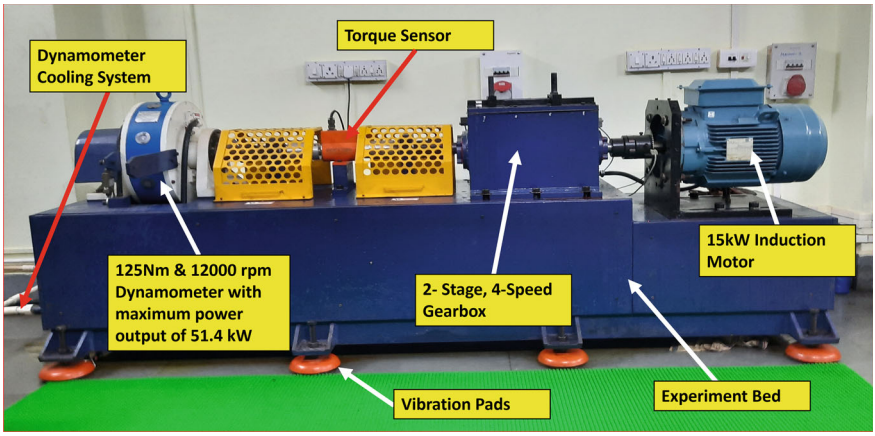


Fig. 1 The experimental setup

Table 1 Experimental details

Gear conditions	Speed (rpm)	Loads
No fault	500	0 Nm (no load)
Pitting fault	1000	20 Nm
Broken tooth	1500	
Pitting and broken tooth		

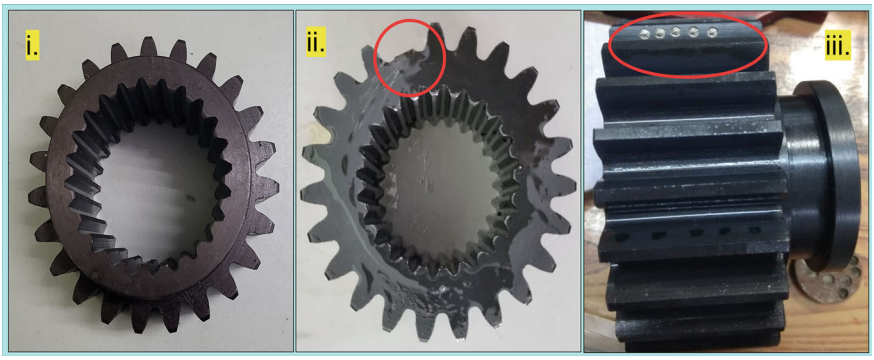


Fig. 2 Gear fauls (i) no fault (ii) broken tooth and (iii) pitting fault

acquisition is performed at a sampling rate of 20 kHz using a National Instruments (NI-9252) DAQ system and the effective vibration sensor is selected based on [4].

The experiment was conducted for three different levels of speeds and two different loading levels to include various operating conditions of a gearbox. This is to develop a model that is redundant to gearbox operating under various conditions. The details of the experiments being conducted is provided in Table 1. So, for every condition of gears there are six experimental runs with three speeds and two loading conditions. The data is collected for 10 s, which accounts for 0.2 million data point at once, as the accelerometers are sampled at 20 kHz.

The faults in gears are artificially seeded using electro discharge machining and electro chemical machining for broken tooth and pitting tooth respectively. The no fault gears and the seeded faults gears can be seen in Fig. 2.

The swapping of the gears in the gearbox are carried by removing a non-faulty gear with the faulty gear at the locations shown in Fig. 3.

3 Methodology

This research employs a structured approach to gear fault diagnosis, which involves data acquisition, image generation, feature extraction, and classification using Convolutional Neural Networks (CNNs) based on VGG16 architecture. The proposed

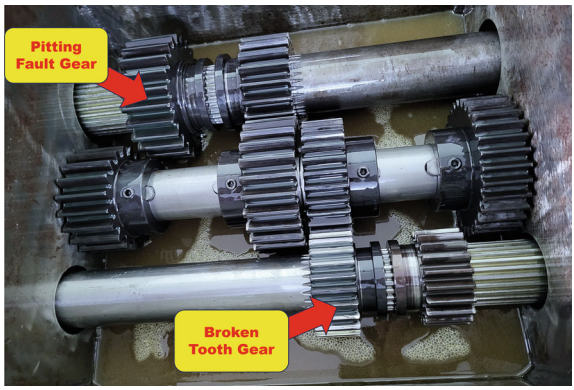


Fig. 3 Gear swapping locations for introducing seeded fault gears for experiments

methodology is shown in Fig. 4. Initiates with vibration data acquisition using the sensors positioned at the strategic locations of the multistage gearbox under varying operational conditions. Subsequent to data acquisition, the collected vibration signals undergo preprocessing that include the application of windowing techniques, which divide the vibration data into smaller segments. With limited size of the dataset, data augmentation technique is employed to enhance the training process.

In particular, striding technique is applied to the vibration data, generating multiple overlapping segments. This approach increases the number of spectrograms images by almost nine to ten folds, significantly improving the training data and mitigating the risk of overfitting.

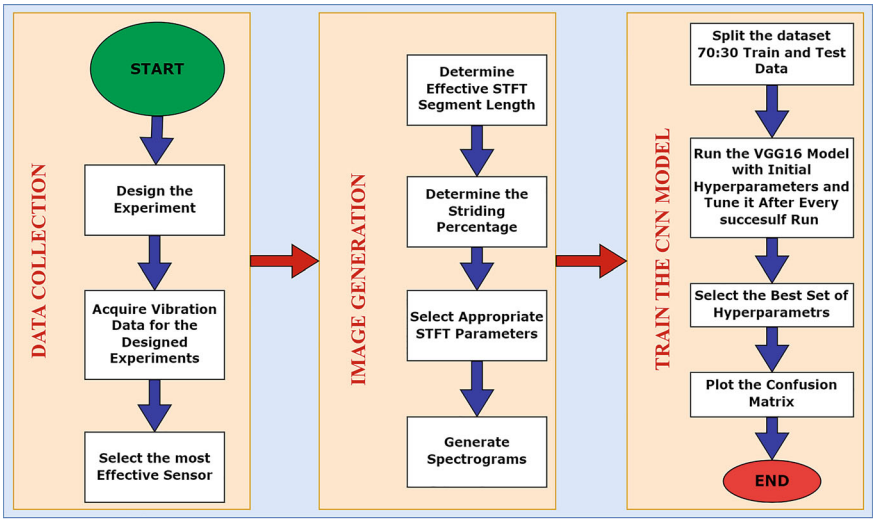


Fig. 4 Proposed methodology for gear fault detection

The Short-Time Fourier Transform (STFT) is then applied to the pre-processed vibration signal segments to extract time–frequency representations (TFR) as spectrograms as shown in Fig. 5. More attention is given to the selection of STFT parameters, such as window length and overlap, in order to optimize the trade-off between time and frequency resolution. The generated spectrograms are then fed as input for the VGG16 CNN architecture. The final fully connected layers of the CNN are replaced with layers specifically designed for multi-fault classification. We modified the VGG16 to incorporate a new dense layer, followed by a Dropout layer followed by a four-class final output layer with SoftMax activation. Therefore, the modified VGG16 now looks like as shown in Fig. 6. The modified VGG16 model is then trained using the augmented dataset. Model training is performed using the Adam optimizer, with hyperparameters such as learning rate, batch size, and number of epochs fine-tuned to maximize model performance. To ensure the model's generalizability, the dataset is divided into training, validation, and test sets, with careful attention paid to the prevention of overfitting during training.

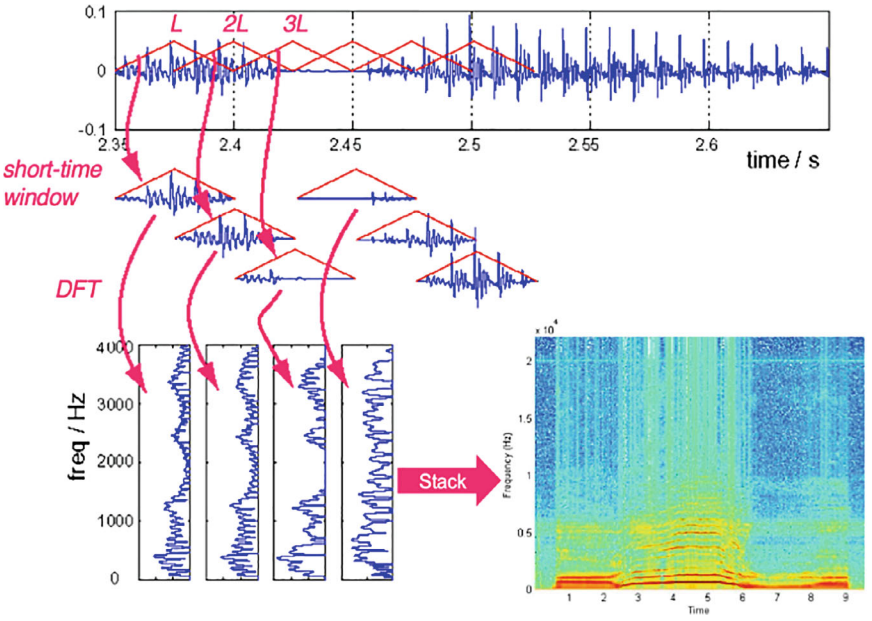


Fig. 5 Process of generating STFT spectrogram [19]

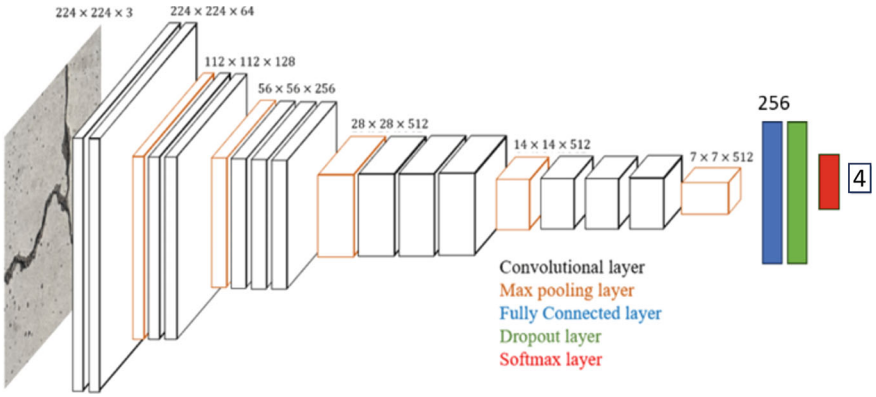


Fig. 6 The modified VGG16 model

4 Results and Discussions

As described in Sect. 3 (Methodology), vibration signals were acquired for four different gear fault classes: no fault, broken tooth, pitting, and a combination of pitting and broken teeth. The data was collected at three different operating speeds (500, 1000, and 1500 rpm) and under two load levels (0 and 20 Nm). Figure 7 shows the time domain (TD) and frequency domain (FD) plots for all fault conditions at a 20 Nm load and 1500 rpm. It is evident from Fig. 7 that no distinct fault characteristics are visible. This could be due to noise in the vibration signals, indicating a need for signal preprocessing. However, signal denoising is a challenging task that involves selecting appropriate parameters for denoising algorithms (such as Wavelet, Minimum Entropy Deconvolution, etc.) to effectively declutter the signals while retaining essential information [20]. To address this issue, we use Convolutional Neural Networks (CNNs) for fault detection, allowing the model to self-learn fault features from the generated spectrograms of various fault and no-fault classes. Additionally, considering diverse operating conditions during data collection helps build a robust deep learning model capable of detecting gear faults under varying operating conditions.

Spectrograms are generated for each class of faults such that 2000 data points are initially being considered based on [21]. However, with 2000 data points per spectrogram with 10 s data having 0.2 million data points, only 100 images could be generated for a given operating (speed and load) condition. So, with three speed and two loading conditions, 600 spectrogram images could be generated at once for a given fault class. With four fault classes only 2400 images could be generated and for such a smaller number of spectrogram images for training the VGG16 model, the training efficiency of the model was found to be very low. To generate more STFT images, we employed a striding technique to divide the vibration signals into smaller fragments. After selecting the number of data points for the initial image, we

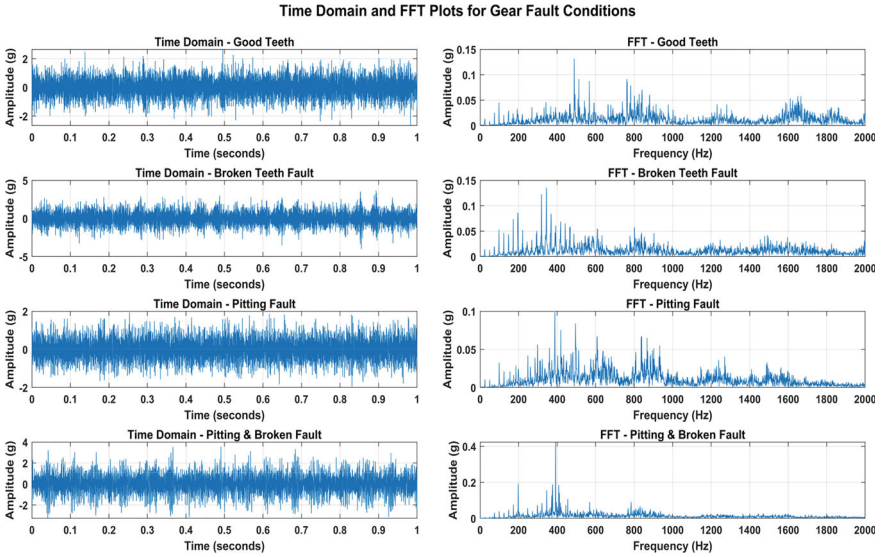


Fig. 7 TD and FD plot for all faults at 1500 rpm and 20 Nm

overlapped a certain percentage of these data points with the subsequent window to create the second image, and so on. The process of striding can be seen in Fig. 8.

By employing the striding technique, we were able to substantially increase the number of images generated for each fault class. While striding enhances data volume by creating overlapping segments, it can also introduce redundant information, which

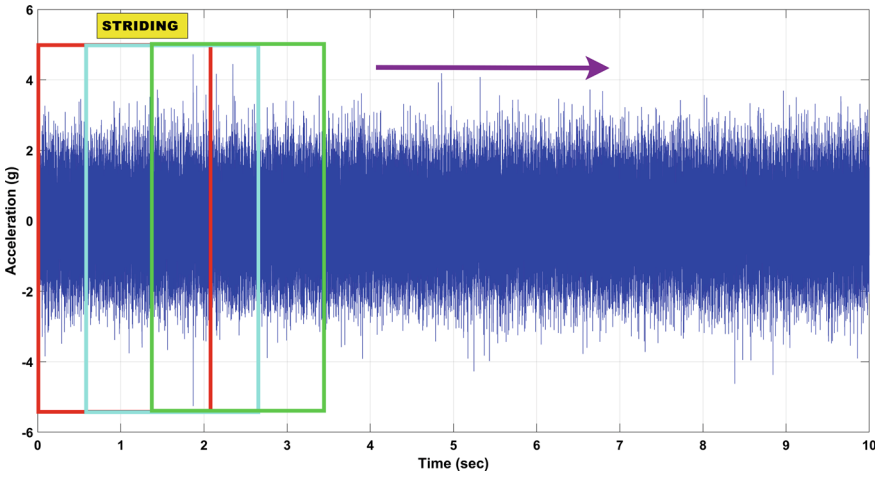


Fig. 8 The process of striding for generating a greater number of spectrograms with the same vibration signal

may result in model overfitting if not managed properly. Additionally, the technique could limit generalizability, as the model might focus on specific features unique to the augmented data. However, careful parameter tuning helps mitigate overfitting and computational strain, balancing the benefits of augmented data with model performance. This can be seen when we used the initial 2000 data points to generate TFR maps, the resulting spectrograms didn't clearly capture the subtle fault features that we needed to extract using the VGG16 model as a result we could only get an accuracy of 40%. From this instance it's well established that, in order to create data segments for generating STFT spectrogram images, the length of the signal (in terms of data points) has a significant impact on the model training and accurate prediction. We discovered that the model's accuracy improves only after a certain minimum window length is reached. This is because a single STFT image needs to capture at least one full rotation of the gear to effectively identify the relevant fault features. In the scenario with the lowest revolution speed of 500 RPM, we calculated 8.2 rotations per second. With a data acquisition rate of 20,000 Hz, this translated to 2440 data points per rotation. Consequently, increasing the window size for data segmentation beyond 2440 data points resulted in a significant improvement in training accuracy. Similar analysis was done for all three speeds (500, 1000 and 1500 rpm) and it was found that 2500 data points per segment is used with 75% overlap. This suggests that capturing at least one complete gear rotation within each STFT image is crucial for effective feature extraction and fault diagnosis. Incorporating striding and adjusting data points to minimum one rotation the number of images generated per class is tabulated in Table 2.

Since the STFT images will be fed to the VGG16 model, an input image size of 224×224 pixels must be maintained. The window size has been adjusted to ensure that at least one rotation is captured in each STFT image. Additionally, a resolution of 2 pixels per time window is desired, resulting in 112 time windows across the image width. To achieve this with a 50% overlap, the window size is set to 224. Similarly, the number of frequency bins is also set to 224, ensuring a 1:1 mapping between frequency bins and image pixels. Based on all the modifications the STFT images now generated now looks different from what it used to be without these modification and same can be seen in Fig. 9.

With all the modification the STFT based spectrograms are fed into the modified VGG16 CNN network. The image dataset was randomly split into 70:30 train

Table 2 STFT images generated before and after striding

Class type	No. of STFT images before striding	No. of STFT images after striding
Good	600	5625
Full Broken Teeth	600	5858
Pitting	600	5820
Full broken teeth and pitting	600	5881

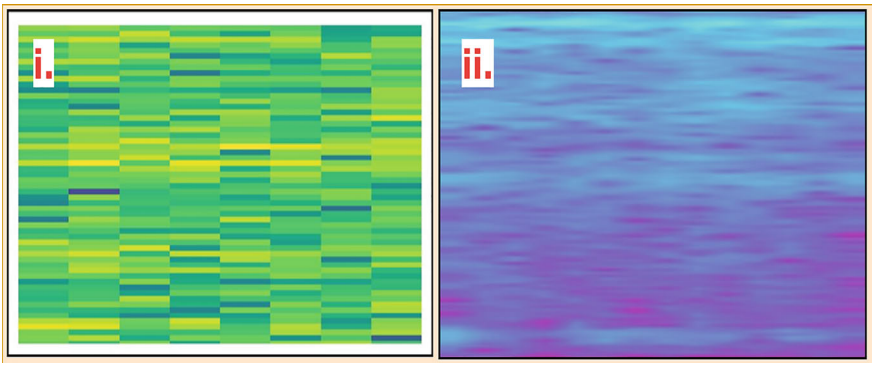


Fig. 9 Spectrogram (i) before modification and (ii) after modifications

Table 3 Hyperparameters of the selected VGG16 model

Hyperparameters	Optimal value
Learning rate	$1e^{-4}$
Batch size	64
Number of neurons in the new dense layer	256
Number of trainable layers	Last 2
Dropout rate	0.6

and test data for the VGG16 network. Through extensive experimentation to identify the optimal hyperparameters for VGG16, the most effective configuration was determined, as detailed in Table 3.

Based on the selected hyperparameters, the VGG16 model was run on an Intel i7 12th Generation CPU with NVDIA GTX 3050Ti 4 GB GPU and 32 GB DDR5 4800 MHz RAM. The efficiency of the model with these modification improved efficiency from 40 to 99.52%. The confusion matrix for training and testing can be seen in Figs. 10 and 11 respectively. Where training accuracy of 99.60% and testing accuracy of 99.52% is achieved using the model.

The augmented dataset is tested against other established CNN models like InceptionV3, MobileNet and a lasted computationally tractable CNN model developed by [22]. The details of which can be found in Table 4. It can be noted that using data augmentation technique helps in generating feature rich images for each fault class, leading to better accuracy even in less depth and computationally tractable CNN models like [22].

The proposed VGG16-based approach is feasible for real-time applications with optimized hardware, such as GPUs, which can handle the intensive computations required for processing spectrograms. The relatively high classification accuracy and robustness to noise suggest it could be deployed in real-time monitoring systems, though latency could vary depending on hardware constraints. While latency concerns might arise in standard CPU environments, further optimizations, such as

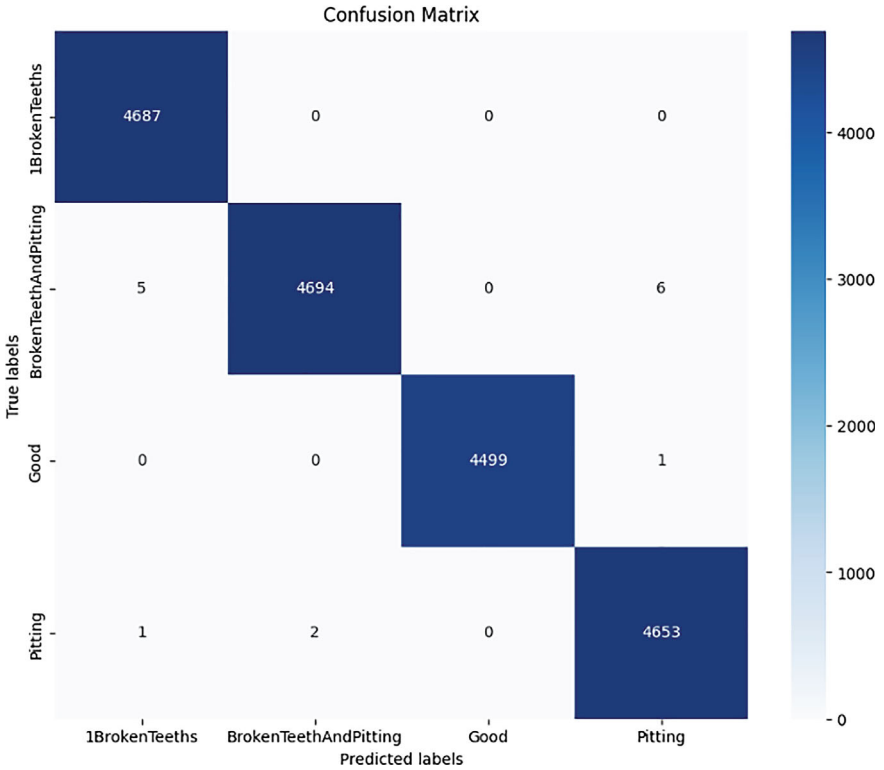


Fig. 10 Training accuracy of the CNN model

model pruning or quantization, can reduce the processing load. However, if hardware limitation arises due to no GPU availability, Goswami and Rai CNN model could be used. As it is a shallow model and can be effectively used for training in CPU environment for fault diagnosis with similar high classification accuracy of 99.5% efficiency. The confusion matrix of InceptionV3, MobileNet Fig. 12 is shown in and Goswami and Rai is shown in Fig. 13.

5 Conclusions

This paper introduces a deep learning-based approach for the detection of multiple and compound faults in industrial gearboxes using a modified VGG16 Convolutional Neural Network (CNN) architecture. By converting vibration signals into STFT spectrograms and employing data augmentation techniques, the proposed method enhances the model's ability to detect faults with high accuracy across various operating conditions. The results show that the VGG16 model achieved a significant

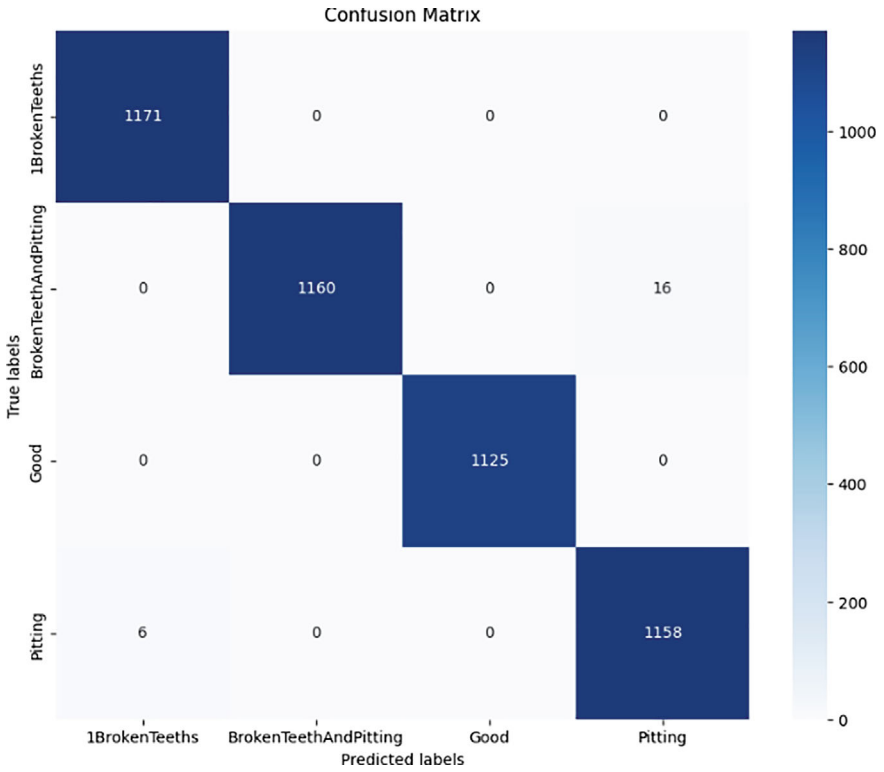


Fig. 11 Testing accuracy of the VGG16 model

Table 4 CNN models comparison after data augmentation

Proposed method	Training accuracy (%)	Testing accuracy (%)
VGG16	99.6	99.52
InceptionV3	86.64	83.61
MobileNet	96.37	96.80
Goswami and Rai [22]	99.42	99.50

improvement in classification accuracy, reaching 99.52%, compared to traditional methods. Moreover, data augmentation led to more feature rich spectrogram images, such that a shallow CNN model like Goswami and Rai could reach an accuracy upto 99.5%. The findings underscore the effectiveness of using CNNs for automatic feature extraction and robust fault classification in complex mechanical systems. This approach offers a scalable and reliable solution for real-world applications, paving the way for more advanced predictive maintenance strategies in industrial environments. Future work could focus on further optimizing the model for real-time fault detection and expanding its applicability to other types of rotating machinery.

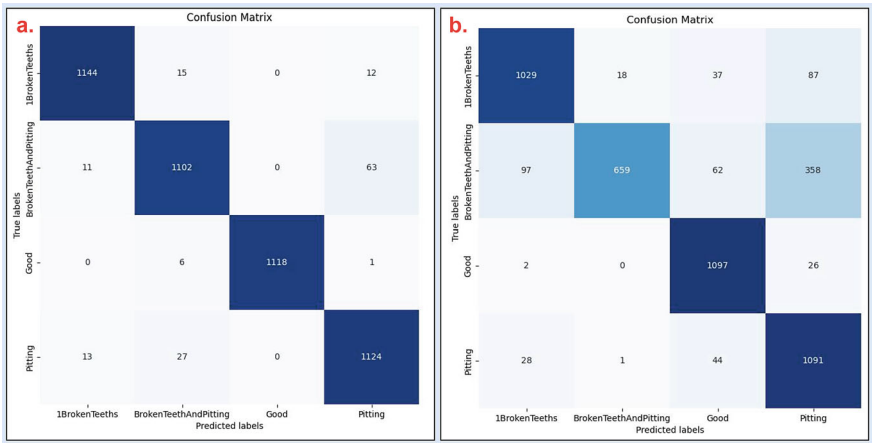


Fig. 12 Confusion Matrix for **a** MobileNet and **b** InceptionV3

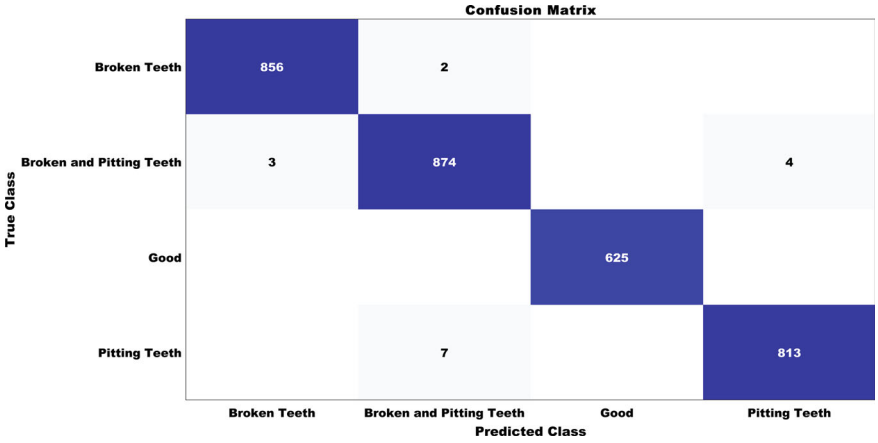


Fig. 13 Goswami and Rai CNN model confusion matrix

Funding This research received no specific grant from any funding agency in public, commercial or not-for-profit sector.

Declaration of Interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Goswami, P., Nandan Rai, R.: A systematic review on failure modes and proposed methodology to artificially seed faults for promoting PHM studies in laboratory environment for an industrial gearbox, Eng. Fail. Anal. **146**, 107076 (2023). https://doi.org/10.1016/j.engfailanal.2023.10707
2. Goswami, P., Rai, R.N.: Water ingressed lubrication oil degradation with cavitation effect detection in gearbox using sound pressure level analysis (n.d). https://ssrn.com/abstract=4948080
3. Nandi, A., Ahmed, H.: Condition Monitoring with Vibration Signals. Wiley (2019). https://doi.org/10.1002/9781119544678
4. Goswami, P., Rai, R.N.: Principal component analysis based vibration sensor selection for fault diagnosis of an industrial gearbox. In: 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6. IEEE (2024). https://doi.org/10.1109/I2MTC60896.2024.10560836
5. Kumar, A., Gandhi, C.P., Zhou, Y., Kumar, R., Xiang, J.: Latest developments in gear defect diagnosis and prognosis: a review. Measurement **158**, 107735 (2020). https://doi.org/10.1016/j.measurement.2020.107735
6. Mo, Z., Zhang, Z., Tsui, K.L.: The variational kernel-based 1-D convolutional neural network for machinery fault diagnosis. IEEE Trans. Instrum. Meas. **70** (2021). https://doi.org/10.1109/TIM.2021.3105252
7. Du Nguyen, T., Nguyen, H.C., Pham, D.H., Nguyen, P.D.: A distinguished deep learning method for gear fault classification using time–frequency representation. Discov. Appl. Sci. **6** (2024). https://doi.org/10.1007/s42452-024-06033-7
8. Randall, R.B.: Vibration-based diagnostics of gearboxes under variable speed and load conditions. Meccanica **51**, 3227–3239 (2016). https://doi.org/10.1007/s11012-016-0583-z
9. Goswami, P., Rai, R.N.: A novel methodology for chipped teeth and worn tooth gear fault detection using minimum entropy deconvolution and Ceemdan. J. Vib. Eng. Technol. **13**, 11 (2025). https://doi.org/10.1007/s42417-024-01622-7
10. Sahoo, S., Laskar, R.A., Das, J.K., Laskar, S.H.: Gear fault diagnosis and classification using machine learning classifier. ACM Int. Conf. Proc. Ser. 69–72 (2019). https://doi.org/10.1145/3325773.3325782
11. Heidari Bafroui, H., Ohadi, A.: Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions. Neurocomputing **133**, 437–445 (2014). https://doi.org/10.1016/j.neucom.2013.12.018
12. Tayyab, S.M., Chatterton, S., Pennacchi, P.: Fault detection and severity level identification of spiral bevel gears under different operating conditions using artificial intelligence techniques. Machines **9** (2021). https://doi.org/10.3390/machines9080173
13. Sharma, V., Parey, A.: Case study on the effectiveness of gear fault diagnosis technique for gear tooth defects under fluctuating speed. IET Renew. Power Gener. **11**, 1841–1849 (2017). https://doi.org/10.1049/iet-rpg.2016.0639
14. Senanayaka, J.S.L., Van Khang, H., Robbersmvr, K.G.: CNN based gearbox fault diagnosis and interpretation of learning features. In: 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), pp. 1–6 (2021). https://doi.org/10.1109/ISIE45552.2021.9576257
15. Gai, J., Zhong, K., Du, X., Yan, K., Shen, J.: Detection of gear fault severity based on parameter-optimized deep belief network using sparrow search algorithm. Measurement **185**, 110079 (2021). https://doi.org/10.1016/j.measurement.2021.110079
16. Dhamande, L.S., Chaudhari, M.B.: Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. Measurement **125**, 63–77 (2018). https://doi.org/10.1016/j.measurement.2018.04.059
17. Xia, M., Mao, Z., Zhang, R., Jiang, B., Wei, M.: A new compound fault diagnosis method for gearbox based on convolutional neural network. In: 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), pp. 1077–1083 (2020). https://doi.org/10.1109/DDCLS49620.2020.9275264

18. Lu, J., Wang, K., Chen, C., Ji, W.: A deep learning method for rolling bearing fault diagnosis based on attention mechanism and graham angle field. Sensors **23** (2023). https://doi.org/10.3390/s23125487
19. Manilow, E., Seetharaman, P., Salamon, J.: Open source tools & data for music source separation (2020). https://source-separation.github.io/tutorial/basics/representations.html. Accessed 1 Sept 2024
20. Goswami, P., Rai, R.N.: Efficient vibration signal denoising techniques for effective condition monitoring of gearbox. In: 2024 11th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 427–432. IEEE (2024). https://doi.org/10.1109/SPIN60856.2024.10512263
21. Goswami, P. Sahu, P.K., Rai, R.N.: An optimum segmentation of gear vibration signals for an effective fault classification using time-domain feature and multi-class support vector machines, pp. 299–312 (2022). https://doi.org/10.1007/978-981-19-0901-6_29
22. Goswami, P., Rai, R.N.: Broken tooth gear fault detection using vibration signals based on convolutional neural network (2024). https://doi.org/10.21203/rs.3.rs-4949296/v1

Experimental Investigation of Vibrational Characteristics of Prosthetic Feet



Preeti Chauhan, Amit Kumar Singh, and Naresh K. Raghuwanshi

Abstract One of the most vital parts of the lower limb prosthesis is the prosthetic foot. The amputees may experience pain from the prosthetic foot's vibrations, which could also harm the residual limb. Taking vibration into account while designing a prosthetic foot can help amputees feel more comfortable. Nonetheless, there are a lot of prosthetic feet in the market, but their vibrational responses have not been thoroughly studied. The purpose of this research is to determine and investigate the vibration properties of prosthetic feet, such as their natural frequency and damping ratio. An experimental modal analysis is conducted on two energy Storing and Return (ESR) prosthetic feet and two conventional prosthetic feet for this study. On prosthetic feet, experiments are conducted using various masses. With varying masses, the Frequency Response Functions (FRFs) are plotted for every foot. Following that, the peak picking approach is used to derive the modal parameters from these FRFs. A comparison is made between the natural frequencies of the prosthetic feet and the walking frequencies of humans. The average human walking frequency is very far from the natural frequencies of Jaipur and ranger foot. The natural frequency of a carbon fiber foot, however, is shown to be quite close to the frequency of human walking. Compared to other prosthetic feet, the ranger foot's damping ratio is the highest. Therefore, compared to other prosthetic feet, the ranger foot has higher shock absorption capability.

Keywords Vibration · Experimental modal analysis · Prosthetic foot · Amputees

P. Chauhan (✉) · N. K. Raghuwanshi
Department of Mechanical Engineering, Malaviya National Institute of Technology Jaipur, Jaipur, India
e-mail: 2020rme9056@mnit.ac.in

A. K. Singh
Department of Mechanical Engineering, National Institute of Technology Calicut, Kattangal, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_21

1 Introduction

An integral component of the lower limb (LL) prosthesis, the prosthetic foot aids in the residual limb's range of motion. The gait symmetry between the residual limb and the sound limb is defined by the prosthetic foot [1]. ESR prosthetic foot can reduce residual limb stress and promote symmetry by storing energy during the early stance of gait and releasing it during the latter stance of gait [2–4]. Many researchers are now focusing on the design of prosthetic feet and how to modify the current design for different uses [5]. The primary goal of a prosthetic foot designer is to create a foot that mimics the natural human foot.

Inertia manipulation, kinetic, metabolic cost, kinematic patterns, inherent stiffness, and mass are the foundation for the design of prosthetic feet [6]. However, the prosthetic feet's modal characteristics have not received much attention from researchers. Every structure has a natural frequency and structural damping. Structures behave differently depending on their structural characteristics and design. Osteoarthritis, loosening of the prosthetic joint, and low back diseases can arise from vibrations transmitted from prosthetic feet to the residual limb [7]. In addition to the vibration's drawback, [6] hypothesized that if amputees' running frequencies aligned with the ESR prosthetic foot's natural frequency, it might provide athletes a mechanical advantage. Modal analysis was carried out on elite blade composite feet used by athletes in the Olympics [6].

For transtibial amputation, a vibratory feedback system based on the open-loop control mechanism of postural control was employed by [8]. Wakeling and Nigg [9] changed the leg's soft tissue's vibration level by engaging in muscle movement. Burnett et al. [7] investigated the composite prosthetic foot and used piezoelectric shunt damping to lessen vibration. Because of their lost limb, LL amputees are a further source of concern. They must naturally dampen the vibrations from impact forces during walking [10] since they lack some muscle compared to healthy persons because of the missing limb, which leads to asymmetry [11, 12]. The remaining limb's tissue may be damaged as a result of the prosthesis's increased vibration transmissibility. The components of the prosthesis, fixation, and suspension determine the frequency and amplitude of vibrations that are conveyed to the tissue [13]. Noroozi et al. [14] used a 53 kg weight to do modal analysis on Ossur Flex-Run prosthetic feet. They came to the conclusion that in order to more accurately forecast its impact on gait patterns and symmetric transitions of human motion, the modal parameter analysis is crucial. Gaining an understanding of these parameters can aid in better tuning the prosthetic foot to the gait of amputees [14].

The literature review mentioned above illustrates that resonance in the prosthetic foot could lead to damage to the residual limb. Because amputees need to wear prosthetic feet constantly and use them for everyday tasks, it is crucial to ensure that the foot's natural frequency does not coincide with the amputee's walking frequency in order to prevent resonance and additional lower limb injury. Few studies are found related to modal parameter analysis of prosthetic feet but these studies are focused on prosthetic foot used in Olympics. They hypothesised that occurrence of resonance

will help athletes to run quickly whereas for daily use prosthetic foot it can be dangerous as vibration transmission to residual may damage it further. Noroozi et al. [14] performed modal analysis at constant mass of 53 kg whereas mass of amputee can vary hence in this study different mass combination as also analysed. Hence in this paper daily use prosthetic feet Jaipur foot, Ranger foot, Nylon foot and Carbon fibre foot are investigated for vibration characteristics with different mass. For this investigation, experimental modal analysis is used which is a very efficient technique to evaluate modal parameters i.e., natural frequency and damping ratio [15, 16]

2 Methodology

In order to analyze the fundamental mode of vibration, the lumped parameter model of a prosthetic foot with a single degree of freedom system is considered in this work, as seen in Fig. 1. Where m is the amputee's mass, k and c is stiffness and structural damping coefficient for prosthetic foot, F is applied force, $\ddot{x}$ and x are acceleration and absolute displacement of mass respectively due to applied force.

Four prosthetic feet, as depicted in Fig. 2, are subjected to experimental modal analysis in order to determine their modal characteristics. One by one, masses weighing 10, 20, 30, 40, and 50 kg are put on various prosthetic feet. Response and im-pulse data are recorded and examined for every mass. Data is acquired using LabVIEW software. The FRFs curve for each recorded data set is generated by

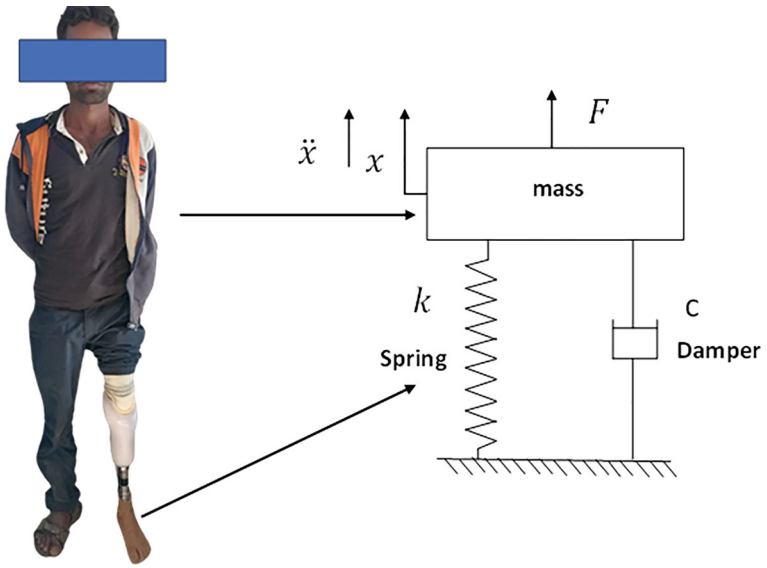


Fig. 1 Lumped parameter model of the prosthetic foot

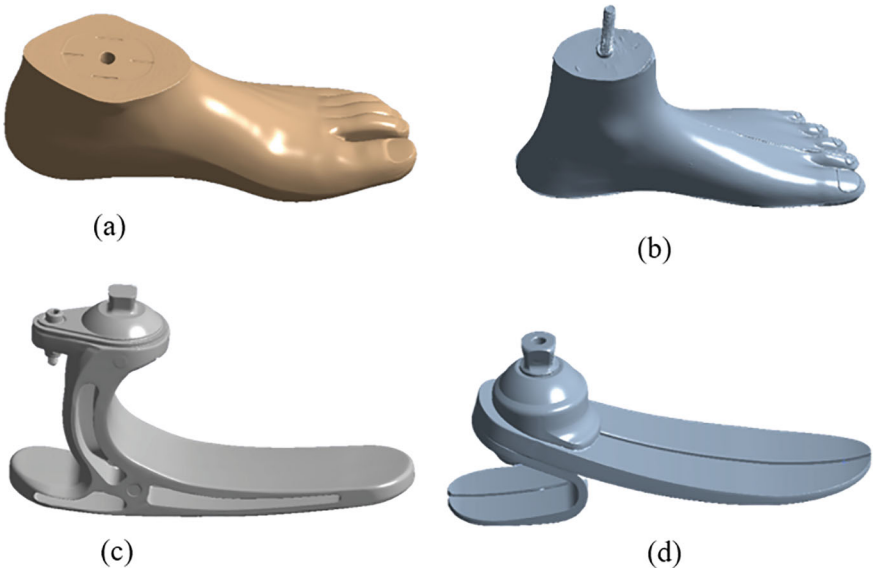


Fig. 2 3D scanned images of **a** ranger foot, **b** Jaipur foot, **c** nylon foot, **d** carbon fibre foot

further processing the data in MATLAB software. A structural response to impulse as a function of frequency is known as a FRF. The response could be displacement, acceleration, or velocity.

$$X(\omega) = H(\omega) * F(\omega) \tag{1}$$

$$H(\omega) = (X(\omega))/(F(\omega)) \tag{2}$$

where $X(\omega)$ is the displacement response, $H(\omega)$ is the transfer function, and $F(\omega)$ is the excitation force as a function of angular velocity ω. Each function is a complex function that may be described by its phase and magnitude. Therefore, each function is a spectral function. Likewise, the transfer function for acceleration and response velocity can be calculated.

The two software programs LabVIEW and MATLAB are used in this paper to determine the modal parameters. Excitation and response signals are obtained using LabVIEW. The measured data are then processed further using MATLAB software, which offers new features specifically designed for modal analysis through the use of a signal processing toolbox. Functions to compute modal parameters and estimate FRF are included in this toolbox. The data exported from LabVIEW is further processed using modal analysis methods found in MATLAB's signal processing package. The FRFs matrix is compiled using the modalfrf function. Regardless of the

sensor, this function always provides the output FRFs in dynamic flexibility (receptance) form. The modal parameters are estimated using the peak-picking approach. For every prosthetic foot, FRFs are plotted for every combination of measured data.

3 Experimental Setup

Any structure's modal parameters can be effectively predicted by experimental modal analysis [19]. Mass is always positioned in the center of the prosthetic foot when conducting trials. The load is supported by fixtures that are designed appropriately. Figure 3 shows several feet with a mass of 10 kg. Other mass combinations are also used in the experiment.

As illustrated in Fig. 4, the experiment's components include the data acquisition system (DAQ) NI-9234, impact hammer, accelerometer LabVIEW 2013, and connecting cables. Figure 5 displays the experimental setup for modal analysis. Various desired places are marked on prosthetic foot. The accelerometer is fixed to the prosthetic foot at various designated locations. A number of studies are conducted before the points are properly marked in order to prevent the possibility that node points will be chosen during the point selection process where no vibration is detected. The impact hammer and accelerometer are connected to separate DAQ system inputs. To acquire the signals, the DAQ equipment is connected to the computer's CPU.

The LabVIEW 2013 software uses graphical programming to provide a graphical user interface (GUI) for data acquisition. The sampling frequency of 1.6516 kHz is utilized to synchronize the data rate from the module for the accurate construction of the data. The foot is excited by the impact hammer. The excitation force is measured by a force transducer in the tip of the impact hammer. This force is transitory, meaning it only lasts for a relatively little time. When the hammer strikes the structure, a double-hitting scenario can occasionally happen. The test must be



Fig. 3 Different prosthetic feet with 10 kg mass **a** Jaipur foot, **b** ranger foot, **c** carbon fibre foot and **d** nylon foot

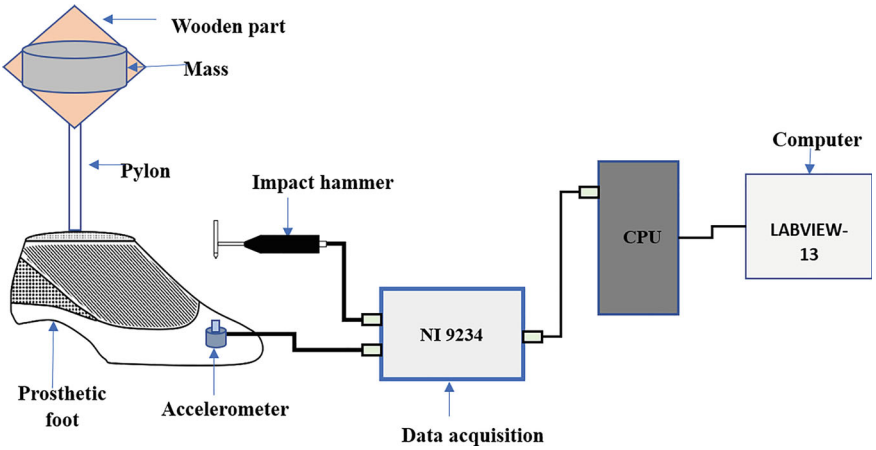


Fig. 4 Schematic diagram of the experimental setup

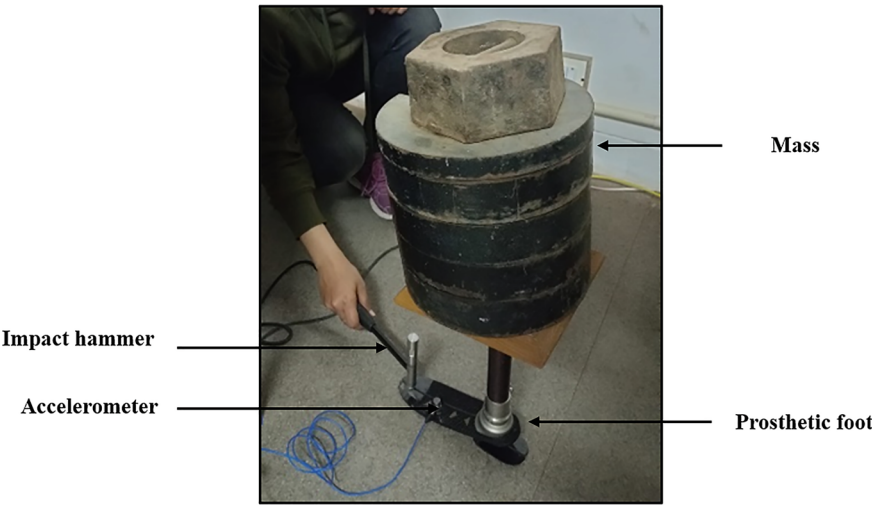


Fig. 5 Experimental setup

redone since the hammer's double hitting generates inaccurate data that should be discarded. For additional analysis, only one hit of data is retained. A variety of tips can be utilized with an impact hammer. Hard metal tips are suitable for high-frequency measurements, while rubber tips are suitable for low-frequency readings. The hammer tip utilized in the current studies is made of plastic, which is neither too hard nor too soft like rubber.

4 Results

Experimental modal analysis is used to create the FRFs plots, as was covered in the preceding sections. Since the resonance events are represented by the peaks, the corresponding frequencies are natural frequencies. The first natural frequency that is analyzed in the current research project is the fundamental frequency. The results demonstrate how altering the mass connected to the prosthetic foot alters the natural frequency and damping ratio. The forcing frequency for our investigation is 2 Hz since this is the average frequency of human walking [17, 18]. Since each person has a different mass, the analysis uses different masses. As a result, different masses produce varied damping ratios and natural frequencies.

The impulse, acceleration, and FRFs graph for the jaipur foot with the attached mass of 50 kg are displayed in Fig. 6. The first natural frequency is 21.3 Hz, as seen in Fig. 6d. Similarly, as Fig. 7 illustrates, further FRF plots are produced for 40, 30, 20, and 10 kg. Only FRFs charts are retained and analyzed in the paper because impulse and acceleration graphs are repeated.

Table 1 also provides an overview of the natural frequency. This indicates that while the natural frequency of the foot lowers with increased mass, it is still far from the forcing frequency of human walking. Therefore, there won't be any resonance conditions for significant mass variation, and the Jaipur foot is safe. The damping ratio

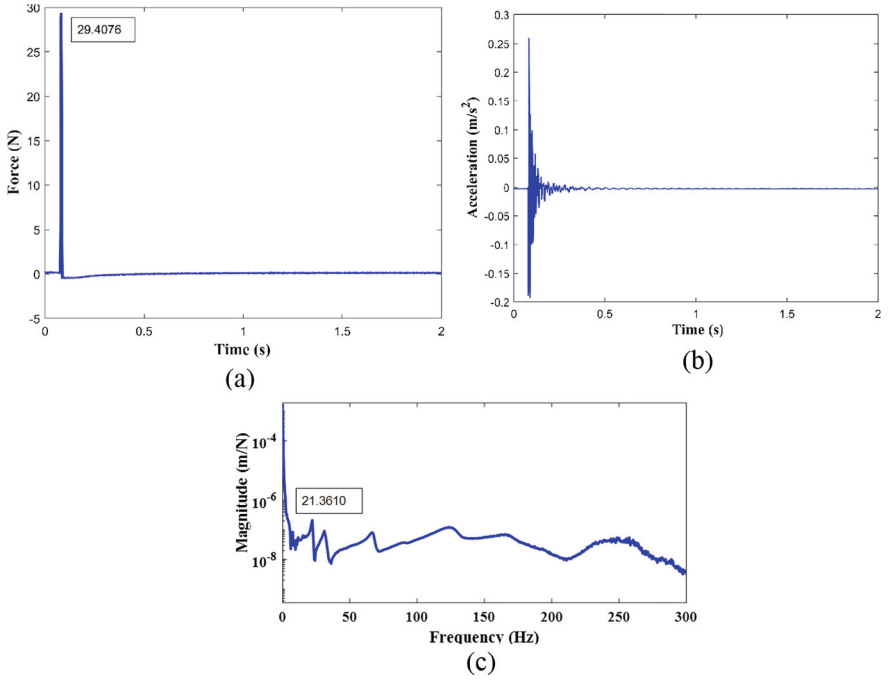


Fig. 6 **a** Impulse, **b** acceleration and **c** FRFs graph for Jaipur foot at 50 kg mass

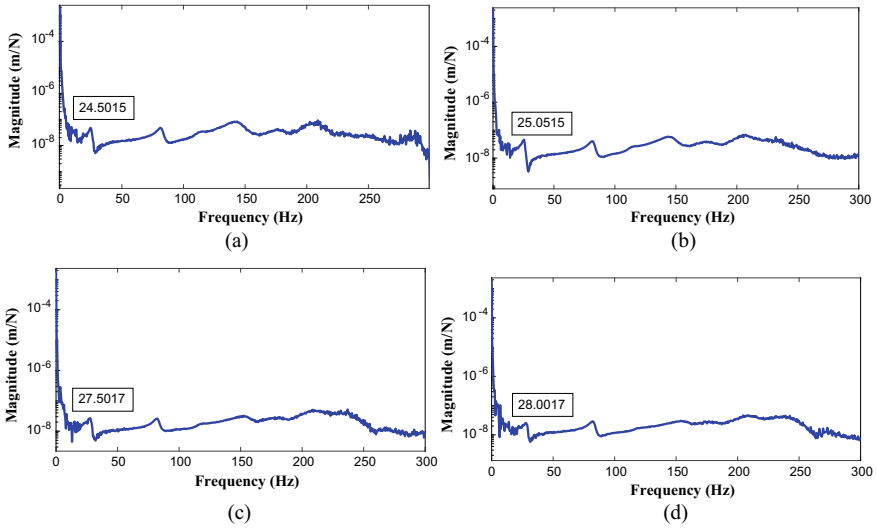


Fig. 7 FRFs graph for Jaipur foot at **a** 40 kg mass, **b** 30 kg, **c** 20 kg and **d** 10 kg

is also displayed in Table 1. Since only structural damping is present, the damping ratio is low. The mass damping ratio also rises as the mass increases. In a similar manner, various masses are attached to ranger, nylon, and carbon fiber feet, and FRFs are calculated for these prosthetic feet during experiments. Table 1 shows the results of the analysis and combination of this data.

The Ranger foot's modal parameters are likewise displayed in Table 1. Table 1 illustrates how changing the mass attached to the ranger foot affects the natural frequency and damping ratio. At 50 kg mass, the damping ratio is 0.1325 and the natural frequency is 6.40. At 10 kg mass, the damping ratio is 0.0795 and the natural frequency is 10.03. Table 1 indicates that when mass attached grows, natural frequency falls. The obtained damping ratio is more than the Jaipur foot, and the variation in the damping ratio does not correspond to the mass increase.

Additionally, Table 1 displays the modal characteristics of a nylon prosthetic foot with varying masses attached. Table 1 displays the relationship between the prosthetic foot's modal characteristics and mass fluctuation. At 50 kg mass, the damping ratio and natural frequency are 0.0804 and 4.14, respectively. At 10 kg mass, the damping ratio is 0.0353 and the natural frequency is 6.63. With an increase in mass from 10 to 50 kg, the natural frequency has similarly dropped. Increasing mass also results in an increase in the damping ratio.

The damping ratio and natural frequency for carbon fiber feet with varying masses are also displayed in Table 1. This parameter's variation with various masses attached to the prosthetic foot is shown in the Table 1. At 50 kg mass, the carbon fiber foot's inherent frequency and damping ratio are 2.50 and 0.0478, respectively. At 10 kg mass, the damping ratio is 0.0206 and the natural frequency is 4.32. As masses rise, the damping ratio increases irregularly.

Table 1 Modal parameters of different prosthetic foot

Prosthetic foot	Mass (kg)	Natural frequency (Hz)	Damping ratio
Jaipur foot	10	28.0017	0.0438
	20	27.5017	0.0590
	30	25.0515	0.0750
	40	24.5015	0.0810
	50	21.3610	0.0920
Ranger foot	10	10.0336	0.0795
	20	9.5369	0.1018
	30	7.1742	0.1110
	40	6.5756	0.1039
	50	6.3976	0.1325
Nylon foot	10	6.6339	0.0353
	20	5.9767	0.0618
	30	5.4994	0.0405
	40	4.6907	0.0733
	50	4.1384	0.0804
Carbon fibre foot	10	4.3219	0.0206
	20	3.6118	0.0250
	30	3.0688	0.0251
	40	2.8385	0.0293
	50	2.4955	0.0478

Nylon foot and Carbon fibre foot are ESR prosthetic feet. At 10 kg mass, the carbon fiber foot's natural frequency is 4.32; as mass increases, its natural frequency falls. At 50 kg mass, its natural frequency is 2.50 Hz. It is a little closer to the frequency of human walking. The natural frequency of a nylon foot is 4.14 Hz at 50 kg mass and 6.63 Hz at 10 kg mass. Compared to conventional prosthetic feet, the natural frequency of these two ESR prosthetic feet is lower. There may be a possibility that the natural frequency of a carbon fiber foot will match the walking frequency of amputees at higher masses, as the natural frequency of the foot may further decrease with increased mass. Additionally, the prosthetic foot is attached to a number of different parts, including a socket, pylon, and adaptor. It is also possible to do additional research to determine whether these components reduce vibration before it gets transmitted to the limb or transfer it directly to the limb. Furthermore, when the natural frequency of a prosthetic foot matches with the frequency of human walking, further research can be done to estimate how much vibration gets transmitted to the human leg and how much damage is caused by vibration transmission to the residual limb.

Experimental measurements of the damping ratio are also made here. A low damping ratio results in less energy dissipation, meaning vibrational oscillations

will fade more slowly if the damping ratio is low, and more quickly if the damping ratio is high. The Ranger foot has the highest damping ratio, followed by the Jaipur, nylon, and carbon fiber types. The Ranger foot's vibrational oscillation will decay more quickly and last shorter periods than the other three prosthetic feet. Effective cushioning at the heel and outside body may be the likely cause of Ranger foot's high damping ratio. When compared to the other three prosthetic feet, the Ranger foot's high damping ratio indicates that it has a good tendency to absorb shock.

5 Conclusions

Since vibrations in the prosthetic foot can cause pain for amputees, vibration characteristics help in the better design of prosthetic feet to minimize vibrations. Reducing the likelihood of additional harm to the amputee's residual limb is also crucial. Different prosthetic feet with varying masses are used for experimental modal analysis. The current study examines a total of four prosthetic feet. These prosthetic feet include the nylon, carbon fiber, Ranger and Jaipur feet. The first two are ESR prosthetic feet, and the other two are traditional prosthetic feet. The natural frequencies and damping ratios are assessed for every foot with varying masses based on the modal analysis experiment.

The maximum natural frequency for a 50 kg mass is 21.36 Hz for the Jaipur foot. 6.39 Hz is the natural frequency found for the Ranger foot at 50 kg mass. The human walking frequency is very distant from the natural frequencies of the two conventional feet. Therefore, there is no possibility that resonance will occur. Additionally, the ranger foot weighing 50 kg has the highest damping ratio 0.13 among all four prosthetic foot. Therefore, the ranger foot has a high capacity for shock absorption. The natural frequency of a nylon foot at 50 kg for the ESR prosthetic feet is 4.14 Hz, which is is also distant from the walking frequency. The natural frequency of a carbon fiber foot at 50 kg is 2.49 Hz, which is extremely near to the frequency of human walking. Therefore, using it on amputees with higher mass could result in resonance. Additional research can be conducted to determine the harm that resonance in prosthetic feet is causing to a residual limb.

References

1. Chauhan, P., Singh, A.K., Raghuwanshi, N.K., Rathore, R.: Experimental investigation of gait parameters and stability analysis for unilateral transtibial amputee. Sādhanā **48**, 1–18 (2023)
2. Chauhan, P., Singh, A.K., Raghuwanshi, N.K.: Classifying the walking pattern of humans on different surfaces using convolutional features and shallow machine learning classifiers. Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. **238**, 1–12 (2024)
3. Chauhan, P., Singh, A.K., Raghuwanshi, N.K.: Classifying walking pattern on different surfaces by optimising features extracted through IMU sensor data using SSA optimisation. J. Br. Soc. Mech. Sci. Eng. **47**(1), 1–17 (2025)

4. Chauhan, P., Raghuwanshi, N.K., Singh, A.K.: Implementation of supervised machine learning algorithms for gait alteration classification of the human foot. In: International Conference on Mechanical Engineering: Researches and Evolutionary Challenges, pp. 449–460 (2023)
5. Chauhan, P., Singh, A.K., Raghuwanshi, N.K.: The state of art review on prosthetic feet and its significance to imitate the biomechanics of human ankle-foot. Mater. Today Proc. **62**, 6364–6370 (2022)
6. Noroozi, S., Sewell, P., Rahman, A.G.A., Vinney, J., Chao, O.Z., Dyer, B.: Modal analysis of composite prosthetic energy-storing-and-returning feet: an initial investigation. Proc. Inst. Mech. Eng. Part P J. Sports Eng. Technol. **227**, 39–48 (2013)
7. Burnett, J.K., Choi, Y.T., Li, H., Wereley, N.M., Miller, R.H., Shim, J.K.: Vibration suppression of a composite prosthetic foot using piezoelectric shunt damping. IEEE Trans. Biomed. Eng. **68**, 2741–2751 (2021)
8. Rusaw, D., Hagberg, K., Nolan, L., Ramstrand, N.: Can vibratory feedback be used to improve postural stability in persons with transtibial limb loss. J. Rehabil. Res. Dev. **49**, 1239–1254 (2012)
9. Wakeling, J.M., Nigg, B.M.: Modification of soft tissue vibrations in the leg by muscular activity. J. Appl. Physiol. **90**, 412–420 (2001)
10. Druga, C., Barbu, D., Lache, S.: Vibration and the human body. Fascicle Manage. Technol. Eng. **6**, 168–173 (2007)
11. Hobara, H., Baum, B.S., Kwon, H.J., Linberg, A., Wolf, E.J., Miller, R.H., Shim, J.K.: Amputee locomotion: lower extremity loading using running-specific prostheses. Gait Posture **39**, 386–390 (2014)
12. Kiernan, D., Miller, R.H., Baum, B.S., Kwon, H.J., Shim, J.K.: Amputee locomotion: frequency content of prosthetic vs. intact limb vertical ground reaction forces during running and the effects of filter cut-off frequency. J. Biomech. **60**, 248–252 (2017)
13. Lehmann, J.F., Price, R., Boswell-Bessette, S., Dralle, A., Questad, K.: Comprehensive analysis of dynamic elastic response feet: Seattle Ankle/Lite Foot versus SACH foot. Arch. Phys. Med. Rehabil. **74**, 853–861 (1993)
14. Noroozi, S., Ong, Z.C., Khoo, S.Y., Aslani, N., Sewell, P.: Dynamic characterisation of Össur flex-run prosthetic feet for a more informed prescription. Prosthet. Orthot. Int. **43**, 62–70 (2019)
15. Chandravanshi, M.L., Mukhopadhyay, A.K.: Modal analysis of structural vibration. In: ASME International Mechanical Engineering Congress and Exposition, San Diego, California, pp. 1–9 (2013)
16. Raghuwanshi, N.K., Parey, A.: A new technique of gear mesh stiffness measurement using experimental modal analysis. J. Vib. Acoust. **141**, 1–13 (2019)
17. Heinemann, P., Kasperski, M.: Damping induced by walking and running. Proc. Eng. **199**, 2826–2831 (2017)
18. Ji, T., Pachi, A.: Frequency and velocity of people walking. Struct. Eng. **84**, 36–40 (2005)
19. Oudenhoven, L.M., Boes, J.M., Hak, L., Faber, G.S., Houdijk, H.: Regulation of step frequency in transtibial amputee endurance athletes using a running-specific prosthesis. J. Biomech. **51**, 42–48 (2017)

Vibration Trend Monitoring of Indigenous Wankel Rotary Engine

Shah Brijeshkumar, Sadanand Kulkarni, Umesh Kumar Sinha, Santhosh Kumar, and A. J. Steve Mithran

Abstract Wankel rotary engines are found to be useful in unmanned aerial vehicles, powered hang gliders, loitering munition, and other strategic applications. They are also promising candidates for hybrid power trains, due to their high power-to-weight ratio. CSIR-NAL has indigenously developed Wankel engines with different power ratings, such as 65, 55, and 30 hp. Out of these, 30 hp is an air-cooled and single-rotor engine. A detailed study has been undertaken to create a health management framework for this engine. Several parameters, including vibration, temperature, etc., are monitored for health and performance assessment. Typically, compared to their nearest sibling, reciprocating IC engines, Wankel engines have less vibration due to their rotary nature. However, continuous monitoring and trend analysis of vibration data becomes essential to assess the health of the rotating components of the engine. Experimental data was collected during ground testing following specified endurance test cycles. This paper describes the vibration trend monitoring adopted for the 30-hp class Wankel engine. In total, twenty-one frequencies were identified, and variations in amplitude over the period were tracked along with their harmonics. During the testing, all the rotating components were labelled to be healthy based on the established trend. This analysis will help to establish engine endurance and time between overhauls of the indigenous Wankel engine.

Keywords Wankel rotary engines · Vibration analysis · Health management

1 Introduction

Wankel rotary-type combustion engines have found their purpose in UAVs both in defence and the private sector. The major benefit of the Wankel engine has high specific power compared to its reciprocating counterparts. Also, this engine has fewer moving parts resulting in ease of serviceability and high reliability. The lower frontal

S. Brijeshkumar (✉) · S. Kulkarni · U. K. Sinha · S. Kumar · A. J. S. Mithran
CSIR-National Aerospace Laboratories, Bangalore, India
e-mail: shahbrij@nal.res.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_22

area is an added advantage for UAVs from a drag perspective. CSIR-NAL has been involved in the development of Wankel rotary engines and successfully developed of 55 and 65-hp engines for unmanned aerial vehicles was a major milestone [1]. Sensing a need for power plants for the 120 kg class UAV, CSIR-NAL has taken up the development of a 30 hp air-cooled Wankel engine. While there have been plenty of studies on the health assessment of reciprocating-type IC engines [2, 3]; very little literature can be found on rotary-type engines. The general theme of these studies revolves around combustions—fuel selections, emission studies and firing analysis etc. [4–6].

A study was conducted to assess the health of the engine by continuous monitoring of engine vibration signatures along with other parameters. Similar approach is reported in [7] for reciprocating IC engine and in [8] for turbo-prop engine. However, in [7] Tomaszewski et al. focuses more on other moving components' natural frequency and its monitoring and in [8], focus is on overall vibration levels of the engine. While there is a similarity in approach, fundamentally all the engine types are different was far as vibration signatures are concerned and hence, results obtained or procedure established are not directly comparable. In present study, it was planned to generate a database of various failure signatures of the engine and develop a data-driven diagnostic model to identify critical failures that are anticipated to occur during engine operation. Another objective of the study was to optimize the sensor/sensor location required for effective health management. In this paper, a detailed methodology adopted for vibration trend monitoring is presented.

1.1 30 hp Wankel Engine

The 30-hp Wankel engine is an air-cooled, direct drive type engine working on the Otto cycle. The rated speed of the engine is 7000 rpm at ISA-sea level. The Wankel engine, being a rotary-type engine, delivers a high power-to-weight ratio. This engine weighs less than 12 kg. In Table 1, details important engine specifications.

The comprehensive engine test setup is equipped with an array of sophisticated sensors, including accelerometers, dynamic pressure sensors, rotary encoders, strain gauge devices, positive displacement flow sensors, and tachometer, along with several temperature measurements. Based on the dynamic nature of the parameters, the sampling rate is decided. Vibration data is collected at 25.6 k Samples/Sec throughout the testing duration. The location of various sensors on the engine assembly is shown in Fig. 1. While the vibration sensor after the shock mount does not provide explicit information on the engine vibration health, it is very informative about the mounting structures' condition and the effectiveness of shock mounts.

Table 1 Technical specifications of 30 hp Wankel engine

Parameter	Description
Type	Single-rotor Wankel engine
Thermodynamic cycle	Otto cycle
Power	30 hp (22.4 kW) @ 7000 rpm at ISA-sea level (without exhaust muffler)
Max propeller speed	7000 rpm (direct drive)
Cylinder capacity	216 cc
Compression ratio	9.2
Housing cooling	Ram air
Rotor cooling	Ram Air
Lubrication	Total loss forced lubrication system
Ignition	CDI system
Fuel used	AV GAS—100 LL/unleaded gasoline
Fuel supply	Carburetor- Diaphragm type
Specific fuel consumption	335–365 g/kWh (0.55–0.60 lb/hp/h)
Engine weight	< 12 kg

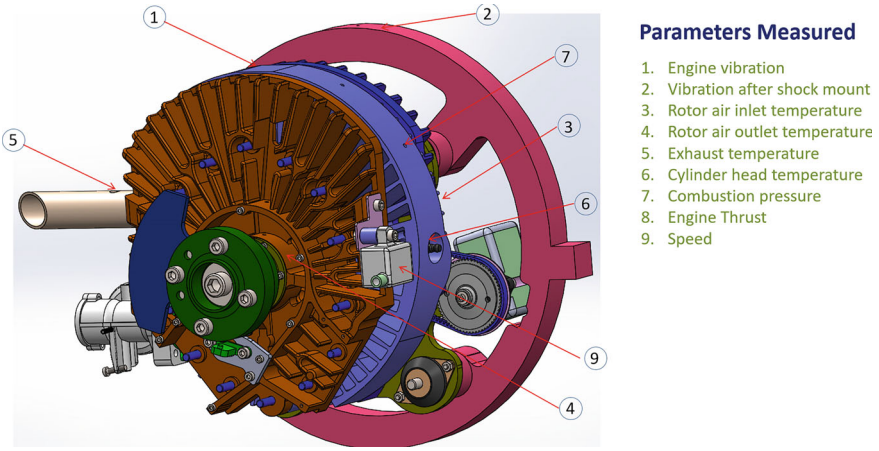


Fig. 1 Engine assembly with location of various sensors

2 Health Management Methodology Approach

In the absence of a large fleet of engines, the most suitable approach to health management is a hybrid of data-driven and computer-based simulation models. In Fig. 2, a flow chart adopted for HM development is shown. Extensive engine testing was planned on the new engine. The data from the initial few runs of the engine was

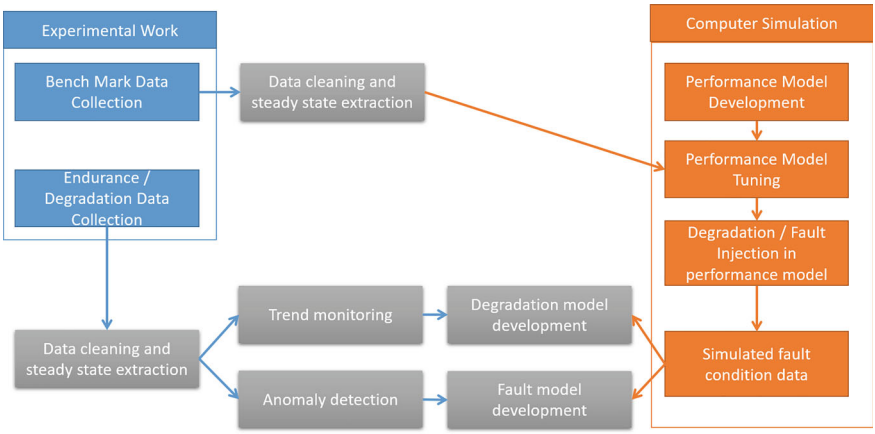


Fig. 2 Methodology for health management system development

considered as benchmark data for the new engine. The same engine is run further to collect data on the degraded engine. Each data collected needs its own data cleaning procedure. Based on data collected during benchmarking, the same is established, and steady-state data is extracted. The same procedure is used for subsequently collected data as well. Thus, processed data is then used for trend-monitoring purposes. In case any anomaly is captured during experimentation, the data can also be used for anomaly detection.

2.1 Bearing Details and Frequencies of Interest

The engine consists of four bearings: two on the drive side, one on the non-drive side, and one on the eccentric shaft. Needle roller bearings are a typical choice for compact bearing arrangement. The rotor bearing, which consists of 27 rollers without inner or outer race, is housed between the rotor and the shaft. As the rotor rotates 1/3 times the shaft speed, the relative speed of the hypothetical inner race is 2/3 of the shaft rotational speed.

Bearing fundamental frequencies is crucial for diagnosing faults caused by bearing defects. The four main frequencies of interest, also known as bearing failing frequencies, are:

1. Ball spin frequency (BSF)
2. Fundamental Train frequency (FTF)
3. Ball pass frequency inner (BPFI)
4. Ball pass frequency outer (BPFO).

There are particular formulas to calculate each of them; in cases where detailed bearing dimensions are unknown. OEMs provide tools to calculate these frequencies

Table 2 Bearing fundamental frequencies

Bearing No	Bearing type	FTF	BSF	BPFI	BPFO
NKS28	Needle roller bearing	0.431	3.542	8.538	6.462
61905-2RZ	Deep groove ball bearing	0.429	3.446	7.424	5.576
NKS25	Needle roller bearing	0.424	3.202	8.068	5.932
K 55X60X30	Needle roller bearing	0.478	11.478	21.391	19.609

[5]. Table 2 details all the frequencies of interest from a health assessment point of view. These frequencies, along with 1x rotational frequency, are monitored up to three harmonics. Based on Table 2 and the rated speed of the engine, it is decided to implement a low pass filter with 2000 Hz bandwidth.

2.2 Data Processing

Raw data collected from the engine test bed needs to be prepared and processed for further analysis. Data acquisition systems take care of sensor signal amplification, unit conversion, etc. and provide test data directly into engineering units. However, as with the present case, data is being collected at 25.6 k samples/s. This results in a considerable file size, making it difficult for DAQ to handle such extensive data. Hence, each 5-min run is stored in separate files. The first step in pre-processing is to merge this data and form a single time series data stream for each vibration channel, followed by filtering of noise or outliers from the data.

As discussed in the section above, all major frequencies of interest are below 2000 Hz. Hence, a low-pass filter was implemented to filter out high-frequency data (Fig. 3) 1 s RMS of the filter data is computed for quick anomaly checks and automated steady-state points extraction.

The one-sided power spectrum was computed using a discrete Fourier transform (DFT) with a Hanning window. Windowing parameters, DFT length, overlapping percentage, etc., were arrived at by a few trials on initial benchmarking data to verify their sensitivity on the spectrum.

2.3 Data Benchmarking

As the new engine underwent the first batch of testing, the data collected was labelled as benchmark data. Benchmarking engine data when the engine is in new condition is crucial. It allows us to access that the engine operates within the expected parameters. Also, the initial few runs and statistically derived parameters therein give an estimate of measurement variation and effect on confined variables or parameters. In other words, benchmarking helps establish baseline data for the new engine. Figure 4 shows

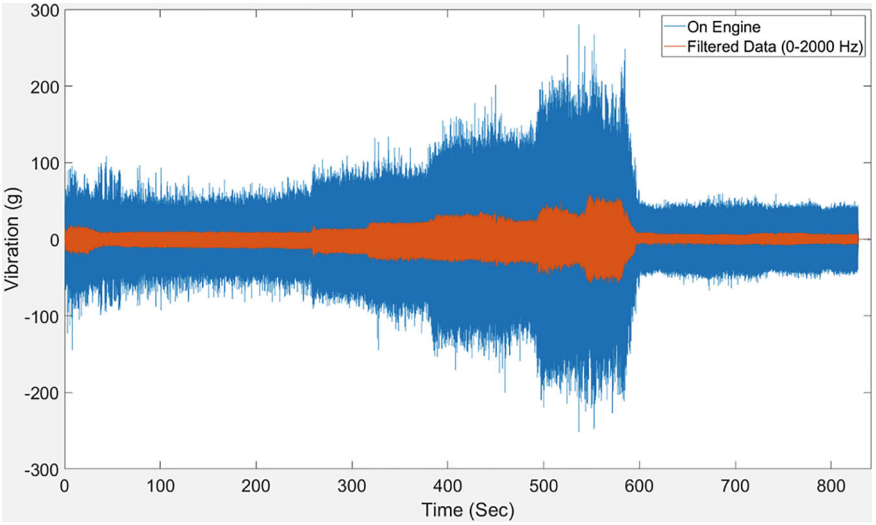


Fig. 3 On-engine vibration data—raw and filtered

initial benchmarking vibration data in the radial direction measured on the engine. Time–frequency domain analysis is shown in Fig. 5. For vibration on the engine in an axial direction during the first performance test runs. Rotation frequency or 1x is clearly visible and dominant for most of the test duration. However, all the 1x multiple harmonics are present in the data, which is a typical characteristic of the Wankel engine.

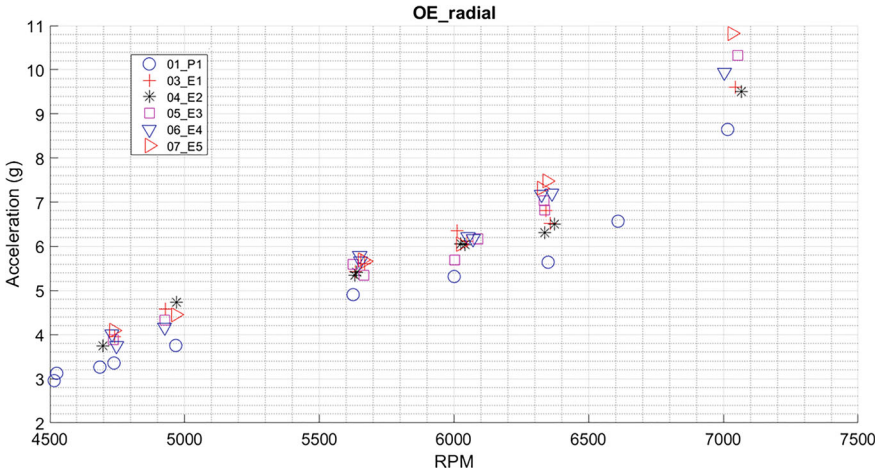


Fig. 4 Baseline benchmarking—1 Sec RMS spread

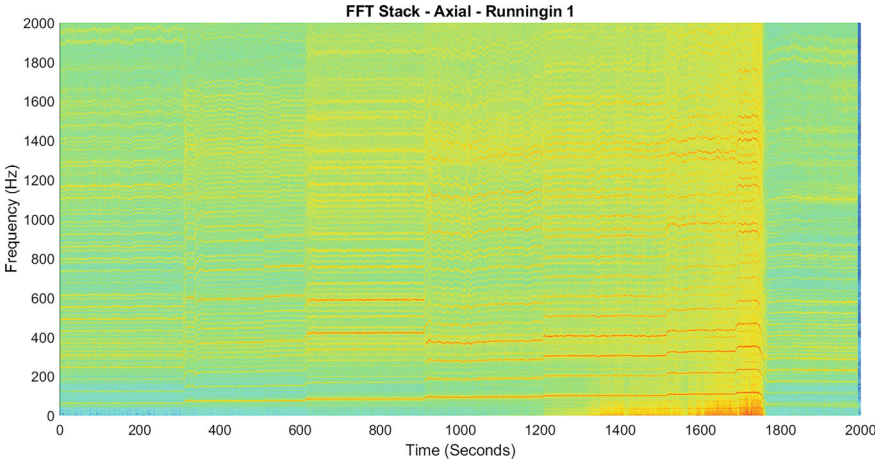


Fig. 5 FFT stack (STFT) of axial vibration

2.4 Monitoring of Bearing Characteristic Frequencies

Vibration analysis from the above sub-section was carried forward to all the bearing frequencies and their harmonics as described in Sect. 3.1. The power spectrum of the 1-s steady-state data was computed, and the amplitude of the specified frequency component was recorded and plotted as shown in the sample plot Fig. 6 for the radial direction. An additional parameter was considered from a health monitoring perspective: the ratio of a given frequency component amplitude to overall RMS vibration amplitude. This non-dimensional unbounded ratio provides an indication of the change in contribution or relation of a given frequency component in overall vibration. This will help in isolating bearing faults with more accuracy. In the plots, this ratio is plotted as bars behind with a vertical axis scale on the right side. From the figure, it is evident that during the complete endurance bearing condition and, in general, the rotating system was healthy.

3 Conclusions

The systematic approach for vibration-based health monitoring is presented in this work. It is important to continuously monitor engine rotor system health. Vibration trend monitoring provides a handy tool to compare present conditions with respect to historical data. Apart from overall vibration (RMS) monitoring, each of the characteristic frequencies also needs to be monitored. In the present case being a direct drive system, each bearing frequencies are monitored. However, if the gearbox is present in the system, additional frequencies need to be monitored as well. During the complete endurance testing of the engine, the rotor-bearing system is reported

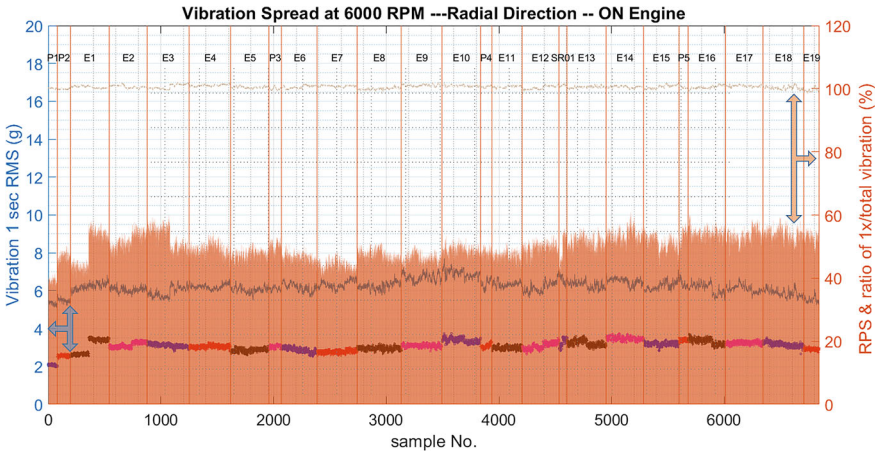


Fig. 6 Vibration trend monitoring of 1x component in radial direction

to be healthy in the present engine. However, it is planned to continue the exercise with other 55 hp and 65 hp engine in future. The above exercise will help in establishing maintenance and over haul interval, as well as total technical life of this indigenous engine. Once the engine goes in service, further trend monitoring will help maintenance team to take pro-active maintenance actions.

Acknowledgements The authors acknowledge the support and encouragement of the Director, CSIR-NAL, Head, Propulsion division, and Dy Head, Propulsion division throughout this activity.

References

1. Kulkarni, S., et al.: Vibration reduction in indigenous Wankel rotary combustion engine with structured layer damping. In: Proceedings of the 6th National Symposium on Rotor Dynamics. Lecture Notes in Mechanical Engineering, Springer, Singapore (2020)
2. Yadav, S.K., Tyagi, K., Shah, B., Kalra, P.K.: Audio signature-based condition monitoring of internal combustion engine using FFT and correlation approach. IEEE Trans. Instrum. Meas. **60**(4), 1217–1226 (2011)
3. Geyer, D., Kick, M., Kraus M.: Monitoring the functional reliability of an internal combustion engine.. USA Patent 8392046 (2013)
4. Smail, B., Mohiuddin, A.K.M.: Combustion chamber design effect on the rotary engine performance—a review. Int. J. Autom. Eng. **11**(4), 200–212 (2020)
5. Levi, O., Eliaz, N.: Failure analysis and condition monitoring of an open-loop oil system using ferrography. Tribol. Lett. **36**, 17–29 (2009)
6. Vorraro, G., Turner, J.W.G., Akehurst, S.: Testing of a modern Wankel rotary engine—part III: firing condition analysis. SAE Technical Paper Series (2022)
7. Tomaszewski, F., Szymański, G.M.: Frequency analysis of vibrations of the internal combustion engine components in the diagnosis of engine processes. Arch. Transp. **24**(1), 117–125 (2012)

8. Fernande, W.G., Vinay, C.A., Priyadarshini, L.: Aero engine vibration measurement. In: iNaCoMM2015-074 (2015)
9. SKF. https://productselect.skf.com/#/size-lubrication/single-bearing

Implementation of Multi Criteria Decision Making (MCDM) Techniques in Selecting Influencing Factors for the Affecting of Artificial Intelligence (AI) in the Manufacturing

Ashish Mogra, Man Mohan Siddh, and Peeyush Vats

Abstract Artificial Intelligence (AI) has become a transformative force in the manufacturing sector, enabling innovations such as predictive maintenance, quality control, and process optimization. However, the adoption of AI requires the consideration of multiple influencing factors, including technological, organizational, economic, and societal aspects. Multi-Criteria Decision-Making (MCDM) techniques provide a robust framework for identifying, prioritizing, and evaluating these factors. This article explores the implementation of MCDM approach in selecting the critical factors influencing AI adoption, offering insights into practical methodologies and their implications for the manufacturing.

Keywords Artificial intelligence · Multi-criteria decision-making · Manufacturing sector

1 Introduction

The manufacturing sector is undergoing an outstanding modification with the incorporation of Artificial Intelligence (AI), which provides advancements in areas such as predictive maintenance, quality control, and supply chain optimization. A wide array of manufacturing techniques has emerged in recent years, driven by increasing

A. Mogra
School of Technology Management and Engineering, SVKM's, NMIMS, Sarangpur, Chandigarh, India

M. M. Siddh (✉)
Department of Mechanical Engineering, Jaipur Engineering College and Research Centre, Jaipur, India
e-mail: manmohansiddh@gmail.com

P. Vats
Department of Mechanical Engineering, Poornima College of Engineering, Jaipur, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_23

competition across various markets and the ever-growing, diverse demands of consumers [1]. As a result, numerous manufacturing methods are now available to meet the specific needs of different production applications. However, the adoption of AI in manufacturing is influenced by a complex interplay of factors, including technological readiness, economic considerations, organizational culture, and regulatory compliance. Effectively navigating these factors requires robust decision-making frameworks, and Multi-Criteria Decision-Making (MCDM) methods have emerged as valuable tools in this context [2].

While various Multi-Criteria Decision-Making (MCDM) approaches and techniques are accessible, selecting the most suitable manufacturing method for a specific application remains a demanding and complex job. This complexity arises from the intricate correlation between the selection process and the unique characteristics of different manufacturing techniques, making the process time-intensive. Additionally, the uneven importance assigned to the criteria in decision-making further complicates the process [3].

As a result, employing an appropriate selection methodology is crucial for addressing specific problems and applications. Conducting a comparative analysis of different manufacturing techniques is essential to identify the most effective approach [4]. In recent years, MCDM tools have gained significant popularity as they provide a structured framework and critical support for such resolution in the manufacturing. These tools enhance the ability to precisely assess and order numerous criteria, making them invaluable for selecting optimal manufacturing techniques [5].

AI adoption in manufacturing is multifaceted, involving several critical factors

Technological Readiness: Technological readiness is a crucial variable in the affecting of Artificial Intelligence (AI) inside the manufacturing. It encompasses an institution potential to effectively implement and integrate AI technologies into existing processes, infrastructure, and systems [6]. A robust technological infrastructure is essential for AI acquisition. This includes advanced hardware, such as high-performance computing systems, and software platforms capable of supporting AI applications. Additionally, the availability of resources like data storage solutions and network capabilities plays a critical role. Organizations must ensure that their infrastructure can handle the computational demands of AI algorithms and facilitate seamless integration into manufacturing operations [7]. The availability and maturity of AI technologies and the existing IT infrastructure to support their implementation [8].

Technological readiness is a multifaceted determinant in the promising acquisition of AI in manufacturing. By thoroughly evaluating infrastructure, data quality, technological competence, integration capabilities, and cybersecurity measures, organizations can develop a comprehensive understanding of their preparedness. Utilizing MCDM techniques facilitates a structured assessment, enabling informed decision-making and strategic planning for AI integration [9].

Organizational Culture and Readiness: Organizational culture and readiness are pivotal in determining the successful acquiring of Artificial Intelligence (AI)

surrounded by the manufacturing. A tradition that holds revolution, continuous learning, and adaptability significantly enhances an organization's capacity to integrate AI technologies effectively [10]. The alignment of AI initiative with organizational goals, employee readiness, and the presence of a culture that supports innovation. A culture that fosters innovation encourages employees to explore new technologies and methodologies. Such an environment supports, experimentation and tolerates calculated risks, which are essential for AI adoption [11]. Organizations that promote innovative thinking are more likely to identify and leverage AI applications that enhance manufacturing processes. Cultivating an institution trend that supports transformation, continuous learning, and moral practices is basic to the successful affectation of AI in manufacturing. Utilizing MCDM techniques to assess and enhance organizational readiness ensures a structured and effective approach to AI integration [12].

Regulatory and Ethical Compliance: The Artificial Intelligence (AI) incorporation into the manufacturing sector offers significant advancements in efficiency and innovation. However, it also introduces complex regulatory and moral provocation that organizations must navigate to make sure accountable adoption [13]. Resolving these issues is essential to upholding legal requirements, public confidence, and business integrity. AI applications in manufacturing are subject to a range of regulations aimed at ensuring safety, privacy, and fairness. In order to protect stakeholders' trust and prevent legal ramifications, compliance with these standards is crucial. Adherence to industry regulation, data privacy laws, and ethical standards governing AI use [14].

Regulatory and ethical compliance are integral to the responsible promotion of AI in manufacturing. Institutions must proactively address these considerations to ensure legal adherence and to uphold ethical standards. Utilizing MCDM techniques facilitates a comprehensive evaluation of these factors, enabling organizations to navigate the complexities of AI integration effectively [15].

Given the complexity and interdependence of these factors, traditional decision-making approaches may fall short. MCDM techniques offer structured methodologies to assess and rank numerous, conflicting standard, and facilitating informed and balanced decision-making [16]. Multi Criteria Decision Making (MCDM) techniques that are often employed include:

Analytic Hierarchy Process (AHP): A technique that decomposes a decision problem into a hierarchy of sub-problems, allowing for pairwise comparisons and priority setting [17].

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS): An approach to ranking options that finds solutions that are closest to the ideal and furthest from the negative-ideal [18].

Decision-Making Trial and Evaluation Laboratory (DEMATEL): A technique used to examine and model complex causal relationships among factors [19].

Fuzzy MCDM: A method for dealing with ambiguity and uncertainty in decision-making processes that uses fuzzy logic [20].

The utilization of MCDM methods in production has been across in numerous learning. For instance, the integration of MCDM approaches with Quality Function Deployment (QFD) has been shown to enhance decision-making in product development by systematically evaluating customer requirements and technical attributes [21].

Additionally, the use of AI in smart manufacturing has been linked to improved multi-criteria decision-making, enabling real-time adjustments based on customer preferences and production capabilities.

This article presents the application of MCDM techniques in selecting influencing factors for AI acquisition in the manufacturing. It provides an overview of key influencing factors and discusses various MCDM methodologies, and presents practical insights through case studies and examples. By leveraging these techniques, manufacturing institutions can direct the complication of AI appropriation and harness its transformative potential effectively.

2 Literature Review

Employed to compute the criteria weights and evaluate supplier scores using an essential element of the Analytic Hierarchy Process (AHP) is the pairwise comparison matrix. The resulting data was subsequently utilized to train an Artificial Neural Network (ANN) model, leveraging a genetic algorithm (GA) to optimize the training process [22]. More recently [23], Genetic Algorithms (GA) has been employed in conjunction with a Multi-Criteria Decision-Making (MCDM) approach that integrates correlation coefficients and standard deviation for enhanced decision-making. Genetic programming (GP) and data envelope analysis (DEA) were used to solve the problem of choosing environmentally friendly suppliers for the apparel manufacturing sector along with MCDM technique. GP was utilized to handle the computational complexity and time-intensive nature of the DEA model. Additionally, MCDM techniques used to rank supplier performance [24]. Instead, then concentrating on a single machine learning algorithm, they compared several approaches, like logistic regression, decision trees, and Support Vector Machines (SVM), in conjunction with DEA. The experimental results revealed that the hybrid model incorporating logistic regression showed the lowest accuracy, decision trees and SVM had the best accuracy among the evaluated methods [25]. Khan and Dhingra [26] concentrated on the pharmaceutical industry's selection of robust and sustainable suppliers to address the challenges posed by globalization and heightened demand volatility.

In contrast, Khan et al. utilized a gradient boosting classifier for improved decision-making after integrating the Supply Chain Operations Reference (SCOR) 4.0 model with the Best–Worst Method (BWM) [27].

3 Understanding Influencing Factors for AI Adoption

A complex web of interrelated elements that fall under the categories of technological, organizational, economic, and regulatory aspects affect the manufacturing areas taking of artificial intelligence (AI). It is essential to comprehend these components in sequence to make wise decisions and successfully integrate AI.

3.1 Technological Factors

Data Infrastructure Readiness: The availability and standard of particulars are fundamental for AI applications. Manufacturers must assess their data collection, storage, and processing capabilities to support AI algorithms effectively. A robust data infrastructure enables accurate predictive maintenance, quality control, and process optimization [28].

Integration with Existing Systems: Seamless integration of AI solutions with current manufacturing systems is essential to avoid disruptions. Compatibility with existing machinery, software, and workflows ensures that AI enhances rather than hinders operations.

3.2 Organizational Factors

Workforce Skills and Training: The successful implementation of AI requires a workforce proficient in AI technologies. Investing in training programs to upskill employees is vital to bridge the knowledge gap and foster a culture of innovation [29].

Leadership Support and Strategic Vision: Strong leadership commitment to AI initiatives is crucial. An understandable calculated vision and support from top management facilitate resource allocation and drive organizational change [30].

3.3 Economic Factors

Cost of Implementation: The financial investment required for AI adoption, including hardware, software, and training, can be substantial. Organizations must evaluate the cost–benefit ratio to ensure economic viability.

Return on Investment (ROI): Assessing the potential ROI is essential to justify AI investments. AI can lead to cost savings through improved efficiency and reduced downtime, but these benefits must be quantified [31].

3.4 Regulatory and Ethical Factors

Compliance with Regulations: Adhering to industry standards and regulations is mandatory. AI applications must observe with data conservation rule, safety excellence, and other relevant regulations to avoid legal repercussions.

Ethical Considerations

Ethical issues, like statistics privacy and the impact on employment, must be addressed. Translucent AI exercises and stakeholder affiance are necessary to mitigate ethical concerns [32].

Manufacturing companies looking to implement AI technologies must comprehend these contributing aspects. Organizations can create well-informed plans that complement their operational objectives and market demands by methodically assessing these characteristics. Effective decision-making in the complex world of AI adoption is made possible by Multi-Criteria Decision-Making (MCDM) tools, which provide structure methodologies to assess and rank these considerations.

4 Application of MCDM Techniques in AI Adoption

The acquisition of Artificial Intelligence (AI) in the manufacturing sector involves evaluating multiple, often conflicting factors such as cost, technological readiness, and organizational impact. Multi-Criteria Decision-Making (MCDM) techniques provide structured methodologies to estimate and prioritize these factors, facilitating informed decision-making. Every MCDM techniques offers unique strengths and is suitable for different contexts within AI adoption in manufacturing. The alternative method subjected on factors such as the nature of the decision complication, accessibility of data, and the precise objectives of the organization. By applying these techniques, manufacturers can systematically evaluate and prioritize the myriad factors influencing AI adoption, most significant to inform and effective excavation.

5 Conclusion

The amalgamation of Artificial Intelligence (AI) into the manufacturing sector holds significant promise for enhancing efficiency, productivity, and innovation. However, the adoption of AI is influenced by a numerous factors, inclusive technological readiness, economic considerations, organizational culture, and regulatory compliance. Effectively navigating these factors necessitates robust decision-making frameworks, and Multi-Criteria Decision-Making (MCDM) techniques have emerged as valuable tools in this context.

MCDM techniques offer structured methodologies to asses and ranking numerous conflicting criteria balanced decision-making. One and all techniques provides unique strengths and is suitable for different contexts within AI adoption in manufacturing. Additionally, the use of AI in smart manufacturing has been linked to improved multi-criteria decision-making, enabling real-time adjustments based on customer preferences and production capabilities.

Despite the advantages, the application of MCDM techniques in selecting influencing factors for AI adoption in manufacturing is not without challenges. These include the complexity of criteria selection and weighting, data availability and quality, the effective essence of the production environment, and the integration of MCDM methodologies into existing organizational processes. Addressing these challenges requires the development of adaptive MCDM models, enhanced data collection and management systems, integration of fuzzy logic and AI, stakeholder engagement and training, and the founding of continuous keep tracking and feedback mechanisms.

In conclusion, MCDM techniques provide a systematic approach to evaluating and prioritizing the myriad variables affecting AI adoption in the manufacturing region. By leveraging these techniques, manufacturing company can traverse the difficulties of AI acquisition and harness its transformative potential effectively. Future research and practice should focus on addressing the identified challenges and exploring innovative applications of MCDM techniques to further enhance decision-making processes in AI adoption.

6 Challenges and Future Directions

Artificial Intelligence (AI) integration in the manufacturing presents a multifaceted challenge and necessitating the evaluation of numerous influencing factors. However, the application of MCDM in this context is not without its challenges, and several future directions can be pursued to enhance its effectiveness.

Challenges

Complexity of Criteria Selection and Weighting: Identifying and assigning appropriate weights to the myriad variables affecting AI adoption like technological readiness, cost implications, organizational culture, and regulatory compliance can be intricate. The subjective nature of weighting can introduce biases, potentially skewing decision outcomes.

Data Accessibility and Quality

MCDM techniques rely heavily on accurate and comprehensive data. In the manufacturing sector, obtaining high-quality data on factors like workforce readiness or technological infrastructure can be challenging, leading to potential inaccuracies in decision-making.

Dynamic Industry Environment: The quick advancement of artificial intelligence innovations and manufacturing processes means that influencing factors can change swiftly. Static MCDM models may not adequately capture these dynamics, necessitating frequent updates to decision frameworks.

Integration with Organizational Processes: Embedding MCDM methodologies into existing organizational decision-making processes can be complex. Resistance to replace, lack of expertise, and resource constraints can hinder effective implementation.

Future Directions

Development of Adaptive MCDM Models: To address the dynamic nature of the manufacturing sector, there is a need for adaptive MCDM models that can evolve with changing industry conditions. These models' responsiveness can be improved by integrating machine learning techniques and real-time data analytics.

Enhanced Data Collection and Management: Investing in robust data collection and management systems is crucial. Utilizing the precious devices and advanced analysis of data can improve the quality and availability of data, thereby enhancing the accuracy of MCDM applications.

Integration of Fuzzy Logic and AI: Incorporating fuzzy logic into MCDM can help manage uncertainties and ambiguities inherent in decision-making. Additionally, integrating AI techniques can automate and optimize the decision-making process, making it more efficient.

Stakeholder Engagement and Training: Engaging stakeholders across all organizational levels and providing comprehensive training can facilitate the integration of MCDM techniques. For AI adoption to be effective, a culture that prioritizes data-driven decision-making must be established.

Continuous Monitoring and Feedback Loops: Establishing mechanisms for continuous monitoring and incorporating feedback loops can ensure that MCDM

models remain relevant and effective. Regular assessments and updates can help organizations adapt to new challenges and opportunities.

By addressing these challenges and pursuing these future directions, manufacturing organizations can enhance their decision-making processes regarding AI adoption. Leveraging advanced MCDM techniques will enable them to direct the difficulties of the modern manufacturing landscape effectively.

References

1. Saaty, T.L.: How to make a decision: the analytic hierarchy process. Eur. J. Oper. Res. **48**(1), 9–26 (1990)
2. Triantaphyllou, E.: Two new cases of rank reversals when the AHP and some of its additive variants are used. J. Multi Crit. Decis. Anal. **10**(1), 11–25 (2001)
3. Zavadskas, E.K., Turskis, Z., Vilutiene, T.: Multiple criteria analysis of foundation instalment alternatives by applying additive ratio assessment (ARAS) method. Arch. Civil Mech. Eng. **10**(3), 123–134 (2010)
4. Govindan, K., Rajendran, S., Sarkis, J., Murugesan, P.: Multi criteria decision making approaches for green supplier evaluation and selection: a literature review. J. Clean. Prod. **98**, 66–83 (2015)
5. Kusiak, A.: Smart manufacturing. Int. J. Prod. Res. **56**(1–2), 508–517 (2018)
6. Ishizaka, A., Labib, A.: Review of the main developments in the analytic hierarchy process. Expert Syst. Appl. **38**(11), 14336–14345 (2011)
7. Yu, W., Chavez, R., Jacobs, M.A., Feng, M.: Data-driven supply chain capabilities and performance: a resource-based view. Transp. Res. Part E Logist. Transp. Rev. **114**, 371–385 (2018)
8. Lu, W., Ip, W.H.: A multi-criteria decision analysis approach to sustainable supplier selection. Int. J. Prod. Econ. **135**(1), 523–529 (2011)
9. Sarkis, J., Dhavale, D.G.: Supplier selection for sustainable operations: a triple-bottom-line approach using a Bayesian framework. Int. J. Prod. Econ. **166**, 177–191 (2015)
10. Dweiri, F., Kumar, S., Khan, S.A., Jain, V.: Designing an integrated AHP based decision support system for supplier selection in automotive industry. Expert Syst. Appl. **62**, 273–283 (2016)
11. Taylan, O., Karagöz, İ: An integrated fuzzy AHP–DEA approach for sustainable supplier selection. Appl. Soft Comput. **27**, 169–179 (2014)
12. Chai, J., Liu, J.N.K., Ngai, E.W.T.: Application of decision-making techniques in supplier selection: a systematic review of literature. Expert Syst. Appl. **40**(10), 3872–3885 (2013)
13. Xia, W., Wu, Z.: Supplier selection with multiple criteria in volume discount environments. Omega **35**(5), 494–504 (2007)
14. Alinezhad, A., Khalili, J.: Hybrid MCDM techniques for AI-based manufacturing process selection. Math. Comput. Simul **167**, 55–70 (2019)
15. Bevilacqua, M., Ciarapica, F.E., Giacchetta, G.: A fuzzy-QFD approach to supplier selection. J. Purch. Supply Manag. **12**(1), 14–27 (2006)
16. Kuo, R.J., Lin, Y., Lin, C.P.: Integration of fuzzy AHP and fuzzy TOPSIS for evaluating the optimal selection of supplier. Appl. Math. Model. **36**(11), 5273–5284 (2012)
17. Liu, Y., Huang, G., Ke, Q.: Fuzzy DEA-based MCDM method for evaluating AI applications in manufacturing. Int. J. Adv. Manuf. Technol. **73**(1–4), 537–548 (2014)
18. Deng, H., Yeh, C.H., Willis, R.J.: Inter-company comparison using modified TOPSIS with objective weights. Comput. Oper. Res. **27**(10), 963–973 (2000)
19. Ho, W., Xu, X., Dey, P.K.: Multi-criteria decision-making approaches for supplier evaluation and selection: a literature review. Eur. J. Oper. Res. **202**(1), 16–24 (2010)

20. Awasthi, A., Chauhan, S.S., Goyal, S.K.: A hybrid approach integrating DEA and AHP for supplier selection. Comput. Ind. Eng. **62**(1), 121–133 (2010)
21. Cheng, E.W., Li, H.: A review of the application of analytic hierarchy process (AHP) in construction. Int. J. Project Manage. **20**(5), 387–405 (2002)
22. Golmohammadi, D.: Neural network application for fuzzy multi-criteria decision making problems. Int. J. Prod. Econ. **131**(2), 490–504 (2011)
23. Fallahpour, A., Olugu, E.U., Musa, S.N., Khezrimotlagh, D., Wong, K.Y.: An integrated model for green supplier selection under fuzzy environment: application of data envelopment analysis and genetic programming approach. Neural Comput. Appl. **27**, 707–725 (2016)
24. Aggarwal, I., Gunreddy, N., Rajan, A.J.: A hybrid supplier selection approach using machine learning and data envelopment analysis. In: 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1–5. IEEE (2021).
25. Khan, M.M., Bashar, I., Minhaj, G.M., Wasi, A.I., Hossain, N.U.I.: Resilient and sustainable supplier selection: an integration of SCOR 4.0 and machine learning approach. Sustain. Resil. Infrastr. **8**(5), 453–469 (2023)
26. Khan, S., Dhingra, A.: SCOR model and best-worst method for supplier selection in pharmaceuticals. J. Manuf. Syst. **60**, 19–36 (2021)
27. Vahdani, B., Mohammadi, A.: A bi-objective approach for sustainable supplier selection using fuzzy compromise programming. J. Manuf. Syst. **34**, 152–161 (2015)
28. Ustun, O., Demirtas, E.A.: An integrated multi-objective decision-making process for AI-driven supplier evaluation. Omega **36**(5), 760–775 (2008)
29. Yazdani, M., Zarate, P., Coulibaly, A.: Multi-criteria decision making for sustainable supplier selection. J. Clean. Prod. **142**, 3728–3740 (2017)
30. Azadi, M., Farzipoor Saen, R.: A hybrid AI and MCDM approach for evaluating manufacturing readiness. Int. J. Adv. Manuf. Technol. **69**(1–4), 585–594 (2013)
31. Song, W., Wang, L.: Using hybrid MCDM methods for AI-driven green supply chain management. J. Clean. Prod. **207**, 1046–1060 (2019)
32. Leão, C.P., Ribeiro, R.A.: Intelligent decision systems in manufacturing. Int. J. Prod. Res. **52**(13), 3974–3994 (2014)

Digital Governance RFP Analysis: BERT Model Adoption for Document Classification



Manisha Tiwari, Padmanabhan Aital, and Padmaja Joshi

Abstract The tender process by client organizations in the Digital governance domain is a well-established mechanism to procure cost effective and high quality products and services in a fair and transparent manner. On the bidder's end, one of the significant steps involves an in-depth requirement analysis of the Request for proposal documents floated by the client in various aspects, which is a time-intensive task. By assessing these fundamental aspects from the service provider's perspective, in the next stage, the bidder frames its bid to be submitted to the client for evaluation. This research proposes a framework to automate the bid development process in response to the Request for proposal for the software development projects in the Digital governance domain. The methodology involves Bidirectional Encoder Representations from Transformers transformation on a curated dataset, with noise reduction followed by meticulous feature extraction with an aim to tokenize the keywords and enhance the interpretability to streamline the selection process. Alongside, the model's performance is rigorously monitored using key metrics to gauge the precision, accuracy, an F1-score and recall. The monitoring outcomes underscore the robustness and the effectiveness of the proposed model and its capability to streamline the Digital governance bidding process by increasing the accuracy and the efficiency of the bid document classification process.

Keywords Digital governance · RFP documents · Machine learning · Text classification

M. Tiwari (✉)
Department of Computer Engineering MPSTME, NMIMS University, Mumbai, India
e-mail: manisha.tiwari@nmims.edu

P. Aital
Department of Technology Management MPSTME, NMIMS University, Mumbai, India
e-mail: padmanabha.aital@nmims.edu

P. Joshi
C-DAC, Mumbai, India
e-mail: padmaja@cdac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_24

1 Introduction

With the evolution of digitization, the e-tendering process has evolved as the digital form of managing the tendering process, typically used in the government sector and by large companies to procure goods and services [1]. The process starts by the company publishing the tender notice, then the interested parties respond to the Request for Proposal (RFP) providing various document requested in the RFP. The response for the RFP includes, company information, technical proposal, costing etc. The tender is awarded based on various evaluations based on the submitted proposal, followed by the acceptance and evaluation of bids leading to the award of contracts, the entire procedure is conducted online [2].

One of the most important aspects of the bidding process is the accurate understanding of the project requirements and the incorporating the details in the bid formulation. However, the criticality of the RFP evaluation for furnishing accurate and relevant information in the bid forms the core activity seeking deployment of considerable human intervention by the bidding firm to understand the requirements and prepare a bid document for submission. As depicted in Fig. 1, the analysis of RFP (Request for Proposal) documents plays a critical role as part of the e-tendering bidding process [3–5] in ensuring the successful bidding for the winning the contract. As part of this framework, the RFP document is first analyzed using document classification feature of classifying text documents into preset classes or categories according to the content [6]. This task involves analyzing content of the RFP document and classifying it into various categories. For example, the documents could be classified into domain-specific categories like “insurance”, “banking” etc., based on their content.

A. *E-Tendering Bidding Process*

The aim of the proposed work is to build a classification system for Digital governance RFP documents by using advanced machine learning techniques.

Natural Language Processing and information retrieval methodologies enables to extract useful information from textual documents, automation using these paradigms enable automatic classification of text documents as well. The advent of deep learning and transformer-based models have resulted in significant advancements in this field. This literature review aims to study the recent developments in the automated document classification sub-domain, with a special emphasis on BERT (Bidirectional Encoder Representations from Transformers) and other similar models. The existing literature on the document classification provides a bouquet of methods for classifying text documents, wherein each of the above methods apply to diverse datasets. To explore the different methods utilized in document classification, we have outlined the following techniques

(1) **Traditional Machine Learning Approaches**

The conventional methodologies for document classification were based on the traditional machine learning algorithms such as Support Vector Machines (SVM),

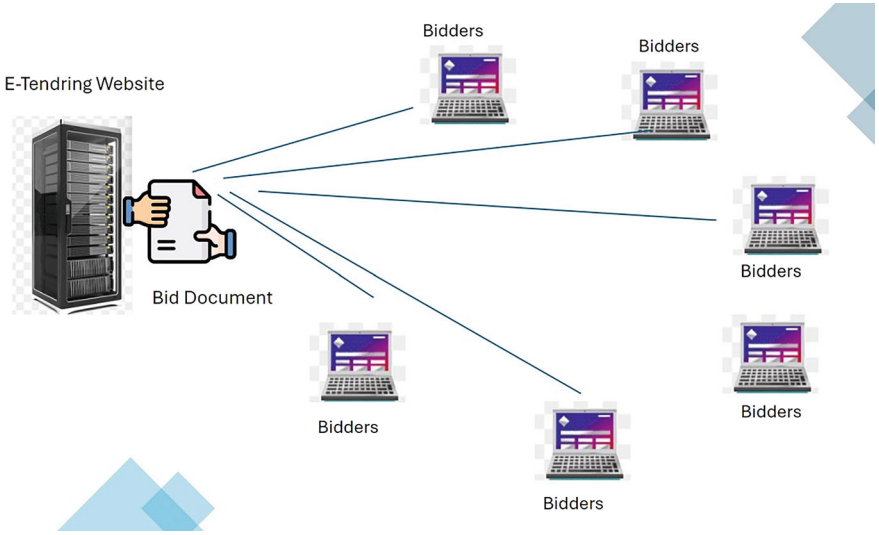


Fig. 1 E-tendering bidding process

Random Forests and Naive Bayes etc. These methods generally used bag-of-words or Term Frequency-Inverse Document Frequency (TF-IDF) representations of the documents and their content as input features.

Joachims [7] established the effectiveness of SVMs for text categorization by demonstrating that they outperformed other classifiers relevant at the time, particularly on datasets with multiple features. McCallum and Nigam [8] provided a wide-ranging comparison of a variety of event models for Naive Bayes-based text classification and validated its efficiency and effectiveness for specific types of text data. In a separate proposition, the frameworks based on Random Forests, introduced by Breiman [9], were applied to text classification tasks. Yang and Pedersen [10] did a comparison of the feature selection methods in statistical learning for text categorization. While these approaches were mostly applicable for many types of activities in this framework, they frequently struggled to capture the contextual details and semantic linkages in text. In most cases, they treated each word as a separate unit, losing track of word order and the underlying context. They also frequently suffered from the curse of dimensionality when dealing with big vocabulary sets [11]. Also, Xu et al. [12] expounded that Random Forests could perform competitively with SVMs on certain document classification tasks, particularly in the case of high-dimensional data. Despite these limitations, the traditional machine learning approaches remain important in certain situations, notably when dealing with tiny datasets or when interpretability is critical [13].

(2) **Deep Learning Models** These models have revolutionized text classification by automatically learning features from raw text, capturing intricate semantic relationships, and excelling at handling large-scale datasets. Deep learning

approaches, particularly those based on transformers and pre-trained language models, generally outperform classical methods in terms of accuracy and ability to understand context. However, they come with increased computational costs, reduced interpretability, and a greater need for large training datasets. The choice between classical and deep learning methods ultimately depends on the specific requirements of the task, available resources, and the nature of the dataset at hand. Deep learning architectures have significantly advanced text classification capabilities. Recurrent Neural Networks (RNNs) [14], particularly their LSTM [14, 15] and GRU variants, excel at capturing sequential information in text. Convolutional Neural Networks (CNNs), originally designed for image processing, have been adapted for text, effectively identifying local patterns and features. Attention mechanisms, especially in Transformer models, have revolutionized the field by allowing models to focus on relevant parts of the input, leading to improved performance on various NLP tasks. Pre-trained language models like BERT, GPT, and XLNet have further pushed the boundaries, leveraging massive amounts of data to learn general language representations that can be fine-tuned for specific classification tasks. Hybrid architectures, combining elements from multiple approaches, often yield state-of-the-art results by capitalizing on the strengths of different model types. These diverse architectures provide a rich toolkit for addressing various text classification challenges, each offering unique strengths in processing and understanding textual data [15]. Analyzes different model types including BERT, ELMo + LSTM, CNN-based and LSTM-based model on Stanford Sentiment Treebank (SST-2), AG News topic classification dataset (Fig. 2).

Minaee et al. [14] provides quantitative comparisons of model performance on several benchmark datasets, showing the progression of state-of-the-art results over time. Overall, deep learning models, especially pre-trained Transformer-based models like BERT, have significantly advanced the state-of-the-art on text classification tasks. “The paper reviews over 150 deep learning models for text classification, including: Recurrent neural networks (RNNs) and LSTMs, Convolutional neural networks (CNNs), Attention mechanisms Transformers and pre-trained language models like BERT, Graph neural networks, Memory-augmented networks Hybrid models combining multiple architectures

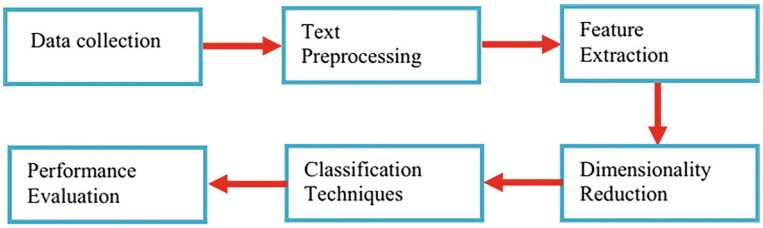


Fig. 2 Stages of text classification [16]

(3) **Feature Engineering Methods** In Machine learning process, features are the granular measurable properties or characteristics of the phenomenon under study.

B. *Traditional Feature Extraction Methods*

(a) Bag of Words: This method represents text data as the word frequency in a document. Despite its simplicity, BoW has been popular in text classification activities [17].
(b) TF-IDF (Term Frequency-Inverse Document Frequency): This measure analyzes the significance of a word in a document as compared to a collection of documents. Salton and Buckley [18] demonstrated its effectiveness in information retrieval tasks, and it has since been widely adopted in document classification.

C. *Word Embeddings*

Word embeddings provide condensed word vector representations to capture semantic meanings. Two prominent approaches are:

(a) Word2Vec: Introduced by Mikolov et al., this method involves neural networks to train on word representations.
(b) GloVe (Global Vectors for Word Representation): GloVe is a count-based model that derives representations from global word-word co-occurrence statistics, Pennington et al.

These embedding techniques have shown significant improvements in various NLP tasks, including document classification [4].

D. *Dimensionality Reduction*

Dimensionality reduction techniques are often employed to simplify models and reduce computational cost while retaining as much information as possible from the original data. Some common methods include:

(a) Principal Component Analysis (PCA): While not specifically designed for text data, PCA has been applied to document classification tasks [19].
(b) Latent Semantic Analysis (LSA): This technique, introduced by Deerwester et al. [20], uses singular value decomposition to identify patterns in the relationships between terms and concepts in unstructured text (Fig. 3).

E. *Advanced Techniques*

(a) Topic Modeling: Approaches such as Latent Dirichlet Allocation (LDA), introduced by Blei et al. [21], can extract higher-level themes from texts. These topics can serve as features for classification tasks.
(b) Contextual Embeddings: Recent advancements in deep learning have led to contextual embedding models like BERT [6], which capture word meanings based on their surrounding context. These models have achieved state-of-the-art results on many document classification benchmarks.

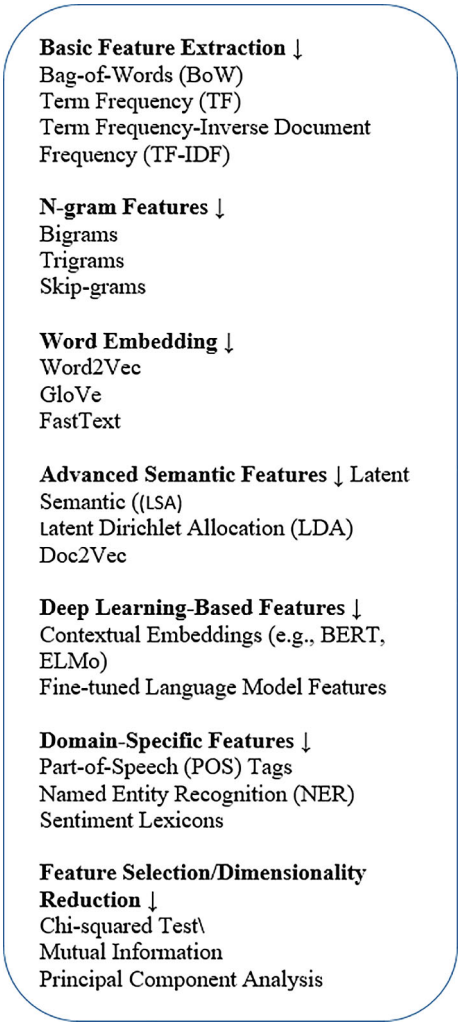


Fig. 3 Feature engineering methods for text classification

(c) Feature Selection: Methods like chi-squared tests or mutual information can help identify the most relevant features. Forman [22] provided a comparative study of feature selection methods for text classification.

(4) **Pre-Trained Models** The evolutionary pre-trained models for text classification have increasingly occupied centre stage due to their strength to leverage knowledge from large datasets and transfer it to specific tasks. A more focused analysis of some of the popular pre-trained models for classifying texts are given below:

(a) BERT (Bidirectional Encoder Representations from Transformers)

- Widely used for various NLP tasks, including text classification
- Excels at capturing context in both directions
- Available in different sizes (Base, Large) and domain-specific versions (e.g., SciBERT for scientific text)

(b) RoBERTa (Robustly Optimized BERT Approach)

- An optimized version of BERT with enhanced training approach
- Generally achieves better performance than BERT on many tasks

(c) DistilBERT

- A lighter, faster version of BERT
- Retains much of BERT's performance apart from being more fitting for application in resource-constrained environments.

(d) XLNet

- Often outperforms BERT on various tasks
- Particularly good at handling longer sequences

e) ALBERT (A Lite BERT)

- A memory-efficient version of BERT
- Suitable for scenarios with limited computational resources

(f) ELECTRA

- More efficient pre-training approach
- Often achieves comparable or better results than BERT with less computational cost

(g) ULMFiT (Universal Language Model Fine-tuning)

- One of the earlier transfer learning methods for text classification
- Still relevant, especially for tasks with limited data

(h) FastText

- While not a transformer model, it's worth mentioning for its speed and efficiency
- Particularly good for large-scale text classification tasks

Table 1 provides the summary of the BERT models used for text classification/ document classification studies.

2 Research Question

This paper aims to address the following questions, making the task of document classification easier to understand and adaptable across different fields of study.

Table 1 Summary of BERT models and datasets in text classification studies

References	BERT model	Dataset	Key findings
[23]	Uncased BERT-base model with 12-layer transformers and hidden size of 768	AAPD, Reuters, IMDB, Yelp-2013	The paper proposes a hierarchical BERT model (HAd-aBERT) that uses BERT as a local encoder to extract high-quality local features from document containers. It also introduces an adaptive fine-tuning strategy to dynamically select BERT layers for fine-tuning. HAdaBERT outperformed several baseline models including DocBERT, RoBERTa, and ALBERT across all datasets. For example, it achieved 3.1%, 1.2%, 1.7%, and 1.9% improvements over DocBERT on Yelp-2013, IMDB, AAPD, and Reuters test sets respectively
[24]	Japanese BERT pre-trained model from Kyoto University with 12/24-layers of multi-head attention	Amazon review dataset (books, DVD, music domains) from Webis-CLS-10	The research proposes using the mean vector of word embeddings generated by BERT as a feature vector for document categorization, rather than using (CLS) token embedding. They combine this BERT-based feature vector with a bag-of-words (BOW) feature vector. Proposed method (vb; vm) (combining BOW and BERT features) achieved best accuracy of 0.8271 averaged across 3 domains Outperformed using just (CLS) embedding (0.7658) or BOW features alone (0.8058) Also outper-formed fine-tuning BERT (0.7904)
[17]	BERT-base and BERT-large models	Yelp 2014 reviews, Reuters-21578, IMDB reviews, arXiv Academic Paper dataset (AAPD)	The research improves BERT's document classification performance by putting a fully linked layer above the (CLS) token's final hidden state. They fine-tune the entire model from end to end. BERT-large obtains cutting-edge performance across all four datasets, followed by BERT-base. They also distill BERT-large into a smaller BiLSTM model (KD-LSTM) that achieves comparable performance to BERT-base with much faster inference
[18]	Used BERT-Base model	CSAT dataset (customer satisfaction prediction), 20 newsgroups (topic identification) Fisher Phase 1 corpus (topic identification)	The paper proposes two extensions to BERT for long document classification: RoBERT: Splits input into segments, gets BERT representations for each, then passes them through an LSTM layer ToBERT: Similar to RoBERT but uses another Transformer instead of LSTM on top of BERT segment representations RoBERT: LSTM trained with Adam optimizer, initial learning rate 0.001 ToBERT: Transformer trained with BERT's Adam optimizer, initial learning rate 5e−5 Fine-tuned BERT on segment-level data before using in RoBERT/ToBERT ToBERT outperformed RoBERT on all datasets Achieved state-of-the-art 95.48Significant improvements over baselines on CSAT and Fisher datasets

(continued)

Table 1 (continued)

References	BERT model	Dataset	Key findings
[19]	Used pre-trained BERT model from ktrain Python module (likely BERT-base)	IMDB movie reviews (sentiment analysis), RealOrNot tweets (disaster classification) Portuguese news articles (topic classification), Chinese hotel reviews (sentiment analysis)	The paper compares BERT to traditional machine learning approaches using TF-IDF features for text classification tasks. They use a pre-trained BERT model from the ktrain Python module and fine-tune it on different datasets. BERT outperformed traditional ML approaches on all datasets: IMDB: BERT 93.87% vs best traditional 90.07% RealOrNot tweets: BERT 83.61% vs H2OAutoML 78.75% Portuguese news: BERT 90.93% vs AutoML 84.80%Chinese hotel reviews: BERT 93.81% vs AutoML 73.99% BERT was also much simpler to implement compared to traditional approaches
[20]	Standard BERT, ECHR-Legal-BERT, Harvard-Law-BERT, RoBERTa, Bi Longformer	ECHR violation dataset (multi-label classification) overruling task dataset (binary classification)	The paper investigates how to effectively adapt BERT for legal document classification, focusing on handling long and extensive documents and the significance of pre-training on in-domain data. It compares different BERT-adpating techniques to large documents and various pre-trained BERT models. For long documents, various techniques are compared: Truncating from rear (RR) or front (RF), Splitting into chunks and using mean/max pooling, No special preprocessing mentioned for shorter documents. Pre-training on in-domain data (legal documents) improved performance For long documents, BigBird and Longformer performed best Truncation hurt performance on long documents For shorter documents, standard BERT variants performed well
[21]	DistilBERT	Internal dataset of customer transaction descriptions (55,038 samples, 10 classes), TREC-6 (5952 samples, 6 classes) AG's News Corpus (127,600 samples, 4 classes)	For multi-class text classification, the paper explores using BERT with various active learning strategies. They fine-tune DistilBERT in each active learning iteration. Fine-tuned for 1 epoch per active learning iteration Learning rate, Batch size: 500 (internal dataset), 100 (public datasets), Number of iterations: 13 (internal dataset), 20 (public datasets), Active learning with BERT reduced labeling cost by 85% while achieving similar F1 scores to full supervised training on the internal dataset No single active learning strategy consistently outperformed others across datasets EGL strategy performed reasonably well across datasets, Random, EGL and DAL captured diverse and representative samples with lower class bias, but EGL and DAL had longer runtimes

(continued)

Table 1 (continued)

References	BERT model	Dataset	Key findings
[22]	RoBERTa-large	AG's news (4 classes), snippets (8 classes), news title (7 classes)	The paper proposes an approach based on prompt-learning for short text classification involving BERT. They use RoBERTa-large as the pre-trained language model and fine-tune it using prompts. The key innovation is expanding the label word space using both the class names and short text itself, retrieving related concepts from an external knowledge graph (Probase). Few-shot setting with 5, 10, 20 labeled examples per class along with 1000 unlabeled examples per class as support set, fine-tuned for 1 epoch per active learning iteration Learning rate: not specified, their method outperforms baselines on all three datasets, especially in few-shot settings. Achieves up to 6 percentage point accuracy improvement over the state-of-the-art peers. More stable performance across varied datasets when equated to other prompt-based methods
[25]	The paper uses RoBERTa, which is known as a robustly optimized BERT pre-training Approach. Specific architectural details are not provided	The dataset consists of 443 documents from ZFW (Fort Worth Air Route Traffic Control Center), mostly Letters of Agreement (LOAs) but also including some other aviation procedure documents. The dataset was split into 222 civil LOA documents and 271 non-civil LOA documents	The paper uses RoBERTa, which is a variant of BERT. It is used as one of the embedding techniques for document classification. The authors also use transfer learning with RoBERTa. RoBERTa and transfer learning with RoBERTa (TL RoBERTa) did not perform well compared to other methods, especially with neural network models. The authors note that while they were able to tune it well for cross-validation, it still overfit on the cross-validation data. The test F1-Macro scores for RoBERTa ranged from 0.359 to 0.838 across different classification models, while TL RoBERTa scores ranged from 0.359 to 0.838

(continued)

Table 1 (continued)

References	BERT model	Dataset	Key findings
[26]	Used pre-trained BERT base as well as large cased models Added a single linear layer on top for classification	PROMISE NFR dataset, Relabeled version of PROMISE NFR dataset, new dataset created by authors for functional requirement subclassification	The authors propose NoRBERT, which fine-tunes BERT for requirements classification tasks. They use both the base and large pre-trained BERT models and then adjust them on various requirements classification datasets. Used AdamW optimizer with learning rate of 2e-05, Batch size of 16, Sequence length of 128 for base model, 50 for large model Experimented with 10–64 epochs. Outperformed or matched state-of-the-art on most tests. Significantly better performance on unseen projects (15 percentage point improvement in F1-score on one task) F1-scores 94% on binary F/NFR classification, F1-scores 87% on NFR subclass classification, F1-scores of up to 92% on new functional requirement subclassification task. The key advantage was improved generalization to unseen projects compared to previous approaches
[27]	Ensemble of 4 BERT-related models were mentioned	Novel text preprocessing method (V3) that groups anchor-context pairs creates structured input with tokens like (CLS), (SEP), (TAR) as-signs scores to each token	Uses an ensemble of 4 BERT-related models: DeBER-TaV3 by He et al., Microsoft's DeBERTa-v3-large, Moritz Laurer's DeBERTa-v3-large-mnli-fever-anli-ling-wanli, Anferico's BERT for patents by Lee and Hsiang for patents, and Google's ELECTRA-large-discriminator, Fine-tunes these models on patent similarity matching task Uses weighted averaging of predictions from individual models Max sequence length: 400, Learning rate: 1e-5, Uses BCELoss and Adversarial Weight Perturbation, fourfold cross-validation, Pearson Correlation Coefficient between predicted and actual similarity scores, Best single model (DeBERTa-v3-large) score: 0.8512, Final ensemble score: 0.8534 The key innovation is the ensemble approach and the unique text pre-processing method specifically used for patent documents

(continued)

Table 1 (continued)

References	BERT model	Dataset	Key findings
[28]	Uses 9 different BERT-based pre-trained models as mentioned above	COVID-19 fake news dataset (10,202 samples), extremist-non-extremist dataset (21,186 tweets) COVID-19 English tweets dataset (6,925 tweets)	Uses an ensemble of 9 BERT-based models including ALBERT-base-v2, BERT-base, BERT-large, Distil-BERT, RoBERTa-base, RoBERTa-large, Electra-small, XLM-RoBERTa-base and BART-large. Fine-tunes these pre-trained models on text classification tasks, COVID-19 fake news: best accuracy 99.71% (RoBERTa-base) COVID-19 tweets: best accuracy 98.83% (BART-large), Extremist-non-extremist: best accuracy 99.71% (BERT-base/large) The BERT-based transfer learning approach outperformed previous methods on all three datasets
[29]	Used standard BlueBERT-Base model (12-layer, 768-hidden, 12-heads)	MIMIC-III discharge summaries Cancer pathology reports from SEER registries	Used BlueBERTBase pretrained on clinical notes from MIMIC-IIIand abstracts from PubMed. Tested 4 methods to adapt BERT for long clinical documents: using first 510 tokens only, max pooling over chunks, target attention, and multilabel attention. BERT generally did not outperform simpler CNN and HiSAN baselines HiSAN achieved best performance on most tasks. BERT suffered with extensive clinical texts, most likely because to attention dilution and difficulties with subword tokens. The key finding is that pretrained BERT models may not be apt for tasks relted to clinical text classification compared to simpler architectures like CNNs or HiSANs

(continued)

Table 1 (continued)

References	BERT model	Dataset	Key findings
[30]	Used standard BERT models from Hugging Face (base and large versions)	PROMISE exp dataset-an augmented version of the tera-PROMISE NFR dataset, contains 444 functional requirements and 525 non-functional requirements	Used BERT for classifying software requirements among the functional or non-functional categories. Compared BERT with RoBERTa and DeBERTa models Used BertForSequenceClassification from Hugging Face for text classification Experimented with bert-base-uncased, bert-large-uncased, deberta-v3-base, roberta-base and roberta-large models. For binary classification, bert-base-uncased and roberta-base achieved 88.6% accuracy. For 12-class classification, bert-base-uncased achieved 76.2% accuracy. Larger models (bert-large, roberta-large) did not improve performance. The key finding is that BERT models, especially bert-base-uncased, perform well for software requirements classification, but larger models do not necessarily improve accuracy

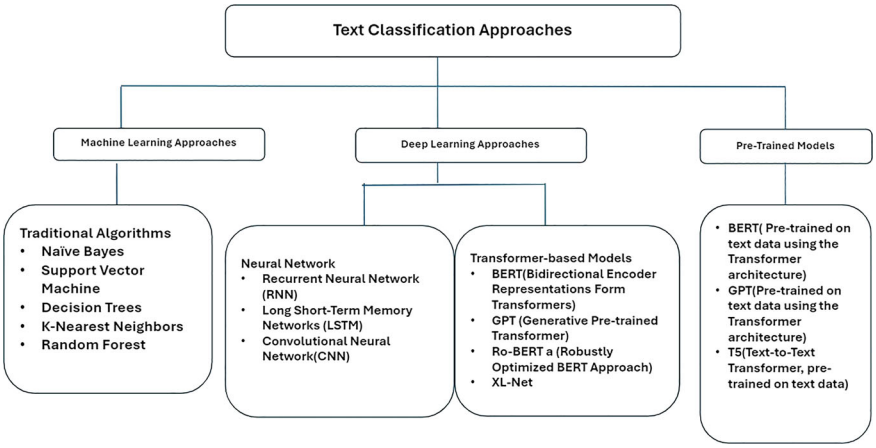


Fig. 4 Text classification approaches

Research Question 1 (RQ1)—What are the various techniques involved in text classification and which one will be suitable for a specific task of e-tendering RFP document classification?

Research Question 1 (RQ2)—Does BERT classification suitable for automated document processing of RFP documents in the e-tendering process?

Following papers elaborate on the various text classification techniques along with usage of BERT classifiers in various document classification (Fig. 4).

Ankit [16] uses Multinomial Naïve Bayes classification for document classification process. It compares different feature representation methods including Binary Vectorizer, Count Vectorizer, and TF-IDF Vectorizer. It uses 20 Newsgroups dataset which contains documents from 20 different categories. TF-IDF vectorizer performed best with 78% accuracy, 0.80 precision, 0.78 recall, and 0.77 F1-score when stop words were removed TF-IDF outperformed Count vectorizer by 4% and Binary vectorizer by 6% in accuracy.

The categorization with KNN and Naive Bayes was performed on the TREC Legal Track dataset containing over 3000 text documents and over twenty categorization classes [13].

A semi-supervised machine learning method known as the co-training strategy is proposed, which uses both labeled and unlabeled data to improve the performance of a learning model. They represented the documents in three ways: TF-IDF using the bag-of-words method, topic distribution using latent Dirichlet allocation (LDA), and neural-network-based document embedding referred to as document to vector (Doc2Vec) [31].

The Arabic text from the document was classified using various methods, including Bernoulli Naïve Bayesian (BNB), Logistic Regression (LR), Support vector classifier (SVC), Multinomial Naïve Bayesian (MNB), Linear SVC, and convolutional neural networks (CNN) [32].

Kadhim [33] explores the supervised machine learning techniques for text classification, focusing on Support Vector Machine (SVM), k-Nearest Neighbors (k-NN) and Naïve Bayes on "Reuters-21578 20 Newsgroups, UCI repository datasets. Movie review datasets", the paper discusses general concepts rather than specific training details. It mentions the importance of sufficient training data and feature selection.

Kim et al. [31] proposes a multi-co-training (MCT) method for classifying documents. It uses three document representation methods (TF-IDF, LDA, Doc2Vec) and combines them in a semi-supervised co-training framework. Three classifiers are trained on the different representations and help improve each other by adding confidently predicted unlabeled examples to the training sets. The Reuters-21578 20 Newsgroup Ohsumed Reuters-50-50 SemEval, MCT outperformed the baseline supervised and semi-supervised methods in most cases. The performance improvement was more evident with Naive Bayes classifier MCT was robust to changes in labeled data ratio and feature dimensions.

A. *Gaps in Research*

Despite the advancements in machine learning (ML) and natural language processing (NLP) models for text classification, several gaps remain in the context of classifying Request for Proposals (RFPs), particularly in distinguishing between Digital governance and non-Digital governance documents. The following key gaps have been identified:

- **Limited Dataset**: One of the notable gaps in research on Digital governance RFP classification is the lack of comprehensive and publicly available RFP datasets specifically for Digital governance RFP. Despite the critical role of RFPs in procurement and project management, a consolidated dataset of RFP documents that can be readily accessed and used for academic and practical purposes is sparse. This scarcity hinders the evolution of advanced classification models.
- **Under-utilization of Transfer Learning and Fine-tuning Techniques**: While the pre-trained models like BERT and ALBERT have been used for text classification tasks, scarce academic research appears to have been conducted on the effective fine-tuning of these models specifically for RFP classification. Many studies do not detail the full capability of transfer learning techniques, especially in adapting models to the specific linguistic characteristics of RFP documents.

3 Methodology

A. *Data sets*

For this study, we gathered a comprehensive dataset of Digital governance RFP (Request for Proposal) documents from e-procurement portals. The dataset was curated with expert intervention to ensure that the most relevant information, which is crucial for making decisions on whether to proceed with a particular RFP document, is extracted.

Considering the curation of dataset from scratch manual intervention is used for feature selection in document classification for RFP's. It involved domain experts choosing specific features based on their knowledge and understanding of the data and the problem. Manually selected features make models more interpretable, as the selected features are often more meaningful and relevant to the problem. The curated dataset includes RFP titles, organization names, and associated scopes of work which consist of 10-20 sentences. To ensure the dataset's robustness, we included documents from different government departments and agencies including Digital governance documents. The collected data was pre-processed to remove duplicates, inconsistencies, and irrelevant information.

B. *Methods used for Experimentation*

For our dataset, we adopted a comprehensive methodology combining the Pre-trained Transformer models and traditional machine learning models.

(1) **Pre-trained Models** Given the advancements in Natural Language Processing (NLP) techniques, we further explored pre-trained transformer models, specifically BERT (Bidirectional Encoder Representations from Transformers) and its variant, ALBERT (A Lite BERT). These models represent the current state of the art for NLP tasks and are particularly effective for small datasets due to their ability to leverage large-scale pre-training on diverse corpora. By fine-tuning BERT on our specific task, we can capture nuanced language patterns that traditional models might miss. We also utilized BERT embeddings as feature vectors and combined them with Bag-of-Words (BOW) features to enhance the representation of documents. The inclusion of ALBERT allows us to compare its efficiency and effectiveness against BERT, given its design for faster training and lower memory consumption.

(2) **Traditional machine learning models:** Machine Learning Models such as Decision Trees, Support Vector Machines (SVM) and Naive Bayes are well-suited for baseline comparisons and offer insights into the effectiveness of standard algorithms on our dataset. To address the challenge of limited data, we applied data augmentation techniques, specifically Synthetic Minority Oversampling Technique (SMOTE), to artificially increase the dataset size and improve the model's generalizability. Additionally, cross-validation was used to provide reliable estimates of model performance and ensure robustness against overfitting.

We used a number of indicators to assess the models' performance:

Accuracy: The percentage of cases properly classified out of all the instances.

Precision: The percentage of actual positive cases among those that were anticipated to be positive. Remember: The percentage of real positive examples divided by the number of true positive examples.

The F1-score is a balanced indicator of model performance that is calculated as the harmonic mean of precision and recall.

In this research, we have used BERT for the implementation for the proposed model as BERT is typically trained on massive datasets and one of its advantages is that it may be fine-tuned on specific jobs with relatively little amounts of data along with leveraging its pre-trained knowledge. It is also suitable for small datasets in many cases, primarily due to its transfer learning capabilities. The pre-training on large corpora allows BERT to incorporate understanding of general languages, which can then be tweaked for individual tasks with limited data. As part of the implementation, We have firstly employed two state-of-the-art transformer-based models, BERT and ALBERT to extract rich contextual representations from our text data.

Subsequently, we utilized the mean vector of word embeddings output by BERT (bert-base-uncased) as a feature vector for document classification, rather than using the (CLS) token embedding further BERT-based model's performance was enhanced by combining the BERT-based feature vector with a bag-of-words (BOW) feature vector, following the strategy proposed by Tanaka et al. [24]. This hybrid approach leverages both the contextual understanding of BERT and the simplicity and effectiveness of traditional NLP techniques

This version of BERT is the fundamental model, with 12 layers (transformer blocks), 768 hidden units per layer, 12 attention heads, and about 110 million parameter values. The "uncased" designation indicates that the model does not distinguish between capital and lowercase letters, treating "apple" and "Apple" identically.

(3) ALBERT v2 as a separate model: In addition to our BERT-based approach, we separately implemented ALBERT [34] as an alternative model. ALBERT was chosen due to its reported suitability for small datasets, which aligns with our small dataset

This version of ALBERT is a more efficient version of BERT with fewer parameters but designed to retain performance. The albert-base-v2 model includes 12 layers, 128 hidden units for each layer, 12 attention heads, and around 11 million parameters, making it even lighter and faster for inference.

C. *Experimentation with BOW features and Word Embedding for the BERT model*

In our experimentation, we employed a combined approach of Bag-of-Words (BOW) and BERT embeddings [24] to enhance the classification accuracy of RFP documents. We also employed ALBERT model. The overall methodology of Text Classification using BERT and Variations used in the experiment is listed in Algorithm 1.

Algorithm 1 Experimentation with BOW features and Word Embedding for the BERT model [24] and ALBERT

Input: Dataset $D = \{(d_i, y_i)\}$ where d_i is the document and y_i is the label
Output: Predicted labels $\hat{y}_i$ for each document i
Step 1: Prepare the Dataset
Load and preprocess the dataset D
Step 2: Base BERT Model
for each document d_i in D **do**
    Compute BERT embeddings: $\boldsymbol{E}^i_{\text{BERT}} = \text{BERT}(d_i)$
    Calculate the mean vector of word embeddings: $\mathbf{v}^i_{\text{BERT}} = \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{E}^i_{\text{BERT}} = [j]$
    Compute BOW features for d_i: $\mathbf{v}^i_{\text{BOW}} = \text{BOW}(d_i)$
    Concatenate mean vector and BOW features: $\mathbf{v}^i = \left[\mathbf{v}^i_{\text{BERT}} || \mathbf{v}^i_{\text{BOW}}\right]$
end for
Step 3: ALBERT for Comparison
for each document d_i in D **do**
    Compute ALBERT embeddings: $\boldsymbol{E}^i_{\text{ALBERT}} = \text{ALBERT}(d_i)$
    Calculate the mean vector of ALBERT embeddings: $\mathbf{v}^i_{\text{ALBERT}} = \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{E}^i_{\text{ALBERT}} = [j]$
end for
Step 4: Training Setup
Separate the dataset into training, validation, and test sets, ensuring consistency across all variations
Train classifiers using feature vectors $\mathbf{v}^i$ from BERT + BOW and ALBERT models
Evaluate all models on the test set utilizing metrics that include accuracy, precision, recall, and the F1-score

D. *E-gov Model for document classification using frequently used words*

Experts have seen that, even when selecting an RFP based only on its specifications, a lot of weight is placed on the title information, which is a quick criterion to evaluate and consider when considering the organization name in the full description. Thus, the attempt to accomplish title categorization is the algorithm that follows.

Algorithm 2 E-gov RFP document classification algorithm

Inputs:
$D = \{d_1, d_2, ..., d_n\}$: Dataset with *Title*, *Scope*, and *Organization*
$Labels = \{y_1, y_2, ..., y_n\}$: Labels for each document, where $y_i \in \{0, 1\}$
Top 20 Words $= \{w_1, w_2, ..., w_{20}\}$: Most frequent words extracted from *Title*
Outputs: Trained classification model combining BERT embeddings, one-hot encoding of top 20 words, cosine similarity,
and one-hot encoded features for *Organization* and *Scope*
1. Data Preprocessing:
Clean and preprocess text in *Title*, *Scope*, and *Organization*
Let T_i, S_i, and O_i be the preprocessed texts:
$T_i = \text{Preprocess}(Title_i)$
$S_i = \text{Preprocess}(Scope_i)$
$O_i = \text{Preprocess}(Organization_i)$
2. Extract Top 20 Words:
Compute word frequencies from *Title* and select the top 20 words:
$W_{\text{top20}} = \{w_1, w_2, ..., w_{20}\}$
Convert each title into a one-hot vector:

$$v_{ij} = \begin{cases} 1, & \text{if } w_j \in T_i \\ 0, & \text{otherwise} \end{cases}$$

The one-hot vector for a document d_i:
$v_i = [v_{i1}, v_{i2}, ..., v_{i20}]$
3. Compute Cosine Similarity:
For each test document, calculate the cosine similarity between the title and the top 20 words:

$$\text{CosineSimilarity}\left(T_i^{\text{test}}, W_{\text{top20}}\right) = \frac{v_i^{\text{test}} \cdot v^{\text{train}}}{\|v_i^{\text{test}}\| \|v^{\text{train}}\|}$$

4. Extract BERT Embeddings:
Generate BERT embeddings for each title:
$e_i = \text{BERT}_\theta(T_i)$
where $e_i \in R^{768}$
5. Feature Concatenation:
Combine BERT embeddings, one-hot vectors for top 20 words, and one-hot encoded *Organization* and *Scope*:
$x_i = [e_i, v_i, o_i, s_i]$
6. Classifier Training:
Train a neural network model f_θ with the feature vector x_i:
$\hat{y}_i = f\vartheta(x_i)$
Loss function: binary cross-entropy **(citation needed)**:

$$L = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

7. Model Evaluation:
Evaluate the model by calculating the accuracy along with confusion matrix, precision, recall, and F1-score:

$$\hat{y}_i^{test} - f\vartheta\left(x_i^{\text{test}}\right)$$

8. Refinement with Cosine Similarity:
Incorporate cosine similarity scores as an additional feature:

$$\text{CosineSimilarity}\left(T_i^{\text{test}}, W_{\text{top20}}\right)$$

4 Evaluation Results

The results of our experiments were analyzed to determine the best-performing model for classification. We compared the models based on their evaluation metrics and discussed the trade-offs between different approaches.

Figure 5 specifies the results of accuracy for all the different feature extraction models which have been involved in the experimentation in the form of bar graphs. Figure 6 visually present the comparison of accuracy across various models such as BOW Features, BERT Embeddings, ALBERT, and e-gov BERT Model.

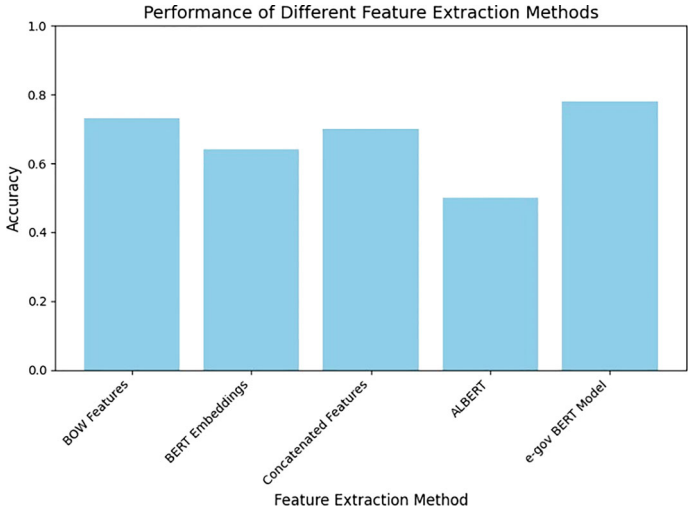


Fig. 5 Chart for the comparison of different model results on the e-gov title dataset

Fig. 6 Accuracy results of different models on the e-gov title dataset

Model	Accuracy
BOW Features	0.73
BERT Embeddings	0.64
Concatenated Features	0.70
ALBERT	0.50
e-gov BERT Model	0.78

A. *Error Analysis*

In the process of classifying RFP documents using both traditional and transformer-based models, several challenges were identified that impacted the accuracy of the models. These challenges primarily stem from the nature of the document titles and their inherent ambiguity. Below, we provide a detailed error analysis, highlighting common error types, their descriptions, and examples that illustrate the issues encountered during classification.

Ambiguous Titles—Titles that use generic terms, which do not clearly specify the scope or nature of the project, making it challenging for the models to accurately classify them as Digital governance or non-Digital governance. some such key-words are "System Integration Project", "Software Development", etc.

Overlapping Terminology—Instances where non-Digital governance RFPs use terminology similar to that found in Digital governance projects, leading to potential misclassification by the model. Examples "Cloud Services", "Data Management Solutions".

Short Titles—Titles that are too brief or lack specific descriptive terms, providing insufficient context for the model to make an accurate classification decision. "IT Services", "Consultancy"

5 Unique Contributions

- Top 20 Words The use of top 20 words from the title as one-hot-encoded features adds explainable and transparency to the model.
- Cosine Similarity Integrating cosine similarity between the test title and the most frequent top 20 words improves the model's ability to gauge textual relevance.
- BERT and Structured Features The combination of dense BERT embeddings with structured data (organization, scope) and engineered features (top words, cosine similarity) offers a rich feature set for classification.

6 Limitations and Challenges

Despite of the contributions, certain challenges and limitations were identified:

(1) Contributions are limited to Digital governance document classification
(2) Ambiguous Titles: Titles with generic terms or insufficient descriptive information posed significant challenges. For example, "System Integration Project" lacked the specificity needed for accurate classification.
(3) Overlapping Terminology: Some non-Digital governance RFPs used terminology similar to that found in Digital governance projects, leading to misclassifications. Terms like "cloud services" and "data management" are common across multiple sectors.

(4) Short Titles: Titles that were too brief, such as "IT Services," did not provide enough context for the models to make accurate predictions. Addressing these issues may involve incorporating additional contextual information from the body or scope of the RFP documents, rather than relying solely on the titles. Advanced NLP techniques, such as contextual embeddings (including the likes of BERT), could further enhance the models' understanding of ambiguous or short titles.

7 Practical Implications

The proposed classification model is poised to act as a handy tool in the hands of practitioners who deal in the tendering process of the software projects in the e-governance domain. The tool is expected to help in the preliminary analysis of the RFP document at the initial stage of procurement process. Early stage screening will also enable better efficiency in terms of the time and effort incurred in the identification process. This framework support the employees/bidders who are involved by saving their time and enable them to take accurate decisions thus improving decision making capabilities. This paper also explores how developing a knowledge repository as part of the text classification process for tenders could benefit both client organizations and bidders in the digital governance domain. Some of the key outcomes of the adoption of the above model are listed below.

(1) Improved efficiency in RFP analysis: A well-structured knowledge repository could significantly reduce the time taken by the bidders to conduct requirement analysis by quickly identifying and categorizing key elements of the RFPs.
(2) Standardization of tender documents: As part of the adoption of the model, the client organizations could use insights from the repository to standardize their RFP language and structure. This makes it easier for the bidders to understand and respond to the requirements.
(3) Historical trend analysis: The repository could enable multidimensional analysis of the past tender documents resulting in identifying patterns in requirements and evaluation criteria over time.
(4) Enhanced decision support: Both clients and bidders could use the repository to inform decision-making processes like project feasibility estimation or identifying competitive advantages.
(5) Automated initial screening: The clients may use the classification system to perform an initial automated screening of bid responses reducing the manual workload in the evaluation process.
(6) Continuous learning and improvement: As more documents are processed, the knowledge repository and classification system could be continuously refined resulting in the improved accuracy and effectiveness over time.
(7) Compliance and transparency: The repository can also help to ensure compliance with procurement regulations and increase transparency in the tender process by providing a clear record of requirements and their interpretations.

(8) Cost reduction: The streamlining of the analysis process could potentially reduce costs associated with tender preparation and evaluation for both clients and bidders.
(9) Knowledge transfer and retention: The repository could serve as a valuable tool for knowledge transfer within organizations by helping the new human resources to understand the past tenders and organizational priorities.

8 Future Work

While the current classification approach yields promising results, there are some areas that require more improvement and research. While reading and analyzing the RFP documents in the tendering process, apart from classifying the documents based on the domain, further analyses is required that involves further classification of documents to sub-domains. This classification mirrors the decision model adopted by the human evaluator while reviewing the RFP document. As a test case, the system determines whether an RFP is related to a domain like digital governance. After this broad categorization initially, it then identifies specific subcategories (e.g., Aadhar-based projects). This enables the model to further explore this hierarchical classification focusing on the related sub-domains and further sub-categories like pre-requisite information, timeline information, requirments etc. Another scope of work could be the refinement of the knowledge repository created as an outcome of the model operations.

9 Conclusion

Our work indicates the usefulness of text classification approaches in the RFP evaluation stage of the tender process for the software projects related to the domain of e-governance. We tried variations of BERT-based algorithms to identify if the document belongs to an digital governance domain. The procurement process through tendering is a multi step process which requires significant human intervention besides fetching RFP documents from the portals. Our work attempts to simplify the early stages of the work and thereby benefiting bidders and officials involved int he process. Another significant contribution of the work is creation of a static knowledge repository related to the given domain that can be utilised at a later stage. This is a stepping stone to a varied number of activities which we can perform in the process and similar such activities and thereby contribute to the larger goal of digital transformation in government processes.

References

1. Sunmola, F.T., Shehu, Y.U.: A case study on performance features of electronic tendering systems. Proc. Manuf. **51**, 1586–1591 (2020). ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2020.10.221
2. Goswami, Y., Agrawal, A., Bhatia, A.: Digital governance: a tendering framework using blockchain with active participation of citizens. In: 2020 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), New Delhi, India, pp. 1–4 (2020). https://doi.org/10.1109/ANTS50601.2020.9342816
3. Rajbhoj, A., Nistala, P., Batra, P., Kulkarni, V.: AI-enabled Project Initiation: an approach based on RFP response document. In: Proceedings of the 15th Innovations in Software Engineering Conference (ISEC '22). Association for Computing Machinery, New York, NY, USA, pp. 1–5 (2022) (Article 22). https://doi.org/10.1145/3511430.3511450
4. Motahari-Nezhad, H.R., Cappi, J.M., Nakamurra, T., Qiao, M.: RFPCog: linguistic-based identification and mapping of service requirements in request for proposals (RFPs) to IT service solutions. In: Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS) (HICSS '16). IEEE Computer Society, USA, pp. 1691–1700 (2016). https://doi.org/10.1109/HICSS.2016.213
5. Nistala, P., Rajbhoj, A., Kulkarni, V., Noronha, S., Joshi, A.: An industrial experience report on model-based, AI-enabled proposal development for an RFP/RFI. Sci. Comput. Program. **233**, 103058 (2024). ISSN 0167-6423. https://doi.org/10.1016/j.scico.2023.103058
6. Tiwari, M., Aital, P., Joshi, P.: A comprehensive survey and review of machine learning techniques in document processing: industry applications and future directions. J. Inf. Optim. Sci. **45**(4), 1177–1188 (2024). https://doi.org/10.47974/JIOS-1701
7. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: Proceedings of ECML-98, 10th European Conference on Machine Learning, Springer, pp. 137–142 (1998). https://doi.org/10.1007/BFb0026683
8. McCallum, A., Nigam, K.: A comparison of event models for Naive Bayes text classification. In: AAAI-98 Workshop on Learning for Text Categorization, PDF link from CMU, pp. 41–48 (1998)
9. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
10. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of ICML, pp. 412–420 (1997). https://www.cs.cmu.edu/~yiming/publications/icml97.pdf
11. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval. Cambridge University Press. Relevant section: "The curse of dimensionality in text classification" (2008)
12. Xu, R., Wang, J., Su, Y.: A comparative study of Random Forest, K-Nearest Neighbor, and Support Vector Machine for the classification of highdimensional data. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), (2012)
13. Rasjid, Z.E., Setiawan, R.: Performance comparison and optimization of text document classification using k-NN and Naïve Bayes classification techniques. Proc. Comput. Sci. **116**, 107–112 (2017. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2017.10.017
14. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J.: Deep learning-based text classification: a comprehensive review. ACM Comput. Surv. **54**(3) (2021) (Article 62) (40 pp). https://doi.org/10.1145/3439726
15. Li, Z., et al.: A unified understanding of deep NLP models for text classification. In: IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 12, pp. 4980–4994 (2022). https://doi.org/10.1109/TVCG.2022.3184186 (keywords: computational modeling; analytical models; visualization; internet; data models; Computer architecture; Adaptation models; explainable AI; visual debugging; visual analytics; deep NLP model; information-based interpretation)
16. Basarkar, A.: Document classification using machine learning. Master's Projects, p. 531 (2017). https://doi.org/10.31979/etd.6jmu-9xdt https://scholarworks.sjsu.edu/etdprojects/531

17. Adhikari, A., Ram, A., Tang, R., Lin, J.: DocBERT: BERT for document classification. (2019). Eprint = 1904.08398, https://arxiv.org/abs/1904.08398
18. Pappagari, R., Zèlasko, P., Villalba, J., Carmiel, Y., Dehak, N.: Hierarchical transformers for long document classification. (2019). Eprint = 1910.10781, https://arxiv.org/abs/1910.10781
19. Garrido-Merchan, Eduardo C., Gozalo-Brizuela, R., Gonzalez-Carvajal, S.: Comparing BERT against traditional machine learning models in text classification. https://doi.org/10.47852/bonviewJCCE3202838
20. Aletras, N., Androutsopoulos, I., Barrett, L., Goanta, C., Preotiuc-Pietro.: Effectively leveraging BERT for legal document classification (2021). https://doi.org/10.18653/v1/2021.nllp-1.22. https://aclanthology.org/2021.nllp-1.22
21. Prabhu, S., Mohamed, M., Misra, H.: Multi-class text classification using BERT-based active learning (2021). https://arxiv.org/abs/2104.14289
22. Zhu, Y., Zhou, X., Qiang, J., Li, Y., Yuan, Y., Wu, X.: Prompt-Learning for Short Text Classification (2022). https://arxiv.org/abs/2202.11345
23. Kong, J., Wang, J., Zhang, X.: Hierarchical BERT with an adaptive fine-tuning strategy for document classification. Knowl. Based Syst. **238**, 107872 (2022). ISSN 0950-7051. https://doi.org/10.1016/j.knosys.2021.107872
24. Tanaka, H., Shinnou, H., Cao, R., Bai, J., Ma, W.: Document classification by word embeddings of BERT. In: Nguyen, L.M., Phan, X.H., Hasida, K., Tojo, S. (eds.) Computational Linguistics. PACLING 2019. Communications in Computer and Information Science, vol. 1215. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-6168-9-13
25. Batra, A., Rajkumar, S., Nielsen, D.L., Clarke, S.S., Kalyanam, K., Tejasen, K., Ohsfeldt, M., Copp, M.: Document Classification Techniques for Aviation Letters of Agreement. AIAA 2024-4358
26. Hey, T., Keim, J., Koziolek, A., Tichy, W.F.: NoRBERT: transfer learning for requirements classification. In: 2020 IEEE 28th International Requirements Engineering Conference (RE), Zurich, Switzerland, pp. 169–179 (2020). https://doi.org/10.1109/RE48521.2020.00028
27. Yu, L., Liu, B., Lin, Q., Zhao, X., Che, C.: Semantic similarity matching for patent documents using ensemble BERT-related model and novel text processing method (2024). https://arxiv.org/abs/2401.06782
28. Qasim, R., Bangyal, W.H., Alqarni, M.A., Ali, A.A.: A fine-tuned BERT-based transfer learning approach for text classification. J Healthc Eng. **2022**, 3498123 (2022). https://doi.org/10.1155/2022/3498123
29. Gao, S., Alawad, M., Young, M.T., Gounley, J., Schaefferkoetter, N., Yoon, H.J., Wu, X.C., Durbin, E.B., Doherty, J., Stroup, A., Coyle, L., Tourassi, G.: Limitations of transformers on clinical text classification. IEEE J. Biomed. Health. Inf. **25**(9), 3596–3607 (2021). https://doi.org/10.1109/JBHI.2021.3062322
30. Xu, H.: Applying the Bert algorithm for software requirements classification (2024). https://doi.org/10.13140/RG.2.2.33317.68320
31. Kim, D., Seo, D., Cho, S., Kang, P.: Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. Inf. Sci. **477**, 15–29 (2019). ISSN 0020-0255. https://doi.org/10.1016/j.ins.2018.10.006
32. Muaad, A.Y., Hemantha Kumar, G., Hanumanthappa, J., Bibal Benifa, J.V., Naveen Mourya, M., Chola, C., Pramodha, M., Bhairava, R.: An effective approach for Arabic document classification using machine learning. Global Trans. Proc. **3**(1), 267–271 (2022). ISSN 2666-285X. https://doi.org/10.1016/j.gltp.2022.03.003
33. Kadhim, A.I.: Survey on supervised machine learning techniques for automatic text classification. Artif. Intell. Rev. **52**(1), 273–292 (2019). https://doi.org/10.1007/s10462-018-09677-1
34. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R.: Albert: a lite bert for self-supervised learning of language representations. International Conference on Learning Representations (ICLR) (2020). Available at: https://openreview.net/forum?id=H1eA7AEtvS
35. Dogra, V., Verma, S., Kavita, Chatterjee, P., Shafi, K., Choi, J., Fazal Ijaz, M.: A complete process of text classification system using state-of-the-art NLP models. https://doi.org/10.1155/2022/1883698

Application of Condition-Based Monitoring for Improvement of Diagnostic Coverage of Safety Mechanisms



Pramit Nag

Abstract Conformance to ISO26262 functional safety standard involves inclusion of safety mechanisms which are necessary to reach defined targets for safety critical applications. 'Diagnostic coverage' (DC) is a measure of the effectiveness of safety mechanism against faults of the E/E components. The ISO 26262 standard provides standard guidelines for choosing possible safety mechanisms for reaching the target safety metrics. But while calculating the DC for these mechanisms, it only considers the base failure rate of the elements. In practical scenario, each safety element undergoes different types of stresses throughout the lifetime of the product. Also, to achieve the required target of safety metrics, multiple safety mechanisms are added in the design which sometimes unnecessarily increases the complexity and cost of the product development. To monitor the effect of stresses and take preventive actions accordingly different condition-based monitoring techniques are widely used in the automotive industry. In this paper, our aim is to propose an additional approach to improve the practical diagnostic coverage of a system by using condition-based monitoring as a safety mechanism to detect the health of the elements which helps the system to transit to a safe state. Thus, depending on the health of the elements, system can take alternate safe state or degradation strategy if required for avoiding the hazard.

Keywords Functional safety · Diagnostic coverage · Condition based maintenance · ISO 26262 · Health monitoring

1 Introduction

With the increase of complexity of electrical and electronic systems used in modern vehicles, the need for enhancing the product reliability, reduction of potential hazards and improving the overall safety performance of the vehicle has become very much

P. Nag (✉)
Eaton India Innovation Center, Pune, India
e-mail: PramitNag@eaton.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_25

crucial for the automakers, suppliers and other stakeholders. ISO 26262 [1] standard provides a structured approach to manage the safety risks in the electrical and electronic systems, which applies to the entire life cycle of a safety critical system, including concept, design, development, production, operation, maintenance and decommissioning. It follows a risk-based classification, known as the automotive safety integrity level (ASIL) to classify the inherent risk in the system, where ASIL range from ASIL A (least stringent) to ASIL D (most stringent). This approach is accepted and used as a benchmark for achieving functional safety in automotive industry globally.

To reduce the safety risks in a system, some additional measures are introduced depending on the type of fault that are needed to be controlled. While measuring the effectiveness of any safety measure, diagnostic coverage plays a key role as it refers to the ability of that particular safety measure to control (detect/prevent/mitigate) safety critical faults. Ensuring robust diagnostic coverage is paramount in safeguarding against unreasonable risks in the system that may occur due to malfunctioning behavior of the system.

Condition-based monitoring (CBM) is a proactive maintenance strategy that utilizes various sensors, data collection techniques and analytics to assess the condition of the safety critical elements in the system and predict the failures before they occur in modern vehicles [2, 3]. Application of CBM in enhancement of automotive safety by preventing unexpected failures has been explored in various research which for instance has mentioned development of advanced mathematical models and artificial intelligent techniques to analyze condition-monitoring data, enabling timely maintenance actions [4]. While significant progress has been made in the field of CBM, there is a lack of standardized guidelines for implementing CBM in automotive systems. Existing research often overlooks the integration of CBM with functional safety standards, which is crucial for ensuring the reliability of safety–critical systems. This gap hinders the widespread adoption and consistent application of CBM technologies across the industry. ISO 9839 international standard addresses some of these gaps and outlines a systematic approach for the integration of CBM with functional safety standards representing a significant advancement in the field [5].

The purpose of this paper is to discuss the approach of using CBM to detect the possible anomalies and deviations from normal behavior for safety critical elements and thus improving the diagnostic capabilities of automotive systems.

2 ISO 26262 Requirements for Diagnostic Coverage

Adherence to ISO 26262 requirements for diagnostic coverage is essential for enhancing the resilience of safety mechanisms against potential failures. It allows to improve the overall safety performance of the system. The standard provides certain guidelines on choosing the most effective safety mechanism depending on the faults that may occur for safety critical elements in a complex automotive system.

Table 1 Random hardware failure rate targets as per ISO 26262:2018-5 [1]

Metric name	Target for ASIL B	Target for ASIL C	Target for ASIL D
Single point fault metric (SPFM) (%)	≥ 90	≥ 97	≥ 99
Latent fault metric (LFM) (%)	≥ 60	≥ 80	≥ 90
Probabilistic metric for random hardware failures (PMHF)	< 100FIT	< 100FIT	< 10FIT

To achieve the overall functional safety compliance for a system it is mandatory to meet the random hardware failure metrics as prescribed by ISO 26262:2018 Part 5 requirements for various ASIL.

Table 1 shows the target values for different random hardware failure metrics as mentioned in the standard for different ASIL.

The mathematical expressions that are used to calculate the random hardware failure metrics clearly emphasize on the importance of diagnostic coverage of safety mechanisms. ISO 26262:2018 Part 5 annexure D provides few guidelines for choosing safety mechanisms by categorizing them into high (99% DC), medium (90% DC) and low (60% DC) according to their capability to detect safety critical random hardware faults.

While these diagnostic coverage values could be used as a starting point for analyzing the effectiveness of the safety mechanisms, these need to be checked by performing proper functional safety tests (e.g., fault injection tests) for getting more accurate estimations.

Although, the random hardware failure metrics focuses only on the random hardware failures, we cannot deny the impact of systematic failures in the context of the overall safety of the system. Most of the systematic failures are often addressed by adhering to standard process guidelines, performing various safety analysis and performing verification tests to check the robustness of the elements. But to ensure the complete safety of the system over the lifetime needs special focus on the life limiting failure modes as well. Although we cannot consider these failure modes as truly random in nature, these can often arise from extreme conditions or increased stress levels. It is equally important to have effective safety measures implemented in the system to protect the system from life limiting failure modes.

3 Fundamentals of Condition-Based Monitoring

Traditional maintenance techniques are used in modern vehicles from a long time which are mainly based on reactive maintenance—where the systems are repaired/replaced in a systematic way once the failure has occurred. Common traditional maintenance techniques include visual inspections, oil changes, filter replacements, tire rotations, brake pad replacements, etc.

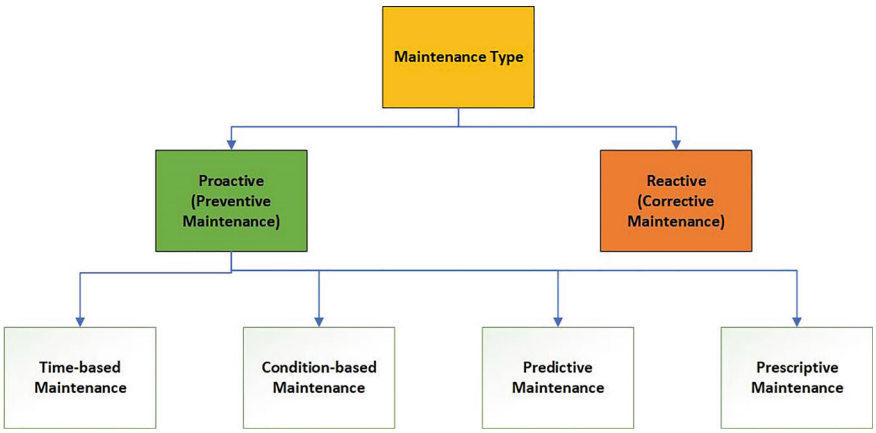


Fig. 1 Different types of maintenance strategies

On the other hand, there are some maintenance techniques which focus on identifying the risks proactively and indicate the user for fixing the issue. As shown in Fig. 1, proactive (preventive) maintenance techniques can be categorized into 4 types:

- Time based maintenance
- Condition-based maintenance
- Predictive maintenance
- Prescriptive maintenance

Condition-based monitoring (CBM) is an important strategy for predictive and preventive maintenance. The term condition-based monitoring refers to monitoring the health or condition of crucial elements to judge whether maintenance is required for the element or not. There are four (4) phases in this kind of maintenance strategy.

- Monitoring and collecting data from different sensors
- Processing the gathered data
- Diagnosing the system
- Making the decision for maintenance strategy

4 Integration of Condition-Based Monitoring with ISO 26262

For calculating the quantitative safety analysis metrics according to ISO 26262:2018 Part 5, base failure rate is used as a primary input. This calculation mainly focuses on random hardware failures only. But most of the techniques which are available for base failure rate estimation do not distinguish between random and systematic

failure modes. As a result, while calculating the random HW failure metrics, we use a more conservative value as base failure rate.

The use of a 'bathtub curve' simplification is very common for most of the base failure rate estimation models. This simplification assumes that infant mortality or early life failure modes are screened by the manufacturer and that end-of-life or wear out failure modes are supposed to occur at very negligible rates during useful mission lifetime. But this simplification is only effective if the failure rate distribution follows constant failure rate, which is more compatible with the computation of random hardware failure metrics as per ISO 26262:2018 Part 5. For the non-constant failure rate distributions, it is recommended by the standard to limit the operating lifespan of the product such that a constant failure rate approximation is more appropriate (Fig. 2).

The standard recommends three different techniques for base failure rate estimations. These are:

(a) Derivation of failure rates based on experimental testing,
(b) Derivation of failure rates from observation of field incidents,
(c) Derivation of failure rates by application of industry reliability data books combined with expert judgment.

Once the base failure rate is properly estimated for the safety critical hardware elements, a quantitative safety analysis is performed to measure the effectiveness of existing safety mechanisms to avoid any unreasonable risks in the system. Earlier we have discussed the various random hardware failure metrics and their dependency on the diagnostic coverage.

To improve the diagnostic coverage of an element, sometimes a combination of multiple safety mechanisms can also be used. In such cases, the final diagnostic coverage is estimated based on the diagnostic coverages of each safety mechanism [6]. Use of multiple safety mechanism is more common now-a-days as the systems are getting more and more complex, but this is associated with higher cost and longer design time. To add to this, the true effectiveness of these safety mechanism relies

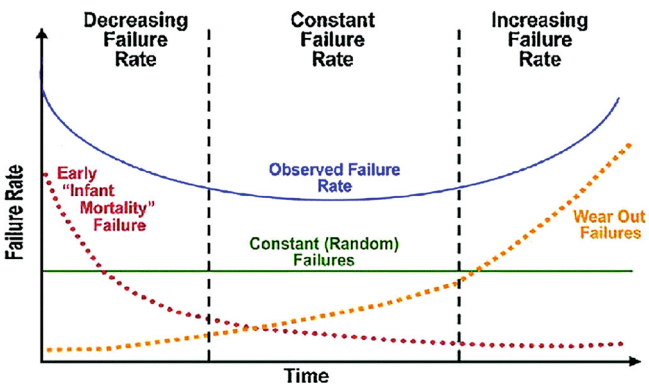


Fig. 2 Bathtub curve—evolution of failure rate over time [1]

on the reliability of the elements used to implement the mechanisms as well as to implement the safe state transition reaction. If the hardware or software components which are added in the design to implement a certain safety mechanism degrades under stress over lifetime, the detectability of faults gradually decreases which results in degradation of the diagnostic coverage of the safety mechanisms over lifetime. To avoid this, various life tests are recommended in the standard for ensuring durability, robustness, and operation under stress.

To achieve the overall functional safety compliance, it is crucial that the system transits to a safe state when a fault is detected by the safety mechanisms to avoid any harm. But it is also equally important that the system is able to transit back to normal state once the fault is mitigated. In both the cases where the system is needed to continue function in the event of fault (fail operational) the designers rely on the redundancy of the elements most of the times. It adds extra cost and complexity in the design process.

Monitoring the condition of the safety critical elements and taking reactions proactively can mitigate this problem altogether as the system does not need to wait till the fault occurs to transit to a safer sate [7]. Implementing different condition-based monitoring measures help the system to prevent few critical faults from occurring, and thus the need of adding extra safety measures to detect those faults can be avoided. Based on the data acquired by the CBM related sensors, the system can predict the faults in the safety critical paths and can alert the driver about the need for maintenance of the parts to avoid any safety hazards.

5 Methodological Approach

The overall approach for using condition-based monitoring to improve diagnostic coverage can be divided into four (4) phases:

i System identification: Identification of critical components whose diagnostic coverage requires enhancement based on its ASIL classification and safety analysis results.
ii Sensor deployment: Selection of appropriate sensor or data acquisition system to monitor key parameters that indicates system health.
iii Data collection & analysis: Collection of raw data from the sensors and analyze them to detect any anomaly.
iv Fault detection & reporting: Prediction of possible faulty behavior of the system and react accordingly to avoid unreasonable risks.

To discuss this approach, we can use a simple example of an E/E system (called as ‘system A’). Let us consider that ‘system A’ is used for connecting high voltage (HV) battery to loads. The system is shown in the Fig. 3, where the control unit is used to control the relay for connecting/disconnecting high voltage path to loads. A current sensor is added to implement a safety functionality—depending on the current in the

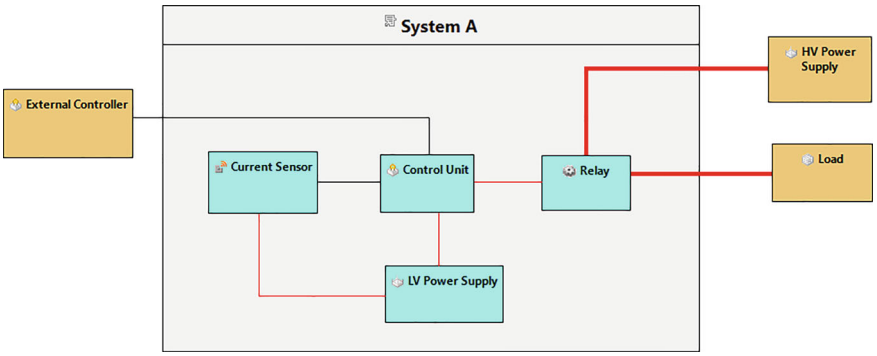


Fig. 3 Architecture diagram of 'system A'

HV path, control unit can take decision for relay control. Low voltage (LV) power supply is used to provide the necessary power input for operation of the elements.

Let us consider the system has mainly two safety goals: *SG01-To disconnect the HV loads from battery when required (ASIL D)* and *SG02-To avoid unintended disconnection of HV loads (ASIL D)*. Figure 4 shows all the possible failure modes of the elements of 'system A', where the failure modes are chosen as per ISO 26262:2018 Part 5 guidelines. For simplicity of the hardware failure rate estimation, we have chosen 100 FIT to be the base failure rate for all the components and distributed the base failure rate to each failure mode equally. The Fig. 4 is a screenshot taken from the hardware element part library created using Ansys Medini Analyze tool for functional safety analysis.

Now let us concentrate on the quantitative safety analysis for *SG02: To avoid unintended disconnection of HV loads (ASIL D)*. To achieve the functional safety of the system, the following safety measures are added to the system as shown in Fig. 5.

Now let us we apply our 4-stage strategy of using condition-based monitoring for this system:

Component	Prediction Mode	Percentage of Parent	Mission Profile	Failure Rate Calculation				Failure Modes		
				Raw Failure Rate (FIT)	Scaling Expression	Apply scaling to children	Failure Rate (FIT)	Potential Failure	Failure Distribution (%)	Failure Fraction (FIT)
System A	User defined			0.0		☑	0.0			
Current Sensor	User defined			100.0		☑	100.0	Sensor stuck out of range	25.0	25.0
								Sensor stuck in range	25.0	25.0
								Sensor drift	25.0	25.0
								Sensor offset	25.0	25.0
Control Unit	User defined			100.0		☑	100.0	Incorrect output (Output stuck high)	50.0	50.0
								Incorrect output (Output stuck low)	50.0	50.0
Relay	User defined			100.0		☑	100.0	Relay stuck closed	50.0	50.0
								Relay stuck open	50.0	50.0
LV Power Supply	User defined			100.0		☑	100.0	LV supply undervoltage	25.0	25.0
								LV supply overvoltage	25.0	25.0
								LV supply drift	25.0	25.0

Fig. 4 Failure modes and their failure rates of HW elements of 'system A'

Safety Mechanism/Measure	Section Key or Identifier	Typical Diagnostic Coverage considered achievable
Failure detection by on-line monitoring	ISO26262-5: D.2.1.1	HIGH
HW redundancy (e.g. Dual Core Lockstep, asymmetric redundancy, coded processing)	ISO26262-5: D.2.3.6	HIGH
Voltage or current control (output)	ISO26262-5: D.2.6.2	HIGH
Sensor Correlation	ISO26262-5: D.2.8.2	HIGH
Sensor Valid Range	ISO26262-5: D.2.8.1	LOW
Failure detection by on-line monitoring	ISO26262-5: D.2.1.1	LOW
Test pattern	ISO26262-5: D.2.4.1	HIGH

Fig. 5 Safety mechanism list for 'System A'

i. System identification

To add the safety mechanisms as mentioned in Fig. 5 in 'system A', dual sensing channels can be used to ensure that sensor correlation can be performed in addition with sensor valid range checks in the control unit. Similarly for control unit HW redundancy (dual core lockstep approach) can be used to achieve high DC. LV power supply output is monitored for ensuring the LV supply is within appropriate range.

Similarly for the relay, voltage sensing points can be installed before and after the relay. Control unit can analyze the voltage sensor data and detect the stuck faults of the relay. Same voltage sensing points can be used to perform periodic tests to check the relays are working properly or not.

With these safety mechanisms if we perform the failure mode, effect and diagnostic analysis (FMEDA) for the system in context of SG02 we will get the results as shown in Fig. 6.

As we can notice that for both single point fault and latent fault metrics the estimated value is not meeting the required target as per ASIL D classification. If we further deep dive into the causes why the metrics are not meeting the target, we would notice that although the relays are equipped with voltage monitoring safety measures it is not sufficient to prevent violation of safety goal. For the case when the relay is stuck open, the control unit can detect the fault by the voltage sensor data but there is nothing the control unit can do to prevent the violation of safety goal (i.e., once the relay is stuck open at run time, unintended disconnection of loads is inevitable). As a result, the effective DC for the safety mechanisms used for the relay is 0% as shown in the Fig. 7.

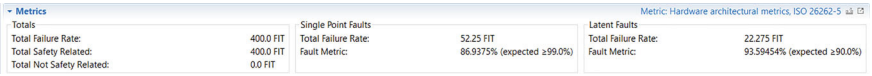
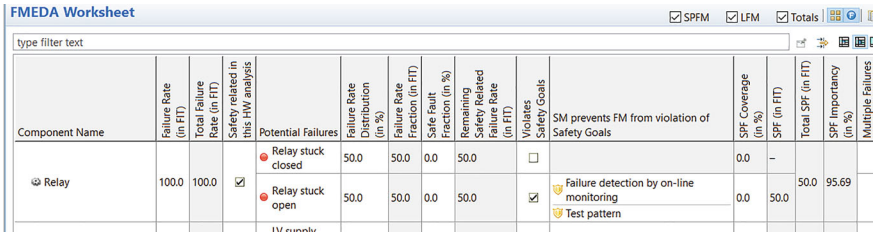


Fig. 6 Screenshot of FMEDA output for SG02


FMEDA Worksheet

SPFM LFM Totals

type filter text

Component Name	Failure Rate (in FIT)	Total Failure Rate (in FIT)	Safety related in this HW analysis	Potential Failures	Failure Rate Distribution (in %)	Failure Rate Fraction (in FIT)	Safe Fault Fraction (in %)	Remaining Safety Related Failure Rate (in FIT)	Violates Safety Goals	SM prevents FM from violation of Safety Goals	SPF Coverage (in %)	SPF (in FIT)	Total SPF (in FIT)	SPF Importancy (in %)	Multiple Failures
				Relay stuck closed	50.0	50.0	0.0	50.0	☐		0.0	–			
Relay	100.0	100.0	☑	Relay stuck open	50.0	50.0	0.0	50.0	☑	Failure detection by on-line monitoring Test pattern	0.0	50.0	50.0	95.69	

Fig. 7 Screenshot of FMEDA sheet for SG02

To solve this problem, one of the most common approaches could be to use redundant relays in the output path, so that if one relay is stuck open, the other can be used to maintain the HV connection. This approach is costly, complex and requires elimination of all the dependent failures between the two relays as well (as there could be some dependent failures which effect both the relays equally so that both fail simultaneously). This makes the relay a suitable candidate for using condition-based monitoring as a safety mechanism to achieve the required diagnostic coverage.

ii. Sensor deployment:

To understand appropriate parameters that can influence the relay performance, the design team shall identify the major factors which causes the relay to stuck open at run time. This activity can involve verification testing, error guessing based on expert judgement, field failure data collection, competitive benchmarking etc.

For relays, we have considered that temperature change can be a crucial factor in degradation of the relay performance over lifetime. There are other factors which can impact the relay performance but let us consider only temperature as the key parameter for simplicity at this point. To monitor the relay condition based on temperature monitoring the design team can deploy temperature sensor in appropriate position of the system. Depending on the monitoring strategy, the choice of elements could be different [8, 9].

iii. Data collection & analysis:

Once the temperature sensor has been deployed, the control unit shall be equipped with an appropriate algorithm which predicts the degradation of the relay w.r.t change in temperature.

In this case, we have considered that the relay coil characteristics has dependency on temperature. Using this relation, the controller shall predict the fault of relay.

iv. Fault detection & reporting:

Once the fault has been predicted by the control unit, it shall decide the reaction strategy. The overall strategy has been depicted in the Fig. 8.

By utilizing the condition-based monitoring the system is capable of detecting current and probable future faults in the relay which helps the control unit to detect

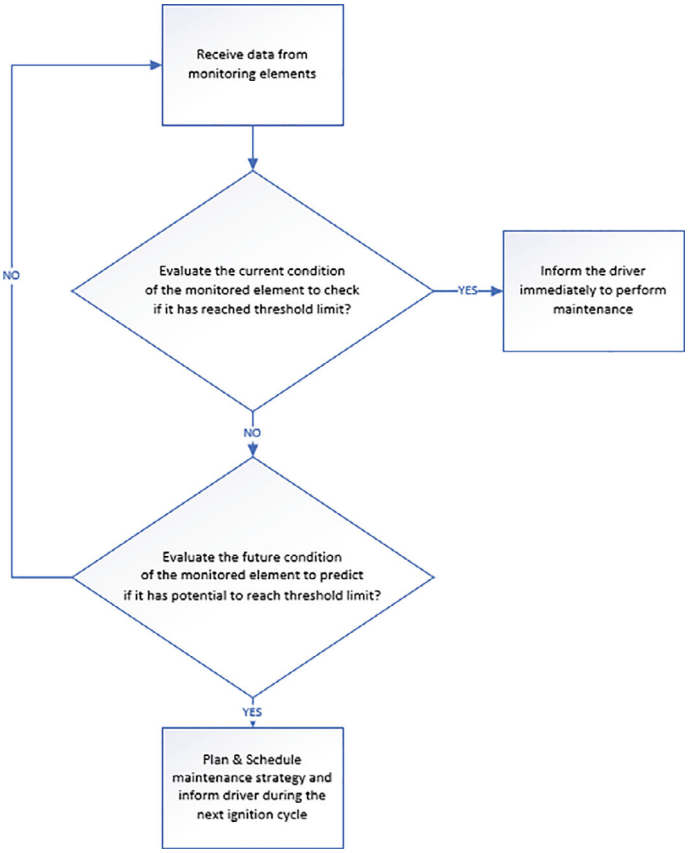


Fig. 8 Decision making flow for current and future faults

'stuck' faults of the relay at the start of every ignition cycle. So, the control unit can warn the driver and move the system to a safe state to avoid any hazardous event. This allows the effective diagnostic coverage for the safety mechanisms (for relay) to be considered as High (99%). Figure 9 shows the updated hardware metrics for the system considering the condition-based monitoring method for relay.



Fig. 9 Screenshot of FMEDA Output for 'system A' after using Condition-based monitoring

6 Discussion

Condition based monitoring is effective for degrading faults whose characteristics are not constant and degrade over time. These faults occur when stimulated after degradation exceeds a critical threshold. In the context of ISO 26262, to meet the safety rating of the technical safety requirements, it is necessary to meet the diagnostic coverage for random hardware faults. It is very important to choose the appropriate CBM technique based on its detection quality (type I and type II errors) to satisfy the target safety metrics.

In application of CBM for enhancing functional safety using the 4-phase method described in this paper, several techniques and tools can be employed for data collection and analysis. Deploying different sensors such as temperature, voltage, vibration, pressure etc. are some common ways of sensor-based monitoring. Other methods like using accelerometers to detect changes in vibration pattern in system or using infrared cameras to detect abnormal heat signatures are also common in the industry. Once data is collected, statistical methods such as regression analysis, statistical process control can be used to identify trends and anomalies in the collected data. Advanced machine learning algorithms, or even signal processing methods like Fast Fourier transform, wavelet analysis are a few state-of-the-art methods for interpreting patterns of potential degrading faults [10].

Despite its benefits, CBM faces challenges such as high implementation costs, the need for specialized expertise, and integration complexities. Research has highlighted the importance of balancing these costs with the operational benefits to maximize the effectiveness of CBM systems [11]. Putting additional safety measures for degrading faults might not be the best solution all the time in terms of design complexity, cost, design time. Thus, it is very much important to identify whether adding CBM techniques can help reduce risk associated with the primary failure mechanisms for the system under consideration or not before choosing it as a safety measure.

7 Conclusion and Future Scope

In conclusion we can say that the effective use of condition-based monitoring can increase the effective diagnostic coverage and help the systems to comply with the ISO 26262 functional safety standard. Proactive maintenance strategy like condition-based monitoring helps to prevent the occurrence of safety critical failures in runtime. But choice of an effective proactive maintenance strategy is equally important. Sometimes, it might be possible that implementing a condition-based maintenance strategy is more costly and complex than using some of the most common safety measures (like redundancy in output stages).

Thus, this strategy is more suitable if multiple elements in the system could be monitored using a single framework without adding extra monitoring elements. Future works could focus on identifying such common factors in modern vehicular

safety critical systems which can be monitored to identify faults in multiple elements together and this could help the safety engineers in achieving the required safety performance without increasing the complexity of the system.

References

1. ISO, ISO 26262:2018 Road vehicles—Functional safety—Part 1 to 10
2. Shao, G., Goldstein, D., Kim, K., Nwadiogbu, E., et al.: Creating a System Architecture for a Vehicle Condition-Based Maintenance System. SAE Technical Paper 2012-01-2097 (2012). https://doi.org/10.4271/2012-01-2097
3. SAE International Recommended Practice, Guidelines for the Development of Architectures for Integrated Vehicle Health Management Systems. SAE Standard ARP6290, Issued May 2023. https://doi.org/10.4271/ARP6290
4. Arena, F., Collotta, M., Luca, L., Ruggieri, M., Termine, F.G.: Predictive maintenance in the automotive sector: a literature review. Math. Comput. Appl. **27**(1), 2 (2022). https://doi.org/10.3390/mca27010002
5. ISO, ISO 9839:2023 Road vehicles—Application of predictive maintenance to hardware with ISO 26262-5
6. KVA Homepage, Calculating Diagnostic Coverage | kVA (kvausa.com)
7. Semiconductor Engineering Homepage, Predictive Health Monitoring in Functional Safety (semiengineering.com)
8. Huang, S., Chen, S., Qiu, Y., Ye, Y., Pan, W.: Online Condition Monitoring Methodology For Relay Protection Based on Self-Test Information. IEEE (2011). https://doi.org/10.1109/APAP.2011.6180525
9. Harshitha, S., Arpitha, B., Shwetha, H., Sinchana, N., Smitha, B., Nagraj, M.: Real-time health monitoring of relays and circuit breakers. In: Shetty, N.R., Patnaik, L.M., Prasad, N.H. (eds.) Emerging Research in Computing, Information, Communication and Applications. Lecture Notes in Electrical Engineering, vol. 928. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-5482-5_40
10. Ali, A., Abdelhadi, A.: Condition-based monitoring and maintenance: state of the art review. Appl. Sci. **12**(2), 688 (2022). https://doi.org/10.3390/app12020688
11. Li, Y., Peng, S., Li, Y., et al.: A review of condition-based maintenance: its prognostic and operational aspects. Front. Eng. Manag. **7**, 323–334 (2020). https://doi.org/10.1007/s42524-020-0121-5

Predictive Maintenance of Hydraulic Systems Using Multi-task Transfer Learning with Multi-layer Perceptron



Tarun Kumar, Mahak Kokate, Pavan Kumar Kankar, and Ankur Miglani

Abstract Condition monitoring and maintenance of hydraulic systems are essential to ensure their proper working in various applications. This paper uses multi-task transfer learning, which employs shared information between the datasets and creates a machine learning model that is accurate and generalized for more than one dataset. This paper uses multi-layer perceptron (MLP) as the main model architecture for multi-task transfer learning because MLP can find complex features and non-linear interactions in the dataset. After finding satisfactory correlations between the internal pump leakage and valve condition datasets, the multi-task transfer model was trained and tested for the predictive maintenance of pump leakage and valve condition. An accuracy of 94.33% for internal pump leakage and 93.65% for the switching behaviour of hydraulic valves was found. This paper demonstrates that knowledge transfer between two related datasets can play a vital role in significantly improving data-driven models' performance.

Keywords Transfer learning · Condition monitoring · Pump leakage · Valve condition · Multi-layer perceptron

1 Introduction

Hydraulic systems are the most used systems in industries such as the automobile, aerospace, and manufacturing industries. They are widely used because they can easily control large forces and do work precisely [1]. Hydraulic systems mainly use pressurized fluids to transfer energy and enable machine operations. Hydraulic

T. Kumar · M. Kokate · P. K. Kankar (✉)
System Dynamics Lab, Department of Mechanical Engineering, Indian Institute of Technology Indore, Indore, India
e-mail: pkankar@iiti.ac.in

A. Miglani
Microfluidics and Droplet Dynamics Lab, Department of Mechanical Engineering, Indian Institute of Technology Indore, Indore, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_26

systems have high power and are flexible, which makes them the best option for applications where heavy machinery is used and large forces are required. Although they have various advantages, hydraulic systems are prone to component degradation over time. Essential components of hydraulic systems, such as pumps, valves, seals, etc., can degrade due to wear, mechanical fatigue, and contamination. These failures can lead to total system inactivity, high-cost repairs, and even more damage if left unchecked. Traditionally, industries have many maintenance strategies, such as reactive and preventive maintenance [2], but these have many disadvantages, such as emergency repairs and unnecessary component replacements. These challenges are addressed using predictive maintenance, which is more effective and highly reliable [3]. Predictive maintenance depends on data-driven models that monitor the system's real-time performance and predict the failures before they happen [4, 5]. Maintenance is easily optimized using this technique. Many hydraulic components, including pumps, valves, accumulators, and coolers, are prone to degradation and failure. Prakash et al. [6] have worked on the switching behaviour of valves using machine learning techniques and continued their work on internal leakage detection of pumps using ensemble learning techniques, exhaustive feature selection, and C-LSTM Neural Network [7, 8]. These papers addressed two critical failures: hydraulic system pump leakages and the other is valve switching conditions, which can drastically impact hydraulic system performance. The models that predict the component failures in hydraulic systems are trained individually and often fail to capture the interconnected features and nature of these systems.

Predicting each component's condition independently can lead to a loss in extracting some features shared commonly by these conditions. This results in low performance in maintenance prediction. This issue can be addressed by providing an approach that considers the shared and correlated features across the system. Transfer Learning allows the model to learn shared knowledge between the two tasks. The study conducted by Wu et al. has shown the effectiveness of transfer learning for fault diagnosis and detection [9]. Deep transfer learning for machinery fault detection and recent developments in transfer learning for fault detection have been presented [10]. The transfer learning-based approach for Fault diagnosis of dynamic systems based on federated neural networks [11]. The survey paper for transfer learning researched many transfer learning techniques from different research papers and their process [12]. Transfer learning can improve the generalization ability of the model and will enable us to capture the shared characteristics between different components in the system [13]. This paper proposes a multi-task transfer learning approach that can simultaneously predict pump leakage and valve switching conditions [14, 15]. The method involves the extraction of standard features that are shared by both tasks. A model can efficiently learn the interrelated aspects by leveraging these standard features and characteristics. This results in generalization across both prediction tasks and improved performance.

2 Experimental Setup

2.1 Datasets

Condition monitoring of hydraulic systems starts with data extraction through the sensors. The dataset used in this study is from the University of California (UCI) Machine Learning Repository [16].

The experimental setup consists of a hydraulic test rig which comprises two circuits:

- Hydraulic circuit
- Cooling circuit

The hydraulic circuit consists of various components like a pump, accumulator, valves, and variable load, as shown in (Fig. 1). The dataset includes numeric sensor data from the hydraulic test rig varying by sensor type and sampling rate. For each sensor, we have 2205 instances. The pressure sensors (PS1-6) and motor power sensors (EPS1) work at 100 Hz frequency and have 6000 attributes. The volume flow sensors (FS1-2) work at a 10 Hz frequency and yield 600 attributes. Temperature sensors (TS1-4), vibration sensors (VS1), efficiency factor (SE), cooling efficiency sensor (CE), and power sensors (CP), operate at a 1 Hz frequency. Each consists of 60 attributes. There is no missing attribute value. The class distribution includes the continuous value, representing the degradation process over time with distinct categories for cooler condition, valve condition, internal pump leakage, hydraulic accumulator pressure, and a stable flag indicating the system stability.

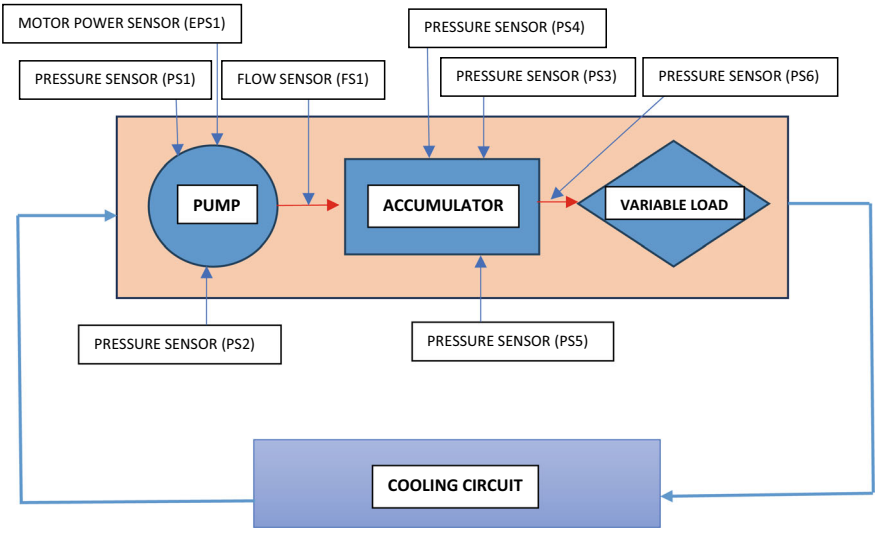


Fig. 1 Block diagram of the hydraulic systems

The model was trained on two distinct Targets. The first Target was Pump leakage, and the second was the Valve Condition dataset. For the Pump leakage dataset, the features that were selected, EPS1 and FS1, were included due to their high correlations with Pump Leakage severity, as shown in Fig. 2, and the labels indicating the pump leakage were No Leakage, Weak Leakage, and Severe Leakage. The next dataset was the Valve Condition dataset. The selected features were pressure sensors such as PS1, PS2, PS3, PS4, PS5, PS6, and the efficiency factor that is SE with labels representing the conditions from optimal to failure. The pre-trained models were initially developed using these datasets, which were then used for multi-task transfer learning to improve performance in predicting pump leakage and Valve condition. We have multiple attributes for each sensor because samples were taken using different sampling rates. The mean values of these attributes for each sensor were calculated and then further utilized for the model training. For model training, the dataset was split into 80:20 for training and test respectively, additionally while training we split our dataset into 80:20 as training and validation sets. Due to this proper monitoring of the model during training can be done.

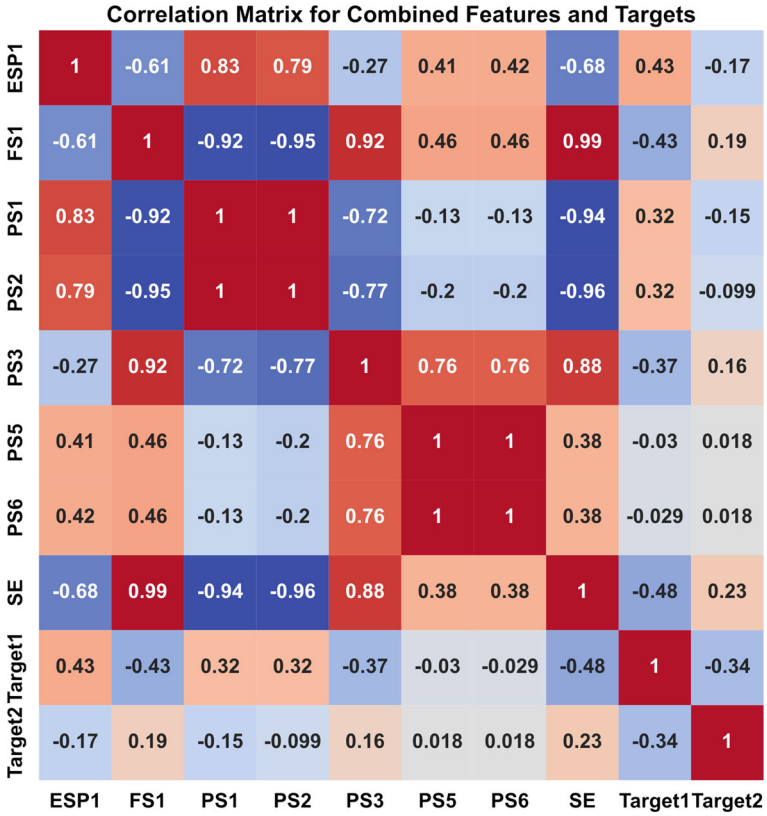


Fig. 2 Correlation matrix

On plotting the graph for the electric motor power sensor and time (Fig. 3), it was observed that the power utilized by no leakage was less than the severe leakage and the weak leakage. Moreover, it is correct that there is some leakage in the pump, so the power required will be high. In a similar case, the flow in the case of no leakage is high, and for weak leakage, it is somewhat low, and for severe leakage, it is lower.

3 Methodology

3.1 Pearson Correlation Analysis

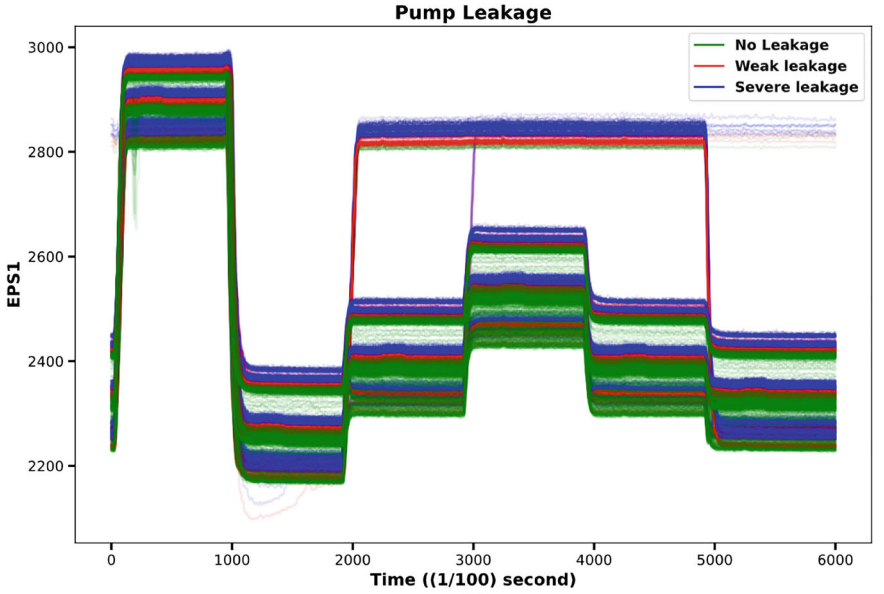
Before combining the datasets into a multi-task transfer model, the Pearson correlation analysis was carried out between the features of the pump leakage dataset and the valve condition dataset. Pearson correlation measures the linear relationships between the variables and is given by Eq. (1) [17].

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} \tag{1}$$

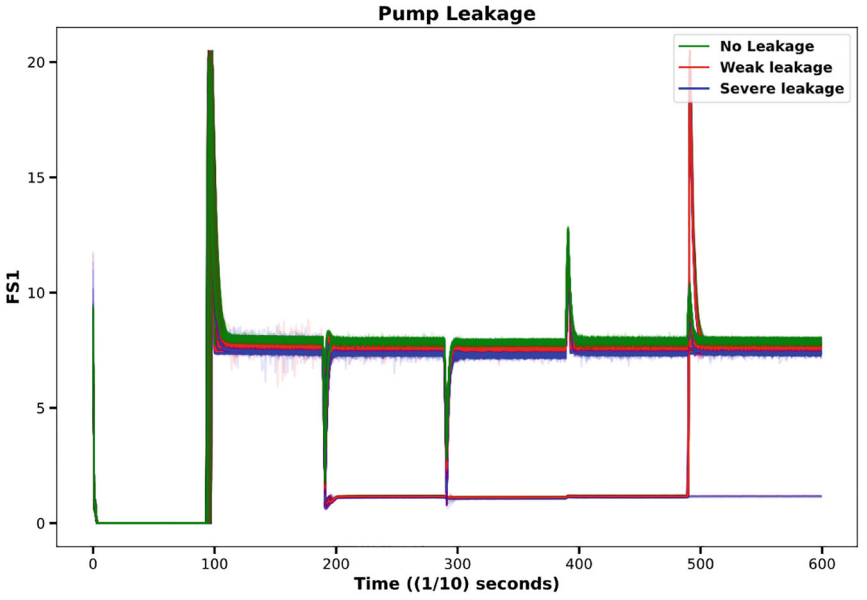
Pearson correlation analysis helps to understand how changes in one dataset might affect the changes in other datasets. The high positive value of correlation indicates a positive relationship, and the high negative value indicates a negative relationship, as shown in Fig. 4. This suggests that the shared representations in the transfer learning model could be good. The correlation value between the pump leakage and valve condition was − 0.34 (Fig. 2). This shows a moderate relationship between the two tasks. The Pearson correlation analysis revealed that many features from both datasets showed significant correlations. This justifies the use of shared feature extraction layers since the relationships between the tasks let the model capture joint patterns, making the predictive performance of both functions better.

3.2 Transfer Learning Framework

Transfer learning uses pre-trained models from one domain and applies the learned knowledge and features to related tasks. Here, two pre-trained models were utilized. The first pre-trained model was pump leakage data, and the second was pre-trained on valve condition data. Then, both models were stripped of their final classification layer, and only feature extraction layers were retained. These feature extraction layers were concatenated into a single multitask model to benefit from shared representation learning.



(a)



(b)

Fig. 3 **a** and **b** shows the variation of motor power and flow respectively for all instances which consists of No leakage, weak leakage and Severe leakage

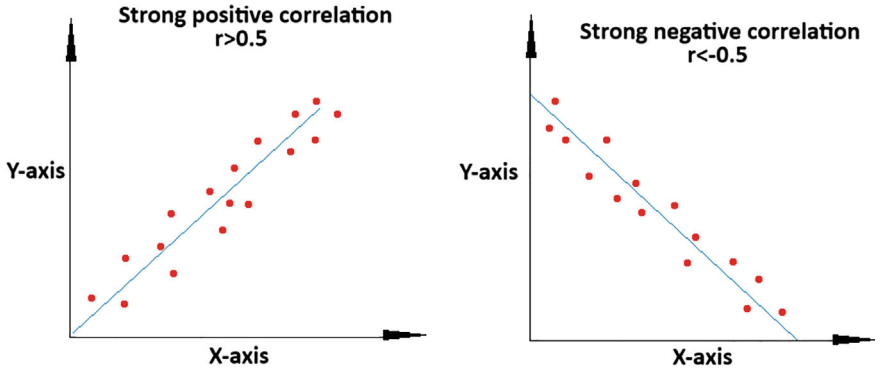


Fig. 4 Strong positive and negative correlation

3.3 Model Architecture

The model proposed employs a multi-task transfer learning model that effectively utilizes a shared feature and characteristics to handle two tasks that have some shared features, predicting pump leakage and valve conditions Fig. 5. The Model Contains Common Features Layers and distinct input layers, one for pump leakage data, which consists of two input nodes, and the other for valve condition, which consists of six input nodes. Each input is processed through a pre-trained feature extraction layer that was derived from models that handle their respective input dimensions. The features which we extracted were then concatenated to form a one-feature representation. This combined feature set is then passed through two shared dense layers to capture a joint pattern beneficial for both tasks, where the dense layers consist of 128 nodes with ReLU (Rectified Linear Unit—Eq. (2)) activation function in the first layer and the second layer consists of 64 nodes with ReLU activation functions. These shared layers allow the model to get the correlated natures of data sets. Thus, more efficient learning is done. Following the shared dense layer, our model has task-specific output layers. The first task was to predict pump leakage, which consists of three nodes and a softmax activation function—Eq. (3). The three classes were no leakage, weak leakage, and severe leakage, and the other task was to predict valve condition, which consists of four nodes and a softmax activation function. The four classes were optimal, small lag, severe lag, and total failure. Our model enhances the predictive performance by reducing the training time and using shared learning features. The model compiled with sparse categorical cross entropy as a loss function for both tasks and uses Adam optimization with a learning rate 0.001. The early stopping technique using custom accuracy checkpoints was used to avoid overfitting and ensure that model performance is good. The checkpoint saves the model when the task accuracies are better. The model was trained for 400 epochs, and the saturation point was between 200 and 250 epochs, as shown in Figs. 6 and 7. However, custom accuracy checkpoints were also used to get good accuracy and

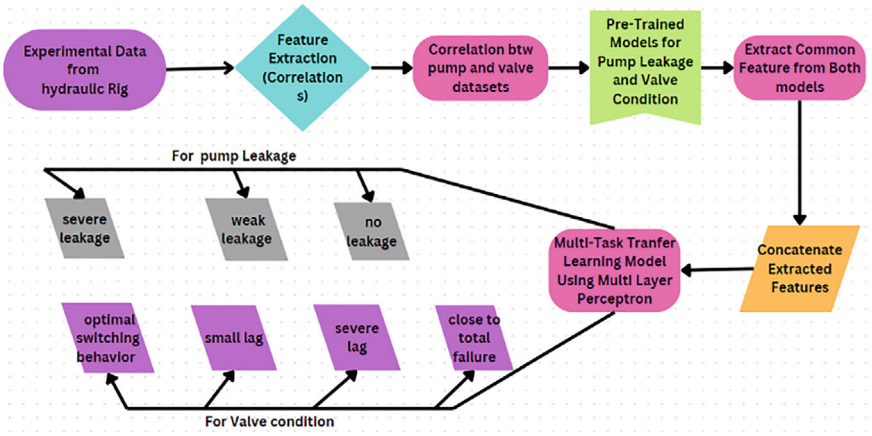


Fig. 5 Flow chart for proposed model architecture

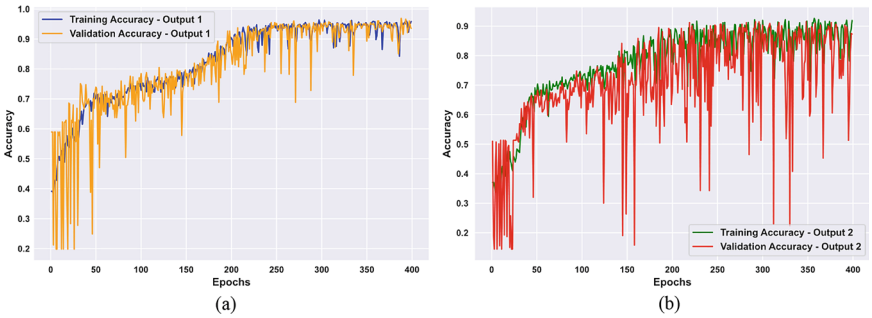


Fig. 6 Epochs versus accuracy graph **a** for pump leakage and **b** for valve condition

ensure overfitting and underfitting.

$$\mathrm{ReLU(x)} = \max(0,\ \mathrm{x}) \tag{2}$$

$$softmax(Z_i) = \frac{e_{z_i}}{\sum_{i=0}^{n} e_{z_j}} \tag{3}$$

4 Results and Discussion

For the evaluation of our presented model, we have used precision, recall, and F1 score. We have used these evaluating values because we have unbalanced data for different conditions and accuracy alone cannot tell the model's performance

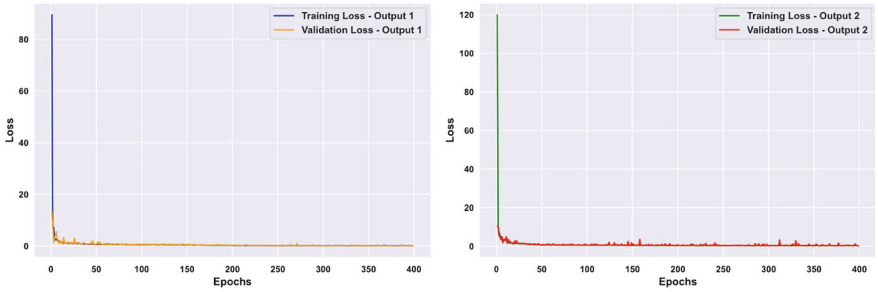


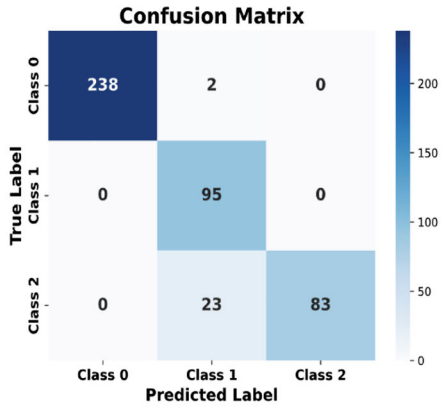
Fig. 7 Epochs versus loss **a** for pump leakage and **b** for valve switching condition

for unbalanced datasets. These evaluating values such as precision and recall also describe how the model has predicted every condition individually. The F1 score tells the combined effect of precision and recall using the harmonic mean of both values—Eq. (4). The model's performance is directly proportional to the F1 score. As the F1 score increases, the model's performance improves.

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

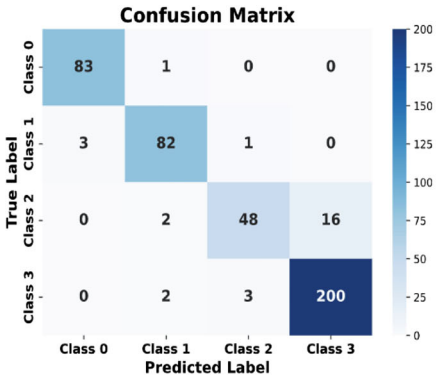
The confusion matrix for both tasks revealed that the model was highly effective in distinguishing between different failure modes for pump leakage and valve conditions (Fig. 8). The class representation is given in tabular form in Fig. 8. The model shows good predictive ability for no leakage with 238 correct predictions and almost no false positives. For valve condition, the model has achieved good accuracy for the optimal switching behaviour with 200 correct predictions and minimal misclassification. The F1 score for no leakage condition in the pump was found to be 0.9958, which is a nearly perfect score, and the F1 score for close to failure and optimal switch for valve condition were found to be 0.9765 and 0.9501. The overall accuracies for the pump leakage and valve condition were 94.33% and 93.65%, respectively.

The main observation from the study is that it significantly improved the performance for both tasks. The shared feature extractor allowed the model to capture the common features between the systems, such as pressure variants that affect both the pump and valve. Besides that, transfer learning had higher performance, as observed from Tables 1 and 2, lower false positives, and faster convergence during training.



Class	Pump leakage
Class 0	No leakage
Class 1	Weak leakage
Class 2	Severe leakage

(a)



Class	Pump leakage
Class 0	Close to failure
Class 1	Severe lag
Class 2	Small lag
Class 3	Optimal switching behaviour

(b)

Fig. 8 Confusion matrix **a** for pump leakage condition and **b** for valve switching condition

Table 1 Classification matrix table for pump leakage

Pump condition	Precision	Recall	F1 score
No leakage	1.0000	0.9917	0.9958
Weak leakage	0.7917	1.0000	0.8837
Severe leakage	1.0000	0.7830	0.8783

Table 2 Classification matrix table for valve switching condition

Valve behaviour	Precision	Recall	F1 score
Optimal switching behaviour	0.9259	0.9756	0.9501
Small lag	0.9231	0.7273	0.8136
Severe lag	0.9425	0.9535	0.9480
Close to failure	0.9651	0.9881	0.9765

5 Conclusion

Multi-task transfer learning can significantly improve the predictive maintenance in hydraulic systems by sharing knowledge between the related tasks, such as pump leakage and valve condition monitoring. The multi-task model had an accuracy of 94.33% and 93.65% for pump Leakage and Valve conditions, respectively, and fast training convergence compared to traditional task-specific models. Utilizing a multi-layer perceptron in our model for transfer learning has shown a significant impact due to its ability to learn non-linear features and interactions. Using transfer learning, the knowledge gained from one data set can be used for another related model, which results in the good performance of the further trained models. In addition to the transfer of knowledge between two datasets, we have also built a generalized model that can classify datasets with good accuracy. In future work, this approach will be extended to other hydraulic systems, such as accumulators and cooler conditions, to improve the system's predictive maintenance capability further.

Acknowledgements This study is supported by Science and Engineering Research Board (SERB), DST, India, under Core Research Grant (CRG), (File No. CRG/2022/007409). The authors wish to acknowledge the funding agency.

References

1. Hoblit, F.: Critical buckling for hydraulic actuating cylinders. In: Stress Engineer, Lockheed Aircraft Corporation. Product Engineering, pp. 108–112 (1950)
2. Safi, S., Mozar, S.: From Reactive Maintenance to Proactive Preventive Maintenance System. ICOMS-2004, pp. 1–8. Sydney (2004)
3. Saxena, A., Saad, A.: Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems. Appl. Soft Comput. **7**(1), 441–454 (2007)
4. Prakash, J., Kankar, P.K.: Health prediction of hydraulic cooling circuit using deep neural network with ensemble feature ranking technique. Measurement **151**, 107225 (2020)
5. Rao, B.K.N.: Handbook of Condition Monitoring, 1st edn. Elsevier Advanced Technology (1996)
6. Prakash, J., Kankar, P.K., Miglani, A.: Monitoring the degradation in the switching behavior of a hydraulic valve using recurrence quantification analysis and fractal dimensions. J. Comput. Inf. Sci. Eng. **21**(6), 061010 (2021)
7. Prakash, J., Kankar, P.K., Miglani, A.: Internal leakage detection in a hydraulic pump using exhaustive feature selection and ensemble learning. In: 2021 International Conference on Maintenance and Intelligent Asset Management (ICMIAM), pp. 1–6, IEEE (2021)
8. Prakash, J., Kankar, P.K., Miglani, A., Tamhankar, R.: Internal leakage diagnosis of a hydraulic cylinder using C-LSTM neural network. In: International Conference on Reliability, Safety, and Hazard, pp. 51–57, Springer Nature Singapore, Singapore (2024)
9. Wu, H., Zhao, J.: Fault detection and diagnosis based on transfer learning for multimode chemical processes. Comput. Chem. Eng. **135**, 106731 (2020)
10. Li, C., Zhang, S., Qin, Y., Estupinan, E.: A systematic review of deep transfer learning for machinery fault diagnosis. Neurocomputing **407**, 121–135 (2020)

11. Chen, H., Chai, Z., Jiang, B., Huang, B.: Data-driven fault detection for dynamic systems with performance degradation: a unified transfer learning framework. IEEE Trans. Instrum. Meas. **70**, 1–2 (2020)
12. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. J. Big data **3**, 1–40 (2016)
13. Torrey, L., Shavlik, J.: Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques, pp. 242–264, IGI Global, Hershey (2010).
14. Zhang, Y., Yang, Q.: An overview of multi-task learning. Natl. Sci. Rev. **5**(1), 30–43 (2018)
15. Zhang, Y., Yang, Q.: A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. **34**(12), 5586–5609 (2021)
16. Helwig, N., Pignanelli, E., Schtze, A.: Condition Monitoring of Hydraulic Systems. UCI Machine Learning Repository (2015). https://doi.org/10.24432/C5CW21
17. Cleophas, T.J., Zwinderman, A.H.: Bayesian Pearson correlation analysis. In: Modern Bayesian Statistics in Clinical Research, pp. 111–118. Springer, Cham (2018)

Condition Monitoring of Hydraulic Systems Using Multi-output Classification Convolutional Neural Networks

Mahak Kokate, Tarun Kumar, Pavan Kumar Kankar, and Ankur Miglani

Abstract Hydraulic systems are crucial in aviation and construction applications, efficiently transmitting heavy loads with minimal effort. Condition monitoring techniques are essential for optimizing the quality and performance of hydraulic systems. It serves as a tool for decision-making in maintenance activities, enabling more effective and informed operations. In modern industrial settings, monitoring machinery has become critical for enhancing the cost-efficiency of hydraulic systems. This paper introduces a deep learning (DL) framework utilizing convolutional neural network (CNN) for condition monitoring of hydraulic systems with raw data for health prediction of multiple components such as valves, cooler, accumulators and pump. This method uses CNN to analyze raw time-series sensor data with varying sampling frequencies (100, 10, and 1 Hz) in which the input matrix of the CNN model is a 3-D matrix and predicts the functional state of the hydraulic system. Demonstrating the effectiveness of CNNs in handling multi-output classification tasks by achieving high accuracy across all targets, and evaluated with accuracy, recall, precision, and F1-score and achieved an overall accuracy of 98.5%. This study illustrates the effectiveness of DL techniques in condition monitoring of hydraulic systems for multi-target classification tasks.

Keywords Hydraulic systems · Condition monitoring · Time-series data · Multioutput-CNN model · Classification tasks

M. Kokate · T. Kumar · P. K. Kankar (✉)
System Dynamics Lab, Department of Mechanical Engineering, Indian Institute of Technology Indore, Indore, India
e-mail: pkankar@iiti.ac.in

A. Miglani
Microfluidics and Droplet Dynamics Lab, Department of Mechanical Engineering, Indian Institute of Technology Indore, Indore, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_27

1 Introduction

Various industries, including manufacturing, aerospace, and construction, use hydraulic systems extensively due to their ability to transmit large amounts of power with high precision. The reliability and efficiency of these systems are critical because their failures can lead to costly repairs and even safety hazards. The early detection of faults and anomalies can prevent costly downtimes. Condition monitoring (CM) observes parameters such as pressure, temperature, vibration, flow, and power consumption to provide insights into system performance. Various industries present a challenge for condition monitoring due to their complex dynamics and multivariable nature [1, 2]. Traditionally, this task relied on manual investigation of sensor data or rule-based systems, which can be time-consuming and prone to errors and can't predict intricate relationships between the sensor data and the system's health state effectively [3]. Machine learning (ML) techniques are continuously increasing in various industry domains. Using Machine learning (ML) models into CM applications, sensor-based systems have become more data-intensive. Supervised machine learning techniques, including decision trees, support vector machines (SVM), and random forests, have been utilized to detect the patterns in the data [4, 5]. Gupta et al. [6] performed a mathematical model-based Prediction of an Axial Piston Pump (APP) leakage fault cylinders. Prakash et al. [7, 8] monitored the condition of the directional control valve (DCV) of the hydraulic system and internal leakage of pump using ML. A deep neural network (DNN) combined with an ensemble feature ranking method was employed for the condition monitoring and prediction of Cooling circuit [9]. Data augmentation and deep learning were used to monitor the hydraulic system [10]. The condition of a 2-stage Hydraulic Cylinder is classified using a semi-supervised approach TS-SVM [11]. The recent advancement in CM using ML directly influences the maintenance and potential losses [12]. With advancements in sensor technology and data analysis, modern CM techniques now employ machine learning and artificial intelligence (AI) models to predict system failures more accurately. Many existing models were proposed for CM that focused on single target prediction, but they lack the ability to monitor multiple component conditions simultaneously. Recent enhanced performance in DL, specifically Convolutional Neural Networks (CNNs), have shown the capability of automatic learning features from raw sensor data for classification tasks [13, 14]. This study investigates and explores the use and application of CNNs for condition monitoring in hydraulic systems, focusing on a dataset consisting of multivariate time-series data sampled at varying frequencies. The present study directly uses the raw sensor data, unlike many conditions monitoring approaches that rely on extensive signal pre-processing or feature extraction (e.g., filtering, Fourier transforms, wavelet analysis). The raw data is reshaped into a 3D matrix and fed as an input to the CNN. This model has the ability to extract features directly from the raw time-series data. This methodology directly processes multivariate time-series data which is collected from multiple sensors operating at different frequencies (e.g. 100, 10, 1 Hz). The data used to train and evaluate the model includes 17 sensor data from sensors such as pressure, temperature, vibration,

flow, and virtual metrics, with 2205 instances. Each working cycle in the hydraulic system lasts 60 s and represents one or a single instance. The nature and description of the data used and utilized in this study are outlined in Sect. 2. This multi-output CNN model predicts the condition of multiple components present in the hydraulic system, which includes valves, accumulators, pump and cooler, along with predicting the system's stability. Since these are the main components in any hydraulic system, our model can monitor their conditions simultaneously. The complete working and architecture of the multi-output CNN model are described in Sect. 3. To evaluate accuracy, losses, precision rates, recall values and F1-score values are presented with the results in Sect. 4. The study concludes with a summary of findings in Sect. 5.

2 Dataset Description

2.1 Experimental Setup

The data used in this study is sourced from a hydraulic test rig designed to simulate various operational conditions. The dataset is retrieved from the UCI Machine Learning Repository [15]. The hydraulic system includes a primary operational circuit (Fig. 1a) and a secondary operational circuit which is a cooling-filtration circuit (Fig. 1b), both are connected to an oil tank. The system contains several sensors measuring process values such as volume flows (FS1, FS2), vibration (VS1), pressures (PS1–PS6), electrical power (EPS1) and temperature (TS1–TS5). In addition, for cooling power, cooling efficiency, and system efficiency which are the 3 virtual sensors and derived from physical sensor values. In total, 17 sensor signals are recorded over a consistent 60 s cycle and which repeats under varying conditions of the hydraulic system components.

2.2 Nature of Data

The data is categorized into three groups based on their sampling frequencies as given below:

- **100 Hz**: 7 sensors measuring pressure (bar) PS1 to PS6 and motor power (W) EPS1.
- **10 Hz**: 2 sensors related to volume flow (l/min) FS1–FS2.
- **1 Hz**: 8 sensors which measures temperature (°C) TS1–TS4, vibration (mm/s) VS1, and virtual metrics which are cooling power, system efficiency factor and cooling efficiency.

This dataset contains 2205 instances, each representing a time series with 43,680 attributes.

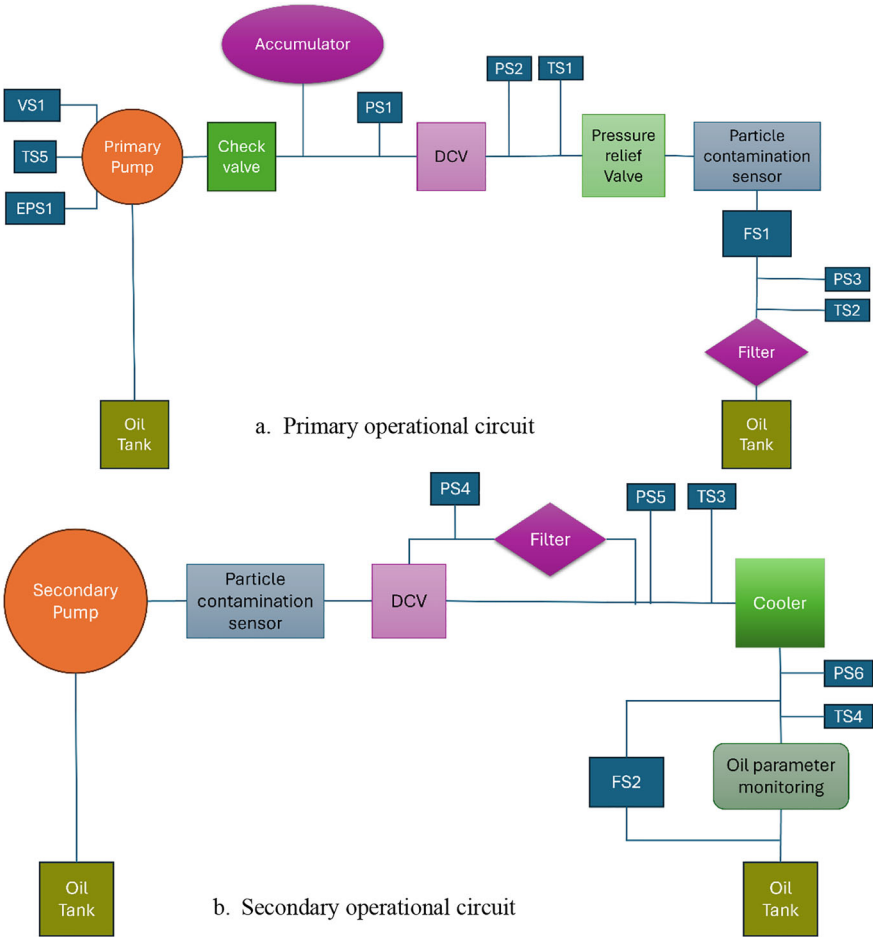


Fig. 1 **a** The primary operational circuit and **b** secondary operational circuit (cooling) of hydraulic system

Calculation of total number of Attributes: $7 \times 6000(100\text{Hz})) + 2 \times 600(10\text{Hz}) + 8 \times 60(1\text{Hz}) = 43{,}680$

The target variables include five indicators reflecting the condition of hydraulic system, as detailed in Table 1.

Table 1 Target classification and component conditions

Target	Component	Condition	Value	Interpretation
Target 1 (3 classes)	Cooler	Cooling power decrease (%)	100	Maximum efficiency
			20	Decreased efficiency
			3	Near complete failure
Target 2 (4 classes)	Valve	Switching degradation (%)	100	Ideal switching behavior
			90	Minor delay
			80	Major delay
			73	Near complete failure
Target 3 (3 classes)	Pump	Internal leakage	0	No leakage
			1	Minor leakage
			2	Major leakage
Target 4 (4 classes)	Accumulator	Gas leakage (bar)	130	Ideal pressure
			115	Moderately reduced pressure
			100	Significantly reduced pressure
			90	Near complete failure
Target 5 (2 classes)		Stability of the system	0	Stable conditions
			1	Static conditions possibly unreached

3 Methodology

3.1 Data Preprocessing

Since the data is multivariate time-series data collected at multiple frequencies, the raw time-series data was resampled to a common frequency of 1 Hz to align the sensors with different sampling rates. The data was then normalized using standard scaling to ensure uniformity across features. The sensor data was padded or truncated to ensure all instances had the same number of features, resulting in a final shape of (2205, 60, 17) for the input data for the CNN model. The target variables were converted into categorical form using one-hot encoding, resulting in five output vectors corresponding to the five targets, which is necessary for the model to classify the target classes.

This pre-processing step ensured that the input data fed into the CNN model was on a comparable scale. The pre-processed data is organized into a three-dimensional

format suitable for input into the CNN, with dimensions corresponding to the number of instances (2205), the number of time steps (60), and the number of features (17).

3.2 Model Architecture

This study employs a 1D Convolutional Neural Network (CNN) to handle time-series data. A flow chart of the multioutput CNN model is illustrated in Fig. 2, in which model consists of following key components:

- **Input Layer**: Receives the input time-series data with a shape of (60, 17).
- **Conv1D Layer**:

$$Output_{Conv1D} = ReLU\,(Conv1D(W \cdot X + b)) \tag{1}$$

  where W and b represents the weights and biases of the convolutional layer. X is the input to the layer. $ReLU\,(z) = max(0, z)$ is the activation function applied elementwise [16–18].
- **Dropout Layer**:

$$Output_{Dropout} = X \cdot M \tag{2}$$

  where M is a binary mask applied during training with a dropout rate [17, 19], where each element is 0 with probability p (the dropout rate) and 1 with probability $1 - p$.

  Each element mi of the mask M is drawn from a Bernoulli distribution:

$$mi \sim Bernoulli(1 - p) \tag{3}$$

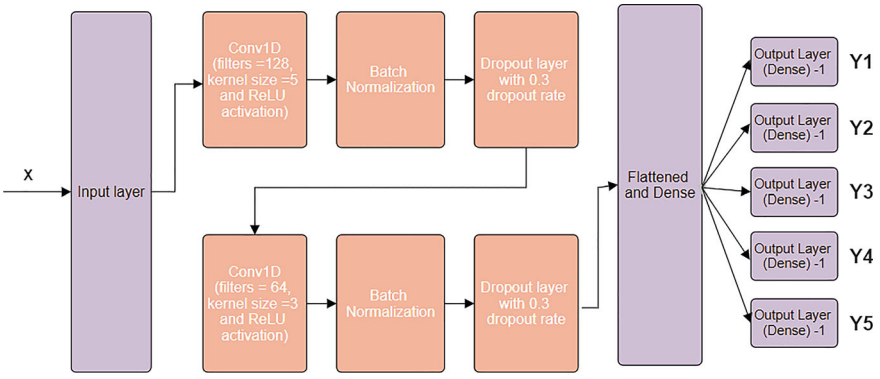


Fig. 2 Multi-output CNN model

- **Flattening Layer**: Transforms the convolved and pooled features into a 1D vector suitable for input into the dense layers.
- **Dense Layers**: Five separate dense layers with SoftMax activation functions were used to predict the five target variables. The output of each dense layer can be formulated as [17, 18]:

$$Output_{Dense} = Softmax(W \cdot X + b) \tag{4}$$

  where W and b denote the weights and biases of the dense layer. SoftMax function transforms logits into probabilities.
- **Output Layer**: An output layer equipped with SoftMax activation generates a probability distribution across the possible classes for each target variable. Five output layers have different numbers of classes to predict the condition of five targets.

3.3 Loss Function and Optimization

Given that the model is a multi-output model, a separate loss function is defined for each target. For a target with C_j classes, the categorical cross-entropy loss for a single instance is given by [18, 20]:

$$L_j = -\sum_{c=1}^{k} y_{jc} \cdot log\left(\widehat{y_{jc}}\right) \tag{5}$$

where y_{jc} is the true label, and $\widehat{y_{jc}}$ is the predicted probability for class c in target j. The total loss L for the model is computed and obtained by aggregating the losses for all target variables:

$$\text{Total loss}(L) = \sum_{c=1}^{k} L_j \tag{6}$$

The Adam optimizer optimized the model, which is well-suited for handling the complex parameter space of deep learning models.

For multi-output models, the accuracy for each output is computed separately. The overall or total accuracy can be calculated as the average of the individual accuracies:

$$\text{Accuracy} = \text{Total number of correct predictions/total number of predictions} \tag{7}$$

$$\text{Total accuracy } = 1/5 * \sum_{k=1}^{5} Accuracy_k \tag{8}$$

3.4 Training and Evaluation

The dataset was separated or partitioned using 80/20 ratio into training and testing set. Additionally, the training data was further partitioned into training and validation subsets following an 80/20 division. The model underwent training for 100 epochs with a batch size of 32. During training, the CNN model optimized by minimizing the categorical cross-entropy loss function to update the model weights. To avoid overfitting, the model's performance was regularly assessed on the validation subset throughout training process. The model's predictive capability to the unseen testing set was evaluated and examined using key performance measures such as accuracy, precision rates, recall and F1-score values for each target variable.

4 Results and Discussion

This multi-output CNN model has achieved high accuracy across all target variables during testing. The classification performance for each target is summarized in Table 2, which displays the accuracies and losses corresponding to each target. Table 3 represents the precision rates, recall values, and F1-score values for individual class within every target. Finally, the overall accuracy and losses for this multi-output CNN are presented and tabulated in Table 4. These accuracies and values are listed using test dataset. To assess the model's training behaviour, accuracy trends for both training and validation are plotted against the number of epochs for each target/output in Fig. 3. To further evaluation, Fig. 4 displays the confusion matrix for each target component to evaluate the model's overall performance.

The results show that the CNN model successfully identified intricate patterns within sensor data, achieving high accuracy for all target variables. This study highlights the capability of DL techniques for condition monitoring in hydraulic systems, particularly when handling multivariate time-series data. One notable challenge was aligning data from sensors with different sampling frequencies. By resampling the data to a common frequency, the information from all sensors could be effectively integrated, and this method does not need any feature extraction techniques like Fourier transform, wavelet analysis, etc. It has reduced the dependency on feature extraction and the time to train and pre-process the model. This approach could be generalized to other industrial applications where sensor data is collected at varying rates.

Table 2 Model's performance on test data for each target component

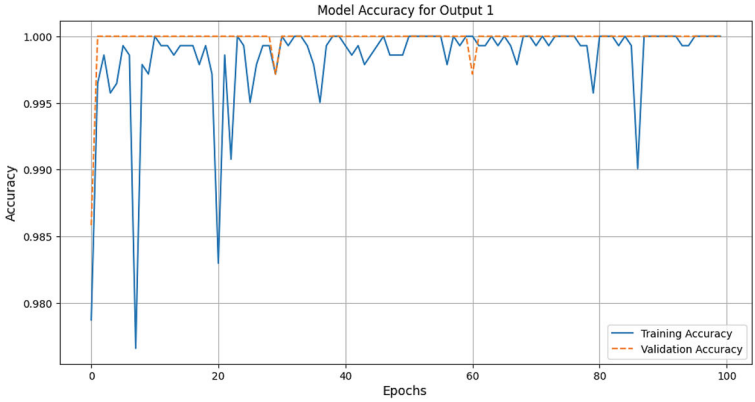
Target	Target 1	Target 2	Target 3	Target 4	Target 5
Accuracy	1.00	1.00	0.9977	0.972789	0.9546
Loss	5.28e−07	0.0092161	0.007400	0.114233	0.159111

Table 3 For each class in each target precision, Recall, and F1-score value

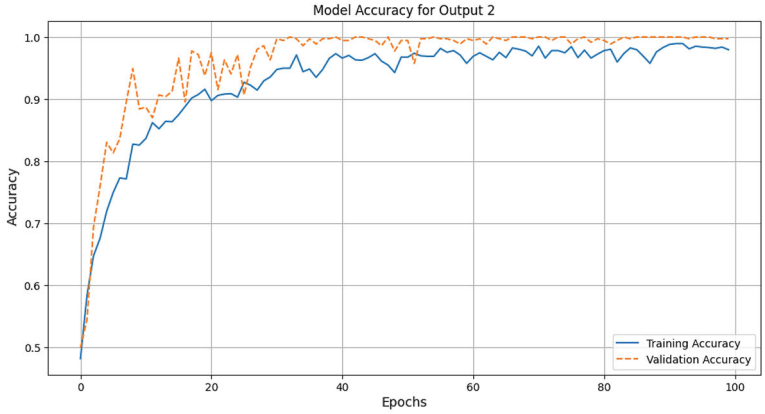
Output/target	Classes	Precision	Recall	F1-score
1	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
	2	1.00	1.00	1.00
2	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
	2	1.00	1.00	1.00
	3	1.00	1.00	1.00
3	0	1.00	1.00	1.00
	1	0.99	1.00	1.00
	2	1.00	0.99	1.00
4	0	0.97	0.99	0.98
	1	0.96	0.93	0.95
	2	0.99	0.95	0.97
	3	0.98	0.99	0.98
5	0	0.94	0.99	0.97
	1	0.98	0.88	0.93

Table 4 Multi-output CNN model's total loss and accuracy on test data

Total loss	0.289962
Total accuracy	0.985034



(a) Accuracy versus Epochs for target 1

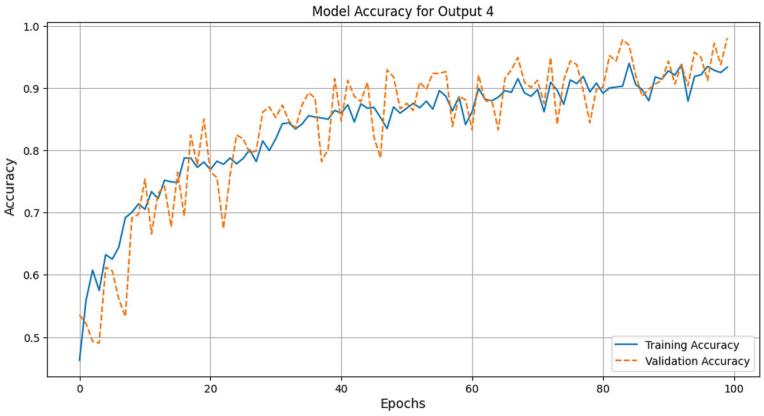


(b) Accuracy versus Epochs for target 2

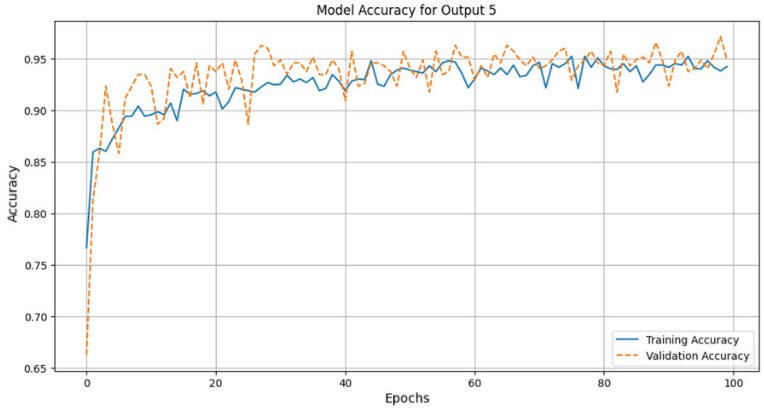


(c) Accuracy versus Epochs for target 3

Fig. 3 **a** Accuracy vs epochs curve for target/output 1, **b** accuracy vs epochs curve for target/output 2, **c** accuracy vs epochs curve for target/output 3, **d** accuracy vs epochs curve for target/output 4, **e** accuracy vs epochs curve for target/output 5



(d) Accuracy versus Epochs for target 4



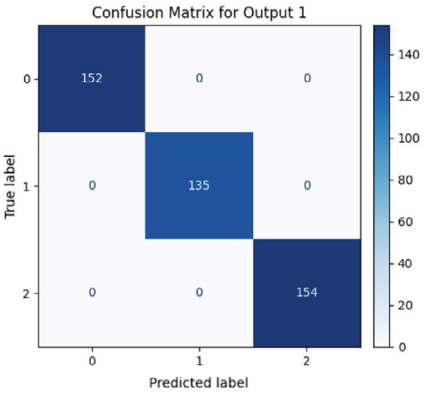
(e) Accuracy versus Epochs for target 5

Fig. 3 (continued)

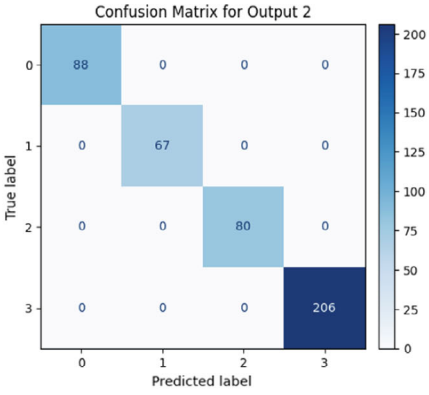
5 Conclusion

The study shows the effectiveness of CNNs for condition monitoring of hydraulic systems. The proposed method effectively handles varying data frequencies. It performs accurate multi-class classification of component health states for cooler, pump, valve, and accumulator and stability of the system using raw data directly and accurately predicts the stability of the hydraulic system. This approach can be adapted to real-world hydraulic system applications, enabling predictive maintenance and improved system reliability. The conclusions from this study are as follows:

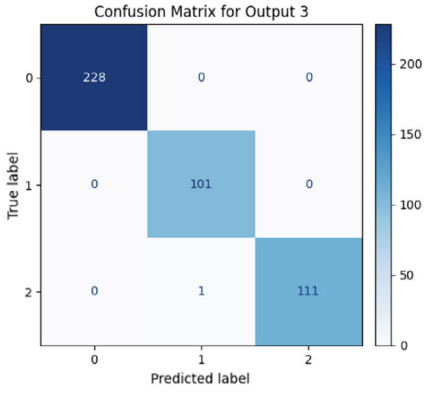
a. The CNN model has successfully learned the raw time-series data using a 3D matrix as input without any feature extraction techniques. Data is resampled to be in the same frequency of 1 Hz and then normalized using a standard scale to ensure uniformity across features.



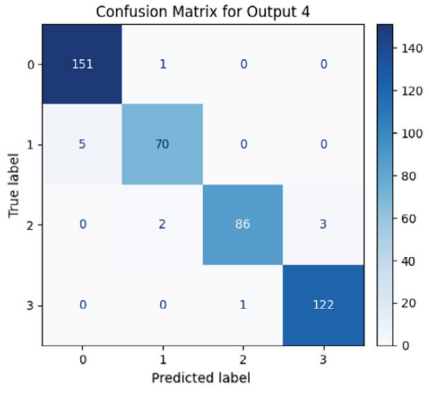
(a) Confusion Matrix for target 1



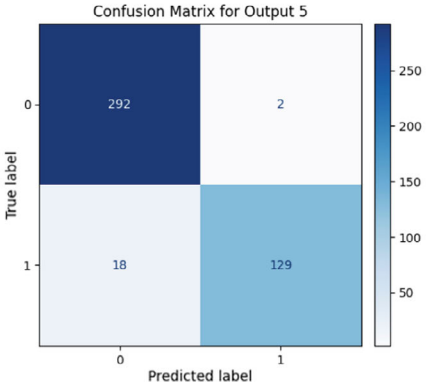
(b) Confusion Matrix for target 2



(c) Confusion Matrix for target 3



(d) Confusion Matrix for target 4



(e) Confusion Matrix for target 5

Fig. 4 **a** Confusion matrix for target/output 1, **b** confusion matrix for target/output 2, **c** confusion matrix for target/output 3, **d** confusion matrix for target/output 4, **e** confusion matrix for target/output 5

b. The multi-output CNN model has achieved high accuracies across all the target variables in the health prediction of components such as cooler (100%), valve (100%), pump (99.77%), and accumulator (97.28%) along with the prediction of system stability (95.46%).
c. This CNN-based approach for multi-output classification of multivariate time-series data from a hydraulic test rig demonstrated a total accuracy of 98.5% overall in condition prediction of hydraulic components and health states.

Acknowledgements This study is supported by Science and Engineering Research Board (SERB), DST, India, under Core Research Grant (CRG), (File No. CRG/2022/007409). The authors wish to acknowledge the funding agency.

References

1. Helwig, N., Pignanelli, E., Schütze, A.: Condition monitoring of a complex hydraulic system using multivariate statistics. In: 2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, pp. 210–215. IEEE, Pisa (2015)
2. Peixin, G.A., Tao, Y.U., Zhang, Y., Jiao, W.A., Jingyu, Z.H.: Vibration analysis and control technologies of hydraulic pipeline system in aircraft: a review. Chin. J. Aeronaut. **34**(4), 83–114 (2021)
3. Keleko, A.T., Kamsu-Foguem, B., Ngouna, R.H., Tongne, A.: Health condition monitoring of a complex hydraulic system using deep neural network and DeepSHAP explainable XAI. Adv. Eng. Softw. **175**, 103339 (2023)
4. Upadhyay, N., Kankar, P.K.: Diagnosis of bearing defects using tunable Q-wavelet transform. J. Mech. Sci. Technol. **32**, 549–558 (2018)
5. Vakharia, V., Gupta, V.K., Kankar, P.K.: Application of chi square feature ranking technique and random forest classifier for fault classification of bearing faults. In: Proceedings of the 22th International Congress on Sound and Vibration, pp. 12–16. International Institute of Acoustics and Vibration (IIAV), Florence (2015)
6. Gupta, R., Miglani, A., Kankar, P.K.: Performance prediction of an axial piston pump with increasing severity of leakage fault in single and multiple cylinders. J. Dyn. Syst. Meas. Contr. **145**(2), 021001 (2023)
7. Prakash, J., Singh, S., Miglani, A., Kankar, P.K.: Pressure signal-based analysis of anomalies in switching behavior of a two-way directional control valve. ASME Open J. Eng. **2**, 021006 (2023)
8. Prakash, J., Miglani, A., Kankar, P.K.: Internal leakage detection in hydraulic pump using model-agnostic feature ranking and ensemble classifiers. J. Comput. Inf. Sci. Eng. **23**(4), 041005 (2023)
9. Prakash, J., Kankar, P.K.: Health prediction of hydraulic cooling circuit using deep neural network with ensemble feature ranking technique. Measurement **151**, 107225 (2020)
10. Kim, K., Jeong, J.: Deep learning-based data augmentation for hydraulic condition monitoring system. Proc. Comput. Sci. **175**, 20–27 (2020)
11. Prakash, J., Miglani, A., Kankar, P.K.: Semi-supervised approach using transductive support vector machine for internal leakage detection in two-stage hydraulic cylinder. J. Comput. Inf. Sci. Eng. **24**(8), 084504 (2024)
12. Surucu, O., Gadsden, S.A., Yawney, J.: Condition monitoring using machine learning: a review of theory, applications, and recent advances. Expert Syst. Appl. **221**, 119738 (2023)

13. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018)
14. Serradilla, O., Zugasti, E., Rodriguez, J., Zurutuza, U.: Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Appl. Intell. **52**(10), 10934–10964 (2022)
15. Helwig, N., Pignanelli, E., Schtze, A.: Condition monitoring of hydraulic systems. UCI Mach. Learn. Reposit. (2015). https://doi.org/10.24432/C5CW21
16. Son, T., Hong, D., Kim, B.: Multi-output classification based on convolutional neural network model for untrained compound fault diagnosis of rotor systems with non-contact sensors. Sensors **23**(6), 3153 (2023)
17. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
18. Goodfellow, I., et al.: Deep Learning. MIT Press (2016)
19. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
20. Huynh, N.: Understanding Loss Functions for Classification (2023). https://medium.com/@nghihuynh_37300/understanding-loss-functions-for-classification-81c19ee72c2a

Securing Enterprise Data with 4D Security Framework in Advanced Cyber-Physical Systems

Meenakshi Malik, Vijay Prakash Sharma, and Surya Prakash

Abstract The advent of Smart Enterprise and office systems has revolutionized modern workplaces, offering enhanced convenience, automation, and vast storage capabilities for sensitive business and professional data. However, these systems are increasingly exposed to sophisticated cyber threats, posing significant risks to the security and privacy of enterprise information. Cyber attackers can exploit vulnerabilities within smart devices, leading to unauthorized access, identity theft, financial fraud, and other detrimental outcomes. To address these concerns, this paper introduces a novel 4D Security Framework specifically designed for Advanced Cyber-Physical Systems. The proposed framework strengthens data security through proactive threat detection, dynamic access control, multi-layered encryption, and real-time response mechanisms. By incorporating these advanced protective measures, the framework mitigates the risks of unauthorized access, data breaches, and other cyber threats, safeguarding enterprise systems from potential financial and reputational damage. The paper also acknowledges the need to address key challenges in the field to further enhance the security and privacy of smart Enterprise and office systems.

Keywords Smart enterprise · Data security · Encryption · Steganography · Cryptography · 6G

M. Malik
BML Munjal University, Gurugram, India

V. P. Sharma (✉)
K.R. Mangalam University, Gurugram, India
e-mail: vijay.bml2017@gmail.com

S. Prakash
Great Lakes Institute of Management, Gurugram, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_28

1 Introduction

Automation systems for an enterprise and workplace are now commonplace. Each business owner dreams of touching the zenith of success and tries to minimize the risk of breach of his company's USPs, business skills, confidential and professional documents, customer data back, and at the same time desires to ensure the data security systems be very user friendly and safe. The term "Smart Enterprise" can be interchangeably referred to by several synonymous terms or secure offices, e.g., "Enterprise or office Automation", "Assistive technology", "digital office", "automated office," "Smart Connected office or Enterprise" and "Intelligent workplace." Smart Enterprise and office systems that connect to the Internet of Things (IoT) are already a part of our daily lives. In a nutshell, smart Enterprises encompass features like Enterprise automation, connected appliances, security and surveillance, energy management, entertainment and media integration, voice control with virtual assistants, and health and wellness devices, all aimed at enhancing safety, ease of work, convenience, efficiency, and security in everyday living. IoT, plays a vital role in smart Enterprises, which is growing quickly because of several things, such as the fact that IoT devices are getting easily available and cheaper and more companies want new amenities now. Sensors, actuators, display screens, electronic appliances, and software all work together to create a "smart Enterprise" or "smart office," a miniature physical world communicating and sharing information with respective company employees and owners. Smart Enterprise and office environments aim to enhance the data security levels and ease of work of employees by providing automated, enterprise, and secure services through the processing of information and data. The goal is to improve the overall existence of the firm including its safety, and security. Even though smart Enterprise and office systems are multiplying, easy to use, and offer benefits, they pose security and privacy risks to their users. Still, attackers started using learning algorithms to break into smart Enterprises and offices' security systems and get Enterprise information. A user's Enterprise information could be the employee's personal and professional details, where they live, where they go, business secrets, or any other information about the company that definitely is not desirable to be shared with any outsider. A smart Enterprise or office backup system stores sensitive information like business data, customer contacts, Enterprise photos, videos, and digital information. Most threats fall into two main categories: First is data privacy and second is context-aware privacy. When exchanging valuable information about the business, Enterprise life, or anything else, data privacy is an important thing to think about. Since everything will be linked and, on the internet, hackers can easily get into the network. By just getting into a part of a network, a hacker can find out everything about an enterprise or business. The context-aware privacy, also called "situational privacy" means seeing, finding, and tracking users' movements, business or Enterprise activities, and actions, and then using this information for bad. Context-aware technology is an essential and useful tool for the business world. Context gives important information about a situation or data and makes security risks worse because people could abuse employees' individuality, location, movement,

and behavior. So, trust and security protection need to be dealt with smartly. There is always a threat of losing system confidentiality or Enterprise data, which can cause a loss of security, privacy, and even financial loss. So, privacy and security are crucial in smart Enterprise security systems due to the handling of sensitive Enterprise data, increased attack surface, potential market harm, privacy concerns, network vulnerabilities, integration with external services, and regulatory compliance. Inadequate security measures can compromise physical safety by allowing unauthorized access to the system. Privacy concerns arise from the collection and transmission of user data by smart Enterprise devices. Network vulnerabilities and insecure integrations can lead to unauthorized access, data interception, and disruptions. The increased integration of smart technologies, IoT devices, and cyber-physical systems (CPS) in enterprises is driving up the requirement for corporate data security. Companies are more vulnerable to cyberthreats like ransomware, phishing, and insider assaults as they produce and retain large volumes of sensitive data. Strong security solutions are in great demand now more than ever since cybercrime losses are expected to exceed billions of dollars yearly by the next year. The business has potential to cross trillions of turnovers by 2025. To guarantee proactive threat detection and multi-layered encryption, there is a special need for AI-driven security solutions, cloud-based protection, or data-centric frameworks. At the same time, compliance with data protection regulations is necessary to avoid legal consequences and maintain trust. Enhancing privacy and security measures in smart Enterprise security systems mitigates these risks and ensures the safety and privacy of Enterprises' data and security. Online facilities are watchfully knotted with the issues regarding obtainability, truthfulness, privacy, and information verification exchanged over communication media which has steered to the development of information hiding techniques such as Steganography and Cryptography [1] for the protection of online and offline data and communication. These can be very useful for the security of smart Enterprises and offices. In Today's era of data security methods, steganography, which dates back to the olden Greeks, has also discovered its way into the computer science arena and is successfully being used single-handedly or together with cryptography.

Our paper proposes a Four-Dimensional Data Security establishment to secure the smart Enterprise and office private data by implanting a Text file in the additional video signal. It is alarming to entrench information in a safe cover Speech in a protected and strong way. This system makes the files safe using impression Compression, Steganography, and Cryptography. Here we have used four methods in one attempt to make enterprise data more secure so that it would be concealed inside security's four layers. To achieve this target, initially, data is compressed, then encoded, and afterward hidden behind a video employing steganography. Recently, it has been suggested that 6G technology be used with the distributed RAN "Radio Access Network and the terahertz (THz) spectrum in order to enhance capacity, lower latency, and advance data safety features of enterprise and office security systems [2]. 6G will give users an ICT infrastructure that will surround them with a "huge artificial brain" that offers almost instant services, unlimited storage, and hugely improved cognitive abilities. The main goals of 6G systems are, (i) lowering the intake of

Table 1 6G comparison with 5G and 4G

Feature	6G	5G	4G
Data rate (per device); peak	1 Tbps	10 Gbps	1 Gbps
Integration with satellite	Fully	No	No
Latency (E2E); end to end	1 ms	10 ms	100 ms
Artificial intelligence	Fully	Partial	NO
Frequency (max)	10 THz	90 GHz	6 GHz

energy utilizing battery-free IoT appliances, (ii) a large number of connected appliances, (iii) global connectivity, (iv) enormously high data rates per device, (v) very low latency, (vi) connected intelligence with machine learning capabilities, and (vii) ultra-high reliable connectivity. Table 1 shows how the 6G communication system compares to 4G and 5G systems.

It is thought that the 6G system will allow 1000 times more wireless connections simultaneously as compared to the 5G system. Unlike 5G, that only has eMBB "Enhanced Mobile Broadband", 6G is supposed to have universal services or uMUB. The goal of 6G is to provide Gbps coverage everywhere, including new places like the sky (10,000 km) and the sea (20 nautical miles). Volume spectral efficiency must be superior in 6G than the more common area spectral efficiency. The 6G system will have batteries that last a very long time and will use advanced battery technology to harvest energy. In the context of 6G systems, the need for individual charging of mobile devices will be eliminated, and this technology will provide an unprecedented scope of data storage and safety. We have opted for 6G as our fourth and most important pillar of the data safety layer proposal.

2 Applications of Machine Learning in Smart Enterprises

Machine learning at present is being used in various applications in smart Enterprises, enabling intelligent and automated systems. A few of the vital appliances and applications of smart Enterprises utilizing machine learning are:

Energy Management: Machine learning algorithms are being used in analyzing energy consumption patterns and optimizing energy usage in smart Enterprises. By learning from historical data, these algorithms can make predictions and adjust energy settings to minimize waste and reduce costs.

Smart Thermostats: Machine learning algorithms learn residents' preferences for temperature settings, occupancy patterns, and weather conditions. This enables smart thermostats to automatically adjust temperature settings, leading to energy savings and enterpriseized comfort.

Predictive Maintenance: Machine learning algorithms analyze sensor data from various smart devices and detect anomalies or potential failures in smart Enterprise

appliances. This allows for predictive maintenance, where repairs or replacements are scheduled proactively, preventing device failures and reducing downtime.

Activity Recognition: Machine learning algorithms are capable of analyzing sensor data from smart devices, such as motion sensors or smart cameras, to recognize and classify human activities within the Enterprise. This can be used for security purposes, Enterprise automation, and even elderly care monitoring.

Voice Assistance: Machine learning is at the core of voice assistants like Amazon Alexa. Such systems utilize natural language processing and ML methodologies to connect to voice commands and respond or execute actions in smart Enterprises.

Security and Intrusion Detection: Machine learning algorithms can analyze sensor data, such as motion or door/window sensors, to detect anomalies and identify potential security breaches. This helps in enhancing Enterprise security and providing timely alerts to enterprise owners. But these systems need multilayered security system to avoid data and privacy intrusion.

Machine learning algorithms can learn from user preferences, historical data, and contextual information to provide Enterpriseized recommendations for entertainment options, lighting settings, or even shopping suggestions based on individual preferences and habits. Now it becomes paramount for us to prevent unauthorized access and data breaches in ML-assisted systems, justifying the possible threats like theft of identity and monetary scams.

3 System Architecture

Data traveling over the smart Enterprise and office security network is undoubtedly not secure. It is obligatory to protect it from viruses, hackers, cybercriminals, intruders, etc. There are diverse methods for ensuring digital data safety. We are banking on the following four-layered security system.

a. Cryptography b. Digital signature c. Steganography d. 6G.

3.1 Cryptography

To ensure that only the intended recipient can read and understand a message, cryptography encrypts it with a secret code. This would ensure that no unauthorized parties could access the information. The word “cryptography” comes from the Greek for “hidden writing” (crypto-) and the suffix “graphy” (writing). Mathematical principles and a system of rule-based calculations called algorithms are the basis of cryptography, a field dedicated to developing methods for safeguarding data by making it computationally challenging to decode transmitted messages. Data privacy, secure web browsing, and sensitive financial transactions like credit card and debit

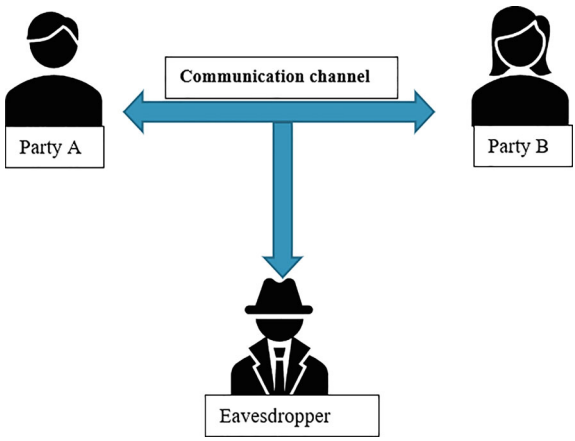


Fig. 1 Basic figure for insecure communication system

card purchases are protected by these algorithms' digital signing, verification, and essential generation functions. The two terms are synonyms: cryptography is used to encrypt data, while steganography is used to hide data. Most importantly, these techniques are used to protect sensitive data [3]. The difference between steganography and cryptography is that the former modifies undisclosed information so that spies cannot read it, while the latter hides the fact that the information exists from unauthorized parties [4, 5]. Data security is of paramount importance for both smart Enterprises and smart offices [6, 7]. An attacker with stolen data can use it for blackmail, defamation, or future attacks [8]. Encryption and decryption of secret writings are what cryptography is all about [9]. Figure 1 demonstrates communication recognized between two parties over an insecure medium that can be effortlessly eavesdropped on Fig. 1.

According to Gollmann [10], cryptography consists of different cryptographic techniques and frameworks such as encryption and decryption, check functions, integrity and digital signature frameworks, and so on. Encryption frameworks convert secret messages into formats that an unauthorized person can't read. In contrast, decryption frameworks are used to decode encrypted messages by a trusted party. A cryptographic hash function [11] is an integrity check function. It's a mathematical relationship used to identify small bits of data that can be used to identify larger digital objects accurately. It is computationally impossible to have an object with the same hash value as a preeminent object [12] because different things have different hash values. Message integrity can be verified after transmission thanks to hash functions. The other type of integrity check function [13] is message authentication codes (MACs), which are derived from the secret cryptographic key and the transmitted message. It's also used to verify a message's authenticity after it's been sent.

3.2 Digital Signature

The authenticity of a message, computer program, or digital document can be confirmed with the help of a mathematical technique called a digital signature. In essence, it's the digital equivalent of a handwritten signature or a stamping seal, but with much higher levels of protection. A digital signature stops people from changing or pretending to be someone else in digital communications. Digital signatures can prove where an electronic document, transaction, or digital message came from, who made it, and it is status. Signers can also use them to say they understand what they are signing. Asymmetric cryptography, which uses a public key, is the basis for digital signatures. A pair of keys is generated using a general key algorithm, such as RSA (Rivest-Shamir-Adleman). There is a mathematical connection between these two keys, a private one and a public one. Public key cryptography uses two separate cryptographic keys to validate a digital signature. All information tied to the digital signature is encrypted with the private key of the creator. This data is encrypted and may only be decrypted utilizing the signer's public key. It indicates a problem with the document or the signature if the recipient cannot translate it using the signer's public key. This is how you can be sure that a digital signature is authentic [14].

3.3 Steganography

Steganography is the science of sending messages in a manner that conceals their transmission. Therefore, a steganographic system covers information in something unremarkable so that an eavesdropper will not notice. We can, for instance, embed text within an image or audio file. [15–17] Steganography is the process of concealing a secret message within a cover message in such a way that it cannot be discovered. The hidden message could consist of an image, plain text, cypher text, or anything else that can be represented by bits [18]. Occasionally, the embedding process is controlled by a secret key, or stego-key, which must be known before the hidden message can be discovered and extracted. A "stego-object" is a message that is concealed within another. Before concealing a secret message beneath a cover, the sender must alter the secret message. Then, they must modify certain cover aspects to create the stego-object [19, 20]. Now, the stego-object is transmitted to the projected receiver via communication. After obtaining the information, the procedure is reversed to uncover the hidden information. If the process employs a secret key, both the sender and the recipient must possess it before the stego-object can be transmitted [21]. Figure 2 depicts a general steganographic model.

The other two methods related to steganography are watermarking and fingerprinting, but they are not the same [22]. Thus, fingerprinting, watermarking, and steganography differ in toughness, application, and smacking capacity [23]. Domain tools and simple systems are used in steganography, including inserting the least significant bit (LSB), manipulating noise and transforming images using algorithms

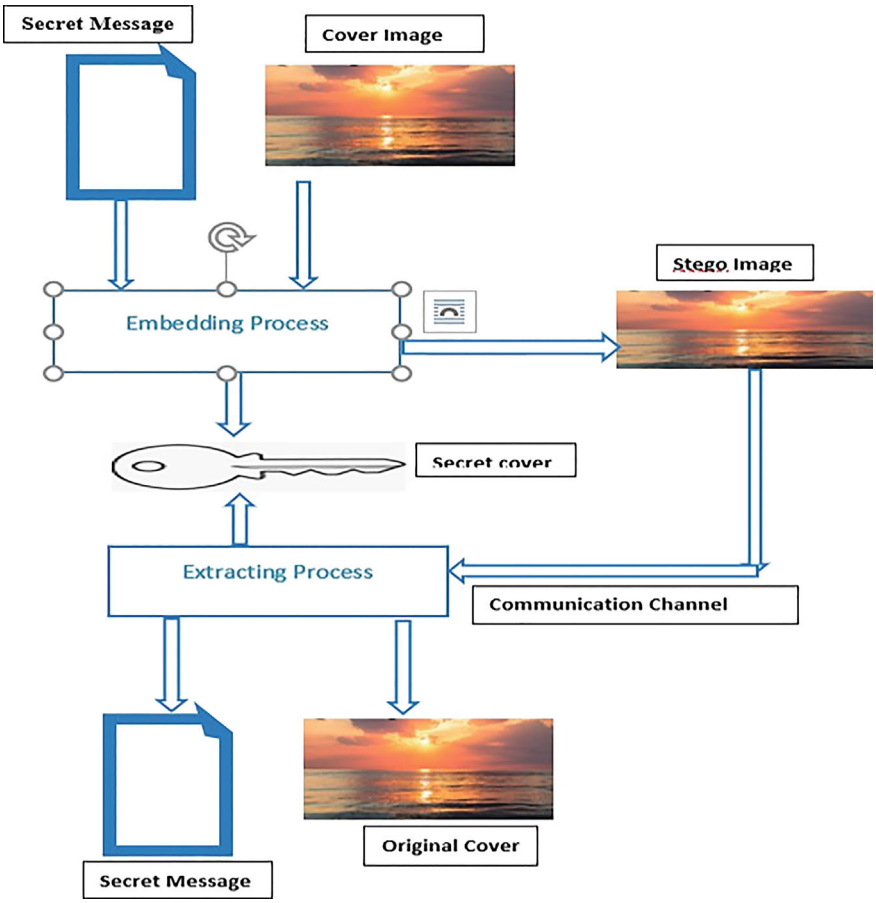


Fig. 2 Basic picture of a steganographic model

like wavelet and discrete cosine transforms. However, some methods share the image and domain tools; for example, patchwork, pattern block encoding, masking and spread spectrum methods.

3.4 6G

Rapidly emerging AI technology, quickly increasing network traffic and new intelligent applications (like self-driving cars, virtual reality, etc.) all need a novel network form that is faster, more reliable, and more flexible. Researchers in industry and universities are now focusing on sixth-generation (6G) communication networks. These are the sixth generation of standards for telecommunications. They are

currently being made for wireless technologies which support cellular data network. It is supposed to come after 5G and may probably be much faster. 6G network is expected to be even more different from each other than their predecessors.

Furthermore, they may be able to support use cases that extend beyond mobile devices, such as VR/AR, universal instant communications, IoT and pervasive intelligence. (ML technology, known as AI's vital most branch, will be used in various intelligent application scenarios in the 6G vision, leading to rich, heterogeneous connections and massive information storage and operations. New possibilities, including many for protected privacy and data security, will arise when ML and 6G are combined. An advanced and highly secure ML structure amalgamated with 6G technology will revolutionize the safety and security of the smart Enterprise and smart corporate world. Hence, we have kept this concept as 4th pillar of our proposed safety and security layered system.

4 Literature Review

Using steganography allows two or more people to have a conversation without being overheard. This is done by hiding a top-secret message on a media cover. Data (called "cover data" or "host data") can be hidden using a variety of steganographic techniques, including audio, text, image, and video steganography. First-dimensional (1D) audio steganography was developed by [24].

This author offers an adaptive Huffman code mapping (AHCM) [25] framework to provide a more secure payload. The paper devised a distortion-limited suppressible code space, that employs entropy codes of equal length to embed data to circumvent the audio codec's frame- offset effect. To further obscure its source and make it more challenging to locate statistically, a stego key has been utilized to generate the Huffman code mapping of every frame in real- time. They then consider combining the intra-frame psychoacoustic model (PAM) with the inter-frame perceptual distortion to achieve the lowest possible total distortion. Finally, syndrome-trellis codes are used for adaptive embedding after being distorted with a function based on the PAM.

Yet another LSB-based audio steganography technique is presented in Chowdhury et al. [26]. They developed a formidable encryption technique by fusing cryptography and steganography. Double encryption was used in this work. The first layer of security involved encrypting the text message with a pattern-matching algorithm that considered where in the file it was stored. The cover file's positional value has been written using the usual LSB method at the second level. If you use a double-encryption technique like this, your data will be safe and secure. Finally, the experimental and comparative results showed that the proposed algorithm is very effective at encrypting text and limiting the size of text it can hold.

The LSB technique is a fantastic way to conceal a message within an illustration or photograph. The image's appearance is not altered in any way, but it still accomplishes its intended purpose [25]. It replaces the least significant bits (LSBs) of specific host

image pixels with hidden message bits [26]. In [27, 28], Muhammad et al. used a stego key to control adaptive LSB replacement through many levels of encryption using a stego key. The secret information and the stego key are encrypted using multi-level encryption and two-level encryption algorithms before embedding. The outcomes demonstrate how the proposed plan improves aesthetics and increases security.

LSB insertion, masking and filtering, and algorithmic transformation were all discussed by Neil et al. [29] as three common methods for concealing digital communications. In their paper, Kevin and his colleagues examined seven different approaches to image concealment [30]. These procedures were variously referred to as "Stego1bit," "Stego2bits," "Stego3bits," "Stego4bits," "StegoPRNG," "Stego Colour Cycle," and "StegoFridrich." For example, in [31], Johnson et al. discovered, analyzed, and attacked various methods of concealing digital data in images, including transform domain methods, substitution systems, statistical steganography, concealment in two-color images, distortion, and cover generation. At long last, Johnson et al. provided an analogous solution. For their adaptive steganographic scheme [32], Wu et al. relied on a pixel-value differencing technique. In this sense, the degree to which each pixel can conceal information may vary. They opted for edge areas or pixels rather than smooth areas to store more secret information because of the greater degree to which an edge area can be changed compared to a smooth area.

Watermarking, steganography, fingerprinting, and signature-based image authentication are all methods described by Lu [33] for protecting digital images and sounds. A new method of 8-bit image embedding was developed by Alwan et al. [34]. The proposed approach utilized edge pixels to combine three images and text into a single picture. Juneja et al. [35] offer several methods, only one of which, LSB insertion, was implemented. The author provides a comprehensive analysis of recent developments in Enterprise Security Systems. It highlights the technical methods, microcontrollers, and sensors enhancing required security systems. The analysis covers networking technologies, user interface, computational methodologies, security systems, and various services like energy management and appliance control. Also, the increased popularity of multiple approach-based and microcontroller-based Enterprise Security Systems, with IoT-based systems emerging as dominant competitors has been depitced. Smartphone application-based interfaces are predicted to play a crucial role in the future due to their accessibility and functionality. Efforts to enhance network and physical security in Homrs are emphasized, indicating a continued focus on improving security measures [36–40].

5 Code Design and Implementation

Our proposed layered approach works on the theme demonstrated in Fig. 3 and acts as a base for the code design.

In this approach, each colored circle signifies the operation performed. The data or value inside the ellipse represents the input data, and the outside suggests the

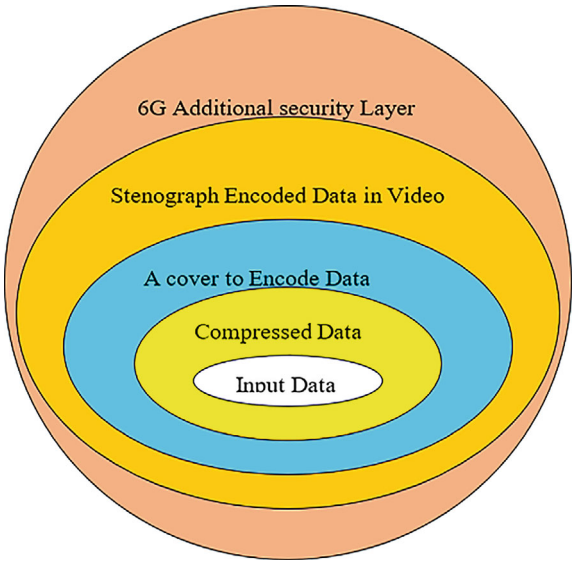


Fig. 3 Layered approach for code design

Output Data. The outermost layer illustrates the role and application of 6S, which has been proposed in this system to make this entire security ecosystem super secure and safe. Step-by-step technique implantation and details of our proposed security system are as under:

5.1 Hiding Technique Implementation

- Read the source file, change each byte to binary equivalent, and save it to array bit.
- Read the carrier file and step size to be calculated.
- In accordance to step size, change the last bit of a byte in the carrier file conforming to the values in the array bit [].
- Results are to be back to the carrier file.
- As it is just a one-bit change in a byte, we cannot see a considerable difference.

5.2 Retrieving Technique

- Retrieving is just the opposite of the hiding process.
- Read stegodata and recover step size.
- According to the step, size retrieves the hidden file and data size.
- Extract and the hidden file to be saved under the name input by the user.

5.3 *Methodology and Implementation*

The Fig. 4 demonstrates the process involved in splitting and reading the video file, followed by compression using GZIP, creation of an input stream with a specified buffer size, applying symmetric key encryption, and finally converting the data back to the video format (Fig. 5).

Initially, the video has to be split. After that, it has to be read, and after reading the video file and splitting it into images, compression is used to compress the data. GZIP compression is to be used, which has been explained below.

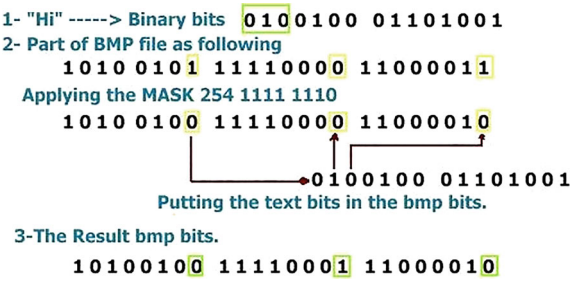


Fig. 4 Hiding technique implementation

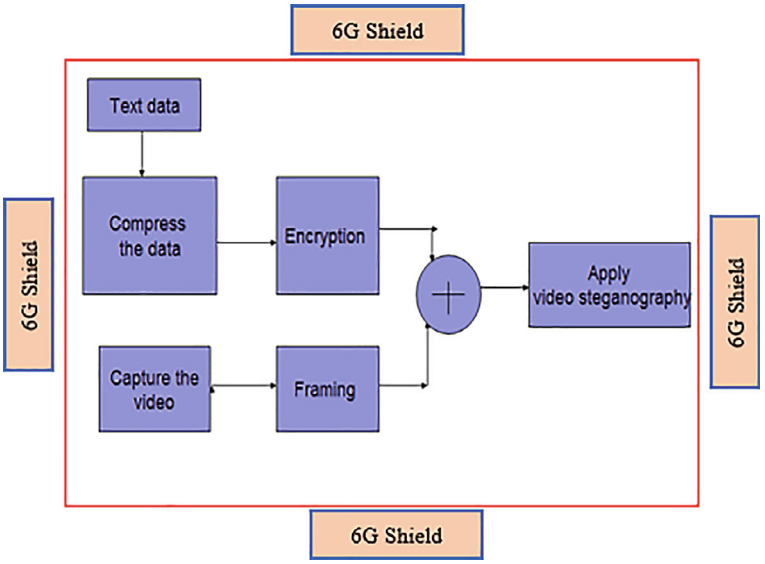


Fig. 5 Methodology of video processing

GZIP Input Stream

A new input stream is to be created with the specified buffer size. The next step would be to apply encryption. Symmetric key cryptography is to be used. Afterward, it has to be converted back to the video.

5.4 Algorithm Steps

Step 1: Select the video that is covered in which we want to hide our data.
Step 2: Select the text or text file which we want to hide.
Step 3: Then compress the image.
Step 4: Apply cryptography or encrypt the file.
Step 5: Then hide the data in the video using the steganography algorithm.
Step 6: Save the result as a new file.
Step 7: Now, to extract data select the file in which we save our final result, then extract data using the extraction algorithm.

5.5 Steganography Algorithm for Hiding Data

Steganography (image, data)

for (y = 0; y < btmap.Height; y + +)//Read Image upto down and left to right

for (x = 0; x < btmap.Width && index < array.Length; x ++)

{

a1 = Get Pixel (x,y);

a2 = a1 + 1;

a3 = a1 - 1;//Get + 1, and − 1 pixel values

Data = Get Bit (data (index))
bit1 = Get Bit(a1,1)
bit2 = Get Bit(a1,2)
bit3 = Get Bit(a1,3)
bit11 = Get Bit(a1,1)
bit12 = Get Bit(a1,2)
bit13 = Get Bit(a1,3)
bit11 = Get Bit(a1,1)
bit12 = Get Bit(a1,2)
bit13 = Get Bit(a1,3)
If (data = 1)

Then
Store according to increment array
Else
Store according to decrement array

TO STORE ZERO DATA

If career pixel bits are 0 0 0
0 0 1
1 0 0
Then store data as it is without any change.
If the career pixel bits are other than this then either increment or decrement 1 as required to get pixel bits in this form.

TO STORE 1 DATA

If career pixel bits are 1 1 1
0 1 1
1 1 1
Then store data as it is without any change.
If the career pixel bits are other than this then either increment or decrement 1 as required to get pixel bits in this form.

TO RETRIEVE DATA

```
for (y = 0; y < btmap.Height; y++).
{
        for (x = 0; x < btmap.Width; x++).
{
    data = get pixel (x,y)
    bit 1 = get bit(data,1)
    bit 2 = get bit(data,2)
    bit 3 = get bit(data,3)
if (bit1 = 0 & bit2 = 0 & bit3 = 0)
{
    Retrieve = 0;
}
        Else if(bit1=0; & bit2=1 & bit3=0)
Then
{
        retrieve=1;
   }}
```

Additional shielding of 6G: Looking at the excellent and bright prospects and features of the upcoming 6G, we have proposed an extra layer of 6G on our secured layered system to provide supplementary protection to the smart Enterprise and smart corporate office networks. Sixth- generation (6G) networks are planned to integrate terrestrial wireless and satellite communications to create a fully connected world.

The design concept of 6G network is to power AI to endorse network services' intelligent and agile development. Services that fall under the category of "intelligent" will inevitably necessitate the storage, computation, and analysis of massive amounts of data. Even though such information could be susceptible towards tamper or bugging by hackers, 6G technology will be advanced enough to deal with these risks. In 6G network, Blockchain-based data security schemes for AI applications can address cyber threats and provide top-notch security to smart Enterprises and smart office systems.

6 Results and Discussion

A series of crucial tests and validations to ensure the robustness and reliability of the implemented system is conducted. These tests specifically focused on three key aspects: compatibility with AVI and WMV video file formats, cryptographic implementation, and code integrity.

Compatibility with AVI and WMV Files: Our system was subjected to rigorous testing to assess its compatibility with AVI and WMV video file formats. This was essential to ensure that the system seamlessly integrates with and functions effectively across a range of file types. The results of these tests indicated that the system successfully accommodates both AVI and WMV formats without any issues. This compatibility is a critical feature for a versatile video processing system.

Cryptography Implementation: Security is paramount in many applications, and to address this concern, we implemented cryptographic methods within our system [41]. The chosen cryptographic techniques, including the one previously mentioned, were thoroughly tested. This step was taken to verify that encryption and decryption processes are executed correctly and securely. Moreover, we extended our evaluation to include alternative cryptographic methods, such as DSA (Digital Signature Algorithm), to determine their effectiveness and suitability for our specific use case. The results of these tests confirmed that our cryptographic implementations meet the required security standards.

Code Testing: To ensure the reliability and robustness of our system, it subjected the code to extensive testing. This encompassed various test cases to identify and rectify any bugs, errors, or unexpected behaviors that may arise under different scenarios or input variations. The objective was to guarantee that the system performs consistently and reliably in real-world scenarios.

The theoretical, managerial, and practical implications: The process of securing enterprise Data with 4D Security Framework which is involving cryptography, digital signatures, steganography, and 6G networks is multi-faceted. Cryptography offers a theoretical basis for secure data exchange, using mathematical algorithms to encrypt data. Managers must prioritize its implementation across communication channels to prevent unauthorized access and data breaches, while practically, encryption tools

should be used in systems such as cloud storage, IoT, and enterprise databases. Digital signatures [42] theoretically authenticate and ensure the integrity of documents. For managers, this technology provides legal backing and ensures that transactions and communications are secure, while practically, integrating digital signatures into workflows helps prevent fraud and tampering, making it essential for contracts and sensitive document handling. Steganography, though less common, offers a theoretical method of concealing the presence of communication, ensuring confidentiality [43]. Managers can use this method to further secure sensitive information, especially in industries where covert communication is necessary, while practically it can be applied by embedding sensitive data within files like images and audio. Finally, 6G networks and their integration with AI offer a theoretical leap in security capabilities, providing ultra-reliable, low-latency connections [44, 45]. For managers, investing in 6G infrastructure will be crucial to handle future AI-driven security needs, while practically, these networks will support advanced encryption, threat detection, and secure IoT environments, revolutionizing data security in corporate settings.

In conclusion, these comprehensive tests and verifications play a pivotal role in our research work. They serve the purpose of instilling confidence in the functionality, compatibility, and security of our implemented system. We have demonstrated its capability to handle diverse video file formats, employ secure cryptographic methods, and maintain reliable performance across a spectrum of use cases. These findings substantiate the robustness and reliability of our system and contribute to the overall credibility of our research. Furthermore, the results of these tests are summarized in Table 2, providing a clear overview of the system's performance across different file sizes and formats. The Fig. 6 present the video steganography file selection on the computer screen.

Table 2 Results for Avi and Wmv files

Text file Size	File size	Format	Result
400 Bytes	100 kb	Avi	Success no file size change
1 kb	100 kb	Avi	Success no file size change
1 MB	100 kb	Avi	Success with file size change
1 MB	3 MB	Wmv	Success no file size change
500 kb	10 MB	Wmv	Success no file size change

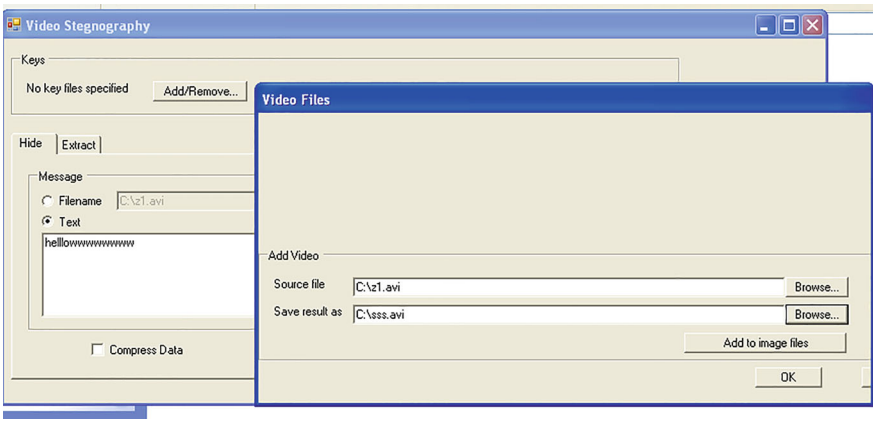


Fig. 6 The video steganography file selection

7 Conclusion

Steganography plays a significant role in data security by complementing cryptography rather than replacing it. By using video steganography to hide messages, the chances of detection are reduced. Even if the message is encrypted or discovered, it remains uncrackable due to an additional layer of protection, and compression adds further security. Video steganography systems ensure that hidden data inside innocent files cannot be discovered, while intended recipients can access the concealed information. Furthermore, we propose incorporating the fourth layer of 6G technology to maintain the scope of futuristic development. By integrating an advanced and highly secure ML structure with 6G technology, smart Enterprise and corporate systems can achieve exceptional safety and security. To keep the scope of development alive, the incorporation of the fourth layer of 6G technology into the system will definitely act as a boon. This integration of an advanced and highly secure machine learning structure with 6G is anticipated to bring superior safety and security to smart Enterprise and smart corporate systems. Therefore, the future scope entails advancing the synergy of video steganography and cryptography to enhance data security. This will involve exploring innovative techniques to further reduce detection risks and bolster uncrackable message protection. Additionally, integrating the fourth layer of 6G technology with advanced machine learning structures promises to elevate safety and security for smart Enterprise and corporate systems. By refining these integrated approaches, pushing the boundaries of data protection in steganography and smart systems, while continually adapting to emerging technological landscapes.

References

1. Almuhammadi, S., Al-Shaaby, A.: A survey on recent approaches combining cryptography and steganography. Comput. Sci. Inform. Technol. (2017)
2. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: Digital image steganography: survey and analysis of current methods. Signal Process. **90**, 727–752 (2010)
3. Moody, G.D., Siponen, M., Pahnila, S.: Toward a unified model of information security policy compliance. MIS Q. **42**, 1 (2018)
4. Acar, A., Aksu, H., Uluagac, A.S., Conti, M.: A survey on homomorphic encryption schemes: theory and implementation. ACM Comput. Surv. **51**, 79 (2018)
5. Hashim, M., Rahim, M., Shafry, M., Alwan, A.A.: A review and open issues of multifarious image steganography techniques in spatial domain. J. Theor. Appl. Inform. Technol. **96** (2018)
6. Diesburg, S.M., Wang, A.I.A.: A survey of confidential data storage and deletion methods. ACM Comput. Surv. **43**, 2 (2010)
7. Matthews, G.J., Harel, O.: Data confidentiality: a review of methods for statistical disclosure limitation and methods for assessing privacy. Stat. Surv. **5**, 1–29 (2011)
8. Seth, D., Ramanathan, L., Pandey, A.: Security enhancement: combining cryptography and steganography. Int. J. Comput. Appl. 3–6 (2010)
9. Joseph A and Sundaram V 2011 Cryptography and steganography–A survey.
10. Gollmann, D.: Computer security. Wiley Interdisc. Rev. Comput. Stat. **2**, 544–554 (2010)
11. Saint-Andre, P., Wild, M., Smith, K., Markmann, T.: Use of Cryptographic Hash Functions in XMPP (2018)
12. Jallad, K., Katz, J., Schneier, B.: Implementation of chosen-ciphertext attacks against PGP and GnuPG. In: International Conference on Information Security, pp. 90–101. Springer, Berlin, Heidelberg (2002)
13. Yadav, H.: Secure multiparty communication with verifiable outsourced decryption for mobile cloud computing (2018)
14. Lyubashevsky, V., Micciancio, D.: Asymptotically efficient lattice-based digital signatures. J. Cryptol. 1–24 (2018)
15. Hashim, M.M., Rahim, M.S.M., Johi, F.A., Taha, M.S., Al-Wan, A.A., Sjarif, N.N.A.: An extensive analysis and conduct comparative based on statistical attach of LSB substitution and LSB matching. Int. J. Eng. Technol. **7**, 4008–4023 (2018)
16. Douglas, M., Bailey, K., Leeney, M., Curran, K.: An overview of steganography techniques applied to the protection of biometric data. Multim. Tools Appl. **77**, 17333–17373 (2018)
17. Mishra, S., Yadav, V.K., Trivedi, M.C., Shrimali, T.: Audio Steganography Techniques: A Survey In Advances in Computer and Computational Sciences, pp. 581–589. Springer, Singapore (2018)
18. Morkel, T.: Image steganography applications for secure communication. Doctoral dissertation, University of Pretoria (2012)
19. Petitcolas, F.A., Anderson, R.J., Kuhn, M.G.: Information hiding–a survey. In: Proceedings of the IEEE 87 1062–1078. 2nd International Conference on Sustainable Engineering Techniques (ICSET 2019) IOP Conference Series: Materials Science and Engineering, vol. 518, p. 052003. IOP Publishing (1999) https://doi.org/10.1088/1757-899X/518/5/052003
20. Cao, Y., Wang, Y., Zhao, X., Zhu, M., Xu, Z.: Cover block decoupling for content adaptive H.264 .steganography In: Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security, pp. 23–30. ACM (2018)
21. Al-Husainy, M.A.F., Uliyan, D.M.: Image steganography technique based on extracted chains from the secret key. J. Eng. Appl. Sci. **13**, 4235–4244 (2018)
22. Petitcolas, F.A., Anderson, R.J., Kuhn, M.G.: Information hiding-a survey. Proc. IEEE **87**, 1062–1078 (1999)
23. Hussain, M., Wahab, A.W.A., Idris, Y.I.B., Ho, A.T.S., Jung, K.-H.: Image steganography in spatial domain: a survey. Signal Process. Image Commun. **65**, 46–66 (2018)
24. Sedighi, V., Cogranne, R., Fridrich, J.: Content-adaptive steganography by minimizing statistical detectability

25. Yi, X., Yang, K., Zhao, X., Wang, Y., Yu, H.: AHCM adaptive Huffman code mapping for audio steganography based on psychoacoustic model. IEEE Trans. Inform. Forens. Secur. **14**(8), 2217–2231 (2019)
26. Chowdhury, R., Bhattacharyya, D., Bandyopadhyay, S.K., Kim, T.: A view on LSB based audio steganography. Int. J. Secur. Appl. **10**(2), 51–62 (2016)
27. Muhammad, K., Ahmad, J., Rehman, N.U., Jan, Z., Sajjad, M.: CISSKA-LSB: color image steganography using stego key-directed adaptive LSB substitution method. Multim. Tools Appl. **76**(6), 8597–8626 (2017)
28. Muhammad, K., Ahmad, J., Rho, S., BaikImage, S.W.: Steganography for authenticity of visual contents in social networks. Multim. Tools Appl. **76**(18), 18985–19004 (2017)
29. Johnson, N.F., Jajodia, S.: Exploring steganography, seeing the unseen. IEEE Comput. Mag. **31**(2), 26–34 (1998)
30. Curran, K., Bailey, K.: An evaluation of image based steganography methods. Int. J. Dig. Evid. **2**(2), 1–40 (2003)
31. Johnson, N.F., Duric, Z., Jajodia, S.: Information hiding, and watermarking—attacks & countermeasures. J. Electron. Imag. **10**(3), 825–826 (2000)
32. Wu, N., Hwang, M.: Data hiding: current status and key issues. Int. J. Netw. Secur. **4**(1), 1–9 (2007)
33. Lu, C.S.: Steganography and digital water marking techniques for protection of intellectual property. In: Multimedia Security, pp. 75–157. Idea Group Publishing, Singapore (2005).
34. Alwan, R.H., Kadhim, F.J., Taani, A.T.A.: Data embedding based on better use of bits in image pixels. Int. J. Signal Process. **2**(1), 104–107 (2005)
35. Juneja, M., Sandhu, P.: An improved LSB based steganography technique for RGB colorimages. Int. J. Comput. Commun. Eng. 513–517 (2013)
36. Chakraborty, A., Islam, M., Shahriyar, F., Islam, S., Zaman, H.U., Hasan, M.: Smart enterprise system: a comprehensive review. J. Electr. Comput. Eng. **2023** (2023) (Article ID 7616683, 30 pages). https://doi.org/10.1155/2023/7616683
37. Almutairi, M., Gabralla, L.A., Abubakar, S., Chiroma, H.: Detecting elderly behaviors based on deep learning for healthcare: recent advances, methods, real-world applications and challenges. IEEE Access **10**, 69802–69821 (2022)
38. Wu, Q., Chen, X., Zhou, Z., Zhang, J.: FedHome: cloud-edge based personalized federated learning for in-home health monitoring. IEEE Trans. Mob. Comput. **21**(8), 2818–2832 (2022)
39. Suman, S., Etemad, A., Rivest, F.: Potential impacts of smart enterprises on human behavior: a reinforcement learning approach. IEEE Trans. Artif. Intell. **3**(4), 567–580 (2022)
40. Hasan, M., Hussain, M.S., Hossain, M., Hasan, M., Zaman, H.U., Islam, S.: A high-speed and scalable XOR-XNOR based hybrid full adder design. Comput. Electr. Eng. **92** (2021)
41. Rajalakshmi, S., Duraisamy, P.: A review on lightweight cryptographic algorithms in internet of things. In: 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1448–1451. IEEE (2023)
42. https://emudhra.com/blog/the-public-key-infrastructure-approach-to-security-a-deep-dive-into-pki. Accessed 16 May 2024
43. Ahmad Bamanga, M., Kamalu Babando, A., Ahmed Shehu, M.: Recent Advances in Steganography. IntechOpen (2024). https://doi.org/10.5772/intechopen.1004521
44. Chataut, R., Nankya, M., Akl, R.: 6G networks and the AI revolution—exploring technologies, applications, and emerging challenges. Sensors **24**(6), 1888 (2024)
45. Rifa-Pous, H., Garcia-Font, V., Nunez-Gomez, C., Salas, J.: Security, trust and privacy challenges in AI-driven 6G networks (2024). arXiv preprint arXiv:2409.10337

Combined Effect of Charge Rate and Temperature on Battery Health of Electric Vehicles

Namrata Mohanty, Neeraj Kumar Goyal, and V. N. Achutha Naikan

Abstract Electric vehicles (EVs) operate under diverse environmental conditions, which significantly influence the performance and lifespan of their batteries. This study investigates the combined impact of charging rates and ambient temperatures on the degradation of lithium iron phosphate (LFP) batteries commonly used in EVs. A novel hybrid machine learning (ML) approach is employed using an extreme gradient boosting random forest (XG-RF) model to estimate the state of health (SOH) of battery, factoring in key parameters such as charge cycles, voltage, and surface temperature. The analysis evaluates battery cycle life across varying charge rates (C/20, 1C, 2C, and 3C) and three different temperatures (5 °C, 25 °C, and 35 °C). The results highlight the superior cycle life of LFP batteries, especially at moderate temperatures, achieving up to 5293 cycles at a 25 °C temperature and a C/20 charge rate. These insights provide critical guidance for enhancing battery performance in EVs and contribute to advancements in battery health monitoring and optimization strategies.

Keywords State of health (SOH) · Lithium-ion batteries (LIBs) · Electric vehicles (EVs) · Battery degradation · Machine learning (ML)

N. Mohanty (✉) · N. K. Goyal · V. N. A. Naikan
Subir Chowdhury School of Quality and Reliability, Indian Institute of Technology Kharagpur, Kharagpur, India
e-mail: namratamohanty@kgpian.iitkgp.ac.in

N. K. Goyal
e-mail: ngoyal@hijli.iitkgp.ac.in

V. N. A. Naikan
e-mail: naikan@hijli.iitkgp.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_29

1 Introduction

The rapid advancement of electric vehicles (EVs) has demanded significant improvements in battery technology, as the efficiency and reliability of lithium-ion batteries (LIBs) are vital to the widespread adoption of EVs [1]. Fast charging technology is designed by applying high current levels to LIBs in EVs, thereby reducing charging time [2]. However, the impact of fast charging on battery degradation remains a critical concern, as it accelerates capacity fade and reduces lifespan by increasing stress on internal components in EVs [3]. Additionally, fast charging also leads to the risk of thermal runaway by generating excessive heat and increasing internal pressure, thereby increasing safety concerns among users [4].

In real-time applications, EVs operate under diverse environmental conditions and charging scenarios, leading to significant fluctuations in both charge rates and ambient temperatures simultaneously [5]. Prior research has revealed several operational characteristics, including charge rate, temperature, and cycling conditions, that affect battery health. For instance, Koo et al. [6] demonstrated that high charge rates lead to lithium plating, which diminishes battery capacity and cycle life. Similarly, Capron et al. [7] highlighted the role of temperature in battery degradation, noting that elevated temperatures accelerate aging and pose safety risks. While these studies provide valuable insights, there remains a significant gap in understanding the combined effects of charge rate and temperature on battery health. Existing research often focuses on one parameter at a time, limiting insights into how these factors interact under real-world conditions, where simultaneous variations in both charge rate and temperature can significantly influence battery degradation, reliability, and safety [8]. Furthermore, while studies on specific battery chemistries like lithium ion phosphate (LFP) [9], nickel manganese cobalt (NMC) [10], and nickel cobalt aluminium (NCA) [11] have documented their distinct degradation behaviors, the interplay between charge rate and temperature in real-world scenarios has yet to be thoroughly explored.

State of health (SOH) assesses battery degradation by indicating the overall status of a battery in relation to its optimal or original state [12]. Existing SOH estimation methods are classified as direct [13], model-based [14] and data-driven machine learning (ML) models [15]. While data-driven approaches may struggle with capturing complex relationships within the data, and model-based approaches may face challenges in accurately representing the non-linear dynamics of LIBs, hybrid models in this case are a better approach of combination of multiple algorithms [16]. Therefore, this study seeks to address this gap by analyzing the combined impact of charge rate and temperature on LIB degradation using a hybrid model, focusing on LFP battery chemistry. The research investigates battery performance across four charge rates (C/20, 1C, 2C, and 3C) and three temperatures (5 °C, 25 °C, and 35 °C). This approach allows for a more nuanced understanding of how various factors jointly affect battery health, which is crucial for enhancing charging protocols and ensuring the reliability and safety of LIBs in real-world applications.

The subsequent sections of this work are organised as follows: Sect. 2 introduces the model framework for SOH estimation. Section 3 describes the joint impact of parameters on battery health, succeeded by results and observations in Sect. 4. The conclusions are summarised in Sect. 5, accompanied by relevant references.

2 Model Structure

This paper presents a dataset of INR21700-M50T LIB cells tested over 23 months at Stanford Energy Control Laboratory, following the Urban Dynamometer Driving Schedule (UDDS) at different charge rates and temperature ranges [17]. This study employs an advanced hybrid ML framework using an extreme gradient boosting random forest (XG-RF) model to predict the SOH of LFP battery under different charging rates and ambient temperatures. The XG-RF model is selected for its capability to handle complex, non-linear relationships between input variables and battery degradation behavior. The key input features for the model include charge cycles, charge rates (C/20, 1C, 2C, and 3C), ambient temperature (5 °C, 25 °C, and 35 °C), and voltage-current profiles across the battery's operational life. These features are essential for capturing the intricate dynamics of battery health under varied conditions. The model is trained on a dataset generated from extensive cycling tests of LFP batteries, where each decision tree within the XG-RF ensemble is iteratively improved, reducing errors and enhancing predictive accuracy. This model provides insights into how charge rates and temperature affect LFP battery life, offering valuable guidance for optimizing charging protocols and improving battery management system (BMS) strategies for EVs.

The battery's end of life (EOL) is reached when SOH falls to 80%, indicating its incapacity to meet the operational requirements of EVs due to reduced capacity and safety concerns. Beyond this threshold, battery performance, especially during fast charging, deteriorates significantly. This hybrid data-driven ML strategy aims to combine the intricate pattern identification abilities of diverse algorithms with the ensemble learning advantages of XGBoost, facilitating a more sophisticated comprehension of battery deterioration.

3 Parameters Effect on Battery Health

This section examines the performance of LFP battery at different charge rates (C/20, 1C, 2C, and 3C) at three distinct temperatures (5 °C, 25 °C, and 35 °C). An in-depth analysis of the interaction among these variables is crucial for formulating effective solutions to reduce degradation and extend the lifespan of LIBs. The correlation between battery performance and the number of charge/discharge cycles under diverse settings can be depicted by SOH versus cycle curve.

Table 1 Results for obtained cycle life for LFP batteries

Charge rates	Cycle life (EOL of first life)		
	5 °C	25 °C	35 °C
C/20	1387	5293	1087
1C	823	4693	866
2C	565	1191	1052
3C	1062	678	283

The experimental results clearly demonstrate the significant impact of charge rates and temperature on the cycle life of LFP battery. As shown in Table 1, LFP batteries exhibit superior performance at lower temperatures and slower charge rates, achieving the longest cycle life under these conditions.

At a C/20 charge rate, the LFP battery achieves an impressive cycle life of 5293 cycles at 25 °C, which is the highest observed in this study. This indicates that moderate temperatures and slow charging significantly extend battery life, making this configuration ideal for applications requiring longevity, such as EVs with predictable usage patterns and longer operational lifespans. In contrast, at 35 °C and the same charge rate, the cycle life drops sharply to 1087 cycles, suggesting that higher temperatures accelerate battery degradation, even at slower charge rates.

For faster charging at 1C, the LFP battery still performs well at 25 °C, reaching 4693 cycles. However, at 5 °C and 35 °C, the cycle life is considerably shorter, with 823 cycles and 866 cycles, respectively. This indicates that while 1C charging is feasible, it becomes increasingly harmful to battery life as temperature deviates from the optimal 25 °C range.

At even higher charge rates of 2C and 3C, the performance of LFP batteries degrades significantly. At 2C, the cycle life at 25 °C is reduced to 1191 cycles, while at 3C, it plummets to 678 cycles, with even worse performance at 35 °C i.e., only 283 cycles at 3C (Fig. 1).

This highlights the detrimental effect of fast charging, especially at higher temperatures, where thermal management becomes critical to prevent excessive wear and tear on the battery.

Next section provides a comprehensive analysis of the results and observations for the battery across different charge rates and temperature conditions.

4 Results and Observations

LFP batteries demonstrate remarkable stability and reliability, particularly at lower temperatures and slower charge rates. However, higher charge rates and temperatures lead to increased degradation, emphasizing the importance of thermal management. Table 1 summarizes results obtained for LFP batteries at different charge rates and temperatures.

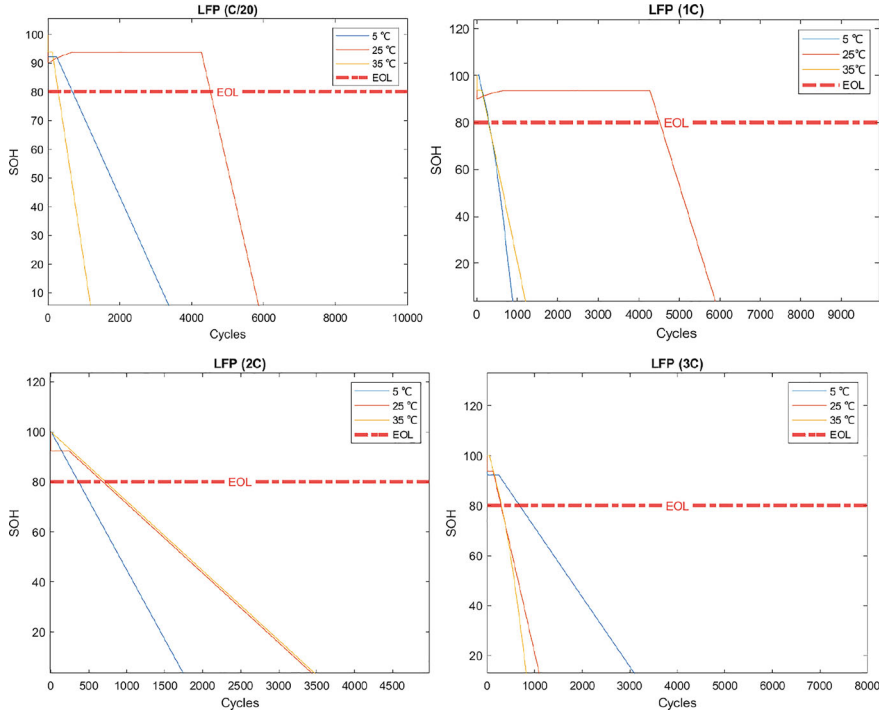


Fig. 1 SOH versus cycle curve for LFP batteries at different charge rates and temperatures

Here, the best performance is observed at 25 °C and C/20 with the longest cycle life, while fast charging (2C and 3C) and high temperatures (35 °C) result in the most rapid degradation. The results underscore the varied performance attributes of these the battery at different charge rates and temperatures. This highlights the necessity of optimizing thermal management and charging procedures for batteries to improve the overall performance and longevity of various battery chemistries in EVs.

Future studies should investigate the combined effects of charge rates and temperatures on the performance of various other battery chemistries, including nickel cobalt aluminum (NCA) and nickel manganese cobalt (NMC), to understand their unique degradation mechanisms. This comprehensive analysis could provide valuable insights for optimizing thermal management and charging protocols tailored to each battery type, enhancing overall EV efficiency and longevity.

5 Conclusion

This study examined the combined effects of charge rates and temperatures on the cycle life of lithium iron phosphate (LFP) batteries. The results reveal that LFP batteries offer exceptional stability and longevity, particularly at a C/20 charge rate and 25 °C, making them highly suitable for applications that demand long life cycles and safety. Their ability to maintain superior performance under moderate temperatures and lower charge rates underscores their advantage in EV applications where reliability is a priority. Future research should focus on optimizing charging protocols and further enhancing thermal management to maximize the efficiency and lifespan of LFP batteries in diverse operating conditions.

References

1. Mohanty, N., Goyal, N.K., Naikan, V.N.A.: Effect of training algorithms in accurate state of charge estimation of lithium-ion batteries using NARX model. Int. J. Heavy Veh. Syst. **30**(2), 232–254 (2023). https://doi.org/10.1504/IJHVS.2023.132335
2. Zhou, H., et al.: Effect of fast charging on degradation and safety characteristics of lithium-ion batteries with $LiFePO_4$ cathodes. Appl. Energy **377**(PA), 124465 (2025). https://doi.org/10.1016/j.apenergy.2024.124465
3. Deb, N., Singh, R., Brooks, R.R., Bai, K.: A review of extremely fast charging stations for electric vehicles. Energies **14**(22) (2021). https://doi.org/10.3390/en14227566
4. Li, Y., et al.: Battery eruption triggered by plated lithium on an anode during thermal runaway after fast charging. Energy **239** (2022). https://doi.org/10.1016/j.energy.2021.122097
5. Gao, Y., Jiang, J., Zhang, C., Zhang, W., Ma, Z., Jiang, Y.: Lithium-ion battery aging mechanisms and life model under different charging stresses. J. Power. Sour. **356**, 103–114 (2017). https://doi.org/10.1016/j.jpowsour.2017.04.084
6. Koo, B., Yi, J., Lee, D., Shin, C.B., Park, S.: Modeling the effect of fast charge scenario on the cycle life of a lithium-ion battery. J. Electrochem. Soc. **165**(16), A3674–A3683 (2018). https://doi.org/10.1149/2.0281816jes
7. Capron, O., Jaguemont, J., Gopalakrishnan, R., Van den Bossche, P., Omar, N., Van Mierlo, J.: Impact of the temperature in the evaluation of battery performances during long-term cycling-characterisation and modelling. Appl. Sci. **8**(8), 1–25 (2018). https://doi.org/10.3390/app8081364
8. Ardeshiri, R.R., Balagopal, B., Alsabbagh, A., Ma, C., Chow, M.Y.: Machine learning approaches in battery management systems: state of the art: remaining useful life and fault detection. In: Proceedings of 2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems, IESES 2020, no. September, pp. 61–66 (2020). https://doi.org/10.1109/IESES45645.2020.9210642
9. Preger, Y., et al.: Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. J. Electrochem. Soc. **167**(12), 120532 (2020). https://doi.org/10.1149/1945-7111/abae37
10. Tanim, T.R., et al.: A comprehensive understanding of the aging effects of extreme fast charging on high Ni NMC cathode. Adv. Energy Mater. **12**(22), 1–17 (2022). https://doi.org/10.1002/aenm.202103712
11. Bhoir, S.S., Brivio, C., Namor, E., Hutter, A.: Experimental comparison of fast-charging protocols for NMC and NCA Li-ion batteries. In: 2021 IEEE Vehicle Power and Propulsion Conference. VPPC 2021—Proceedings, pp. 1–7 (2021). https://doi.org/10.1109/VPPC53923.2021.9699124

12. Ezemobi, E., Silvagni, M., Mozaffari, A., Tonoli, A., Khajepour, A.: State of health estimation of lithium-ion batteries in electric vehicles under dynamic load conditions. Energies **15**(3) (2022). https://doi.org/10.3390/en15031234
13. Vignesh, S., et al.: State of health (SoH) estimation methods for second life lithium-ion battery—review and challenges. Appl. Energy **369**(May), 123542 (2024) https://doi.org/10.1016/j.apenergy.2024.123542
14. Bian, X., Wei, Z., He, J., Yan, F., Liu, L.: A novel model-based voltage construction method for robust state-of-health estimation of lithium-ion batteries. IEEE Trans. Ind. Electron. **68**(12), 12173–12184 (2021). https://doi.org/10.1109/TIE.2020.3044779
15. Oji, T., Zhou, Y., Ci, S., Kang, F., Chen, X., Liu, X.: Data-driven methods for battery soh estimation: survey and a critical analysis. IEEE Access **9**, 126903–126916 (2021). https://doi.org/10.1109/ACCESS.2021.3111927
16. Maleki, S., Ray, B., Hagh, M.T.: Hybrid framework for predicting and forecasting state of health of lithium-ion batteries in electric vehicles. Sustain. Energy, Grids Netw. 100603 (2022). https://doi.org/10.1016/j.segan.2022.100603
17. Pozzato, G., Allam, A., Onori, S.: Lithium-ion battery aging dataset based on electric vehicle real-driving profiles. Data Br. **41**, 107995 (2022). https://doi.org/10.1016/j.dib.2022.107995

Bi-LSTM-Based Model for Software Reliability Prediction in Multi-sprint Agile Based Software



Shikha Dwivedi and Neeraj Kumar Goyal

Abstract Agile software development is characterized by short iterative cycles known as sprints which aims to deliver high-quality software efficiently. Although multi-release reliability models exist, there is limited research on modeling and predicting reliability within a multi-sprint context. This paper presents a neural network-based model for predicting software reliability across multiple sprints. Using historical sprint data from similar projects, the model captures patterns to predict faults in ongoing sprints. The proposed approach outperforms traditional models, especially in early testing phases, by using previous sprint data. The model is trained and tested on the JIRA dataset, demonstrating superior performance compared to existing models, such as the Rawat and Saraf models, across various metrics, including MSE, Adjusted R-squared, and AIC. Results demonstrate the model's applicability and accuracy using real-world software failure data from sprint-based development projects.

Keywords Software reliability · Fault prediction · Agile · Multi-sprint · LSTM

1 Introduction

Agile software development has emerged as a dominant framework, emphasizing flexibility, iterative progress, and customer collaboration. Agile methodologies, such as Scrum, Kanban, and Extreme Programming (XP), prioritize shorter development cycles known as sprints, which typically last from one to four weeks [1]. These sprints allow teams to adapt quickly to changes, deliver incremental updates, and integrate customer feedback continuously. The agility of this approach contrasts

S. Dwivedi (✉) · N. K. Goyal
Subir Chowdhury School of Quality and Reliability, IIT Kharagpur, Kharagpur, India
e-mail: shikhadwivedi1969@iitkgp.ac.in

N. K. Goyal
e-mail: ngoyal@hijli.iitkgp.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_30

sharply with traditional, linear development models like the Waterfall model, which lack the ability to respond dynamically to evolving requirements [2].

In agile projects, software is continuously tested, and feedback from each sprint is used to guide the next development phase. This iterative cycle is beneficial for quickly identifying and rectifying faults, but it also poses challenges for traditional software reliability modeling. Conventional statistical Software Reliability Growth Models (SRGMs) [3–8] were designed for projects with clear boundaries between development and testing phases, typically assuming static environments. SRGMs often rely on the assumption of independence between releases and do not effectively incorporate real-time fault detection and correction data characteristic of agile development.

Karunanithi et al. [9] utilized neural networks to model the cumulative number of failures, developing software reliability models based on feed-forward neural networks. In 2006, Hu et al. [10] enhanced early reliability prediction using deep neural networks by leveraging failure data from previous projects/releases. Bisi and Goyal [11] applied a particle swarm optimization-based ANN model for fault prediction. In 2018, Wang et al. [12] introduced a deep encoder-decoder model to estimate the number of software faults. Samal et al. [13] incorporated DNN using deeper layers to predict software faults. However, the proposed models in above literature have focused on the traditional development approaches where effect of multi sprint modeling is neglected. Also, there are some research that has been caried out on multi-release software reliability modeling using parametric approaches which rely on unrealistic assumptions for modeling. These models considered that each release of the software follow same failure rate and used same failure distribution for modeling them. Although in reality the scenario is totally different, as failure pattern varies from one sprint to another due to addition of new features.

Given these limitations, there is a critical need for models that can adapt to the iterative and dynamic nature across multiple sprints of agile software development. The unpredictability and high variance in the number of faults encountered across sprints make conventional SRGMs less effective in predicting software reliability. This is because SRGMs generally assume a homogeneous failure rate across development phases, which does not hold true in the context of agile, where failure rates vary significantly depending on sprint objectives, developer experience, and testing intensity.

To address these challenges, this paper proposes a LSTM-based model for predicting software reliability in multi-sprint scenarios. LSTM offers a distinct advantage over traditional SRGMs due to their ability to learn from historical patterns and adapt to non-linear relationships within data. Unlike SRGMs, which require predefined distributions and assumptions about fault occurrences, neural networks can capture complex dependencies without these restrictive assumptions.

By utilizing Bidirectional LSTM, the model can dynamically update its predictions as new data from each sprint becomes available, allowing it to refine its accuracy over time. This adaptive capability is essential in agile environments, where failure patterns can change based on evolving project requirements, sprint goals, and team dynamics. The neural network model leverages cumulative failure sequences from

previous sprints as training data, providing a robust mechanism to predict reliability for subsequent sprints.

This approach not only enhances prediction accuracy but also offers more stability in results, especially during early testing phases when data is sparse. In contrast to SRGMs, which may struggle with limited data, the proposed model compensates by incorporating patterns from historical sprints, thereby reducing early-stage prediction errors and ensuring more reliable fault estimation.

The remainder of this paper is organized as follows: Sect. 2 introduces the proposed neural network model for multi-sprint software. Section 3 describes the implementation process and data sources used for validation. Section 4 presents the results and compares them with existing models, highlighting the superior performance of the neural network approach. Finally, Sect. 5 concludes with discussions on the applicability of the model and directions for future research.

2 Proposed Neural Network Model for Multi-sprint Software

2.1 Model Overview

The proposed neural network model leverages data from previous sprints to predict reliability in the current sprint. Agile software development, characterized by iterative sprints, benefits from the availability of historical data to model and forecast software reliability trends effectively. By integrating fault data from prior sprints, the model captures patterns and dependencies that influence the fault behavior of the current sprint, enhancing the accuracy of predictions. Unlike traditional software reliability growth models, which often treat each release or sprint as an independent event, this model uses the correlation between sprints to inform predictions. By learning from previous sprint data, the model becomes more robust and capable of adapting to evolving software development environments.

2.2 Model Implementation

The neural network model employs a Bidirectional Long Short-Term Memory (Bi-LSTM) architecture to capture temporal dependencies in sprint data as shown in Fig. 1. The Bi-LSTM is particularly suitable for this task because it processes information in both forward and backward directions, allowing it to learn complex relationships and long-range dependencies from sequential sprint data. This architecture enhances the model's ability to identify patterns that may not be apparent when using unidirectional LSTM or other traditional methods. The implementation process involves several key steps:

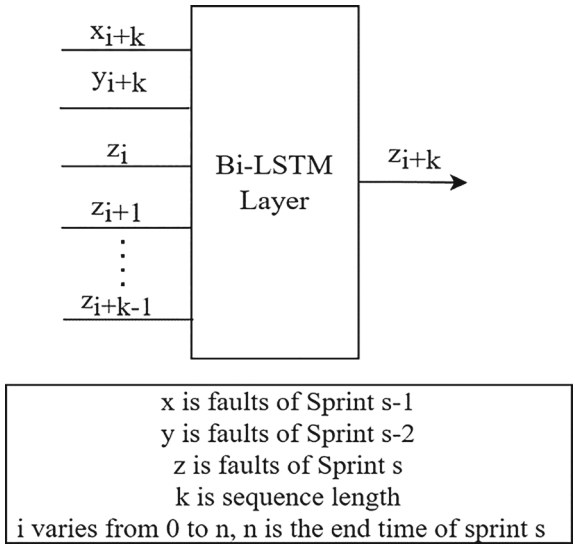


Fig. 1 Model architecture

Data Normalization: Before training, the fault data from previous and current sprints is normalized using a Min–Max Scaler. Normalization ensures that all input values are scaled within a range of [0, 1], improving the convergence rate and stability of the training process. This step is crucial as it prevents bias due to varying magnitudes in fault data across different sprints.

Input Layer: The model's input consists of cumulative fault data sequences from previous sprints and the initial faults observed in the current sprint. By using this structured input, the model learns the relationship between the past and present fault behaviors, enabling it to predict the evolution of faults during the current sprint.

Bi-LSTM Layer: The Bi-LSTM layer processes the sequential input data from previous and current sprints bidirectionally, which allows the model to understand dependencies in both forward and backward time directions. This bidirectional nature is particularly valuable in agile development, where patterns from past sprints significantly influence the ongoing sprint, and knowledge of future data (in a reverse manner) helps refine predictions.

Output: The output layer provides the predicted fault count for the next interval of the current sprint. The dense layer after the BiLSTM utilizes a linear activation function to map the features extracted by the BiLSTM to the output space.

Optimizer: The model is trained using the Adam optimizer, which combines the advantages of both adaptive learning rate and momentum. Adam is well-suited for training deep learning models like Bi-LSTM, as it adapts the learning rate dynamically based on past gradient information, ensuring faster convergence and improved accuracy.

2.3 Model Operation and Data Flow

The operation of the proposed model follows these steps:

Data Collection and Normalization: Fault data from previous sprints and initial fault observations from the current sprint are collected. The data is then normalized using the Min–Max Scaler to ensure consistent scaling and facilitate efficient learning.

Model Training: The neural network is trained with historical sprint data, where sequences of cumulative faults from previous sprints serve as input patterns. The target output is the predicted cumulative faults for the current sprint. The Bi-LSTM processes these sequences bidirectionally, allowing the network to capture and leverage the complex, time-dependent relationships between sprints.

Prediction of Faults in the Current Sprint: Once trained, the model uses data from the previous sprint and the first few intervals of the current sprint to make predictions. As new fault data becomes available during the sprint, the model continuously updates its inputs to predict the subsequent fault values, ensuring that predictions are dynamically refined.

Evaluation Metrics and Optimization: The Adam optimizer is used during training to minimize the mean squared error (MSE) between the predicted and actual faults.

3 Application and Results

3.1 Data and Evaluation Criteria

The proposed model is validated using the Jira Datasets [14] consisting of 6 sprints. The Proposed LSTM-based multi-sprint model (LSTM-MS) is compared against two existing models Rawat Model [15] and Saraf Model [16]. The evaluation of these models is based on their performance across six sprints from the JIRA dataset. The comparison uses five error metrics: Mean Squared Error (MSE), Adjusted R-squared, Bias, Variation, and Akaike Information Criterion (AIC). The aim is to demonstrate the accuracy and robustness of the proposed model in predicting software faults when contrasted with the existing approaches.

3.2 Results

The proposed model shows significantly lower MSE values across all sprints compared to both the Rawat and Saraf models. This indicates that the proposed model has a higher accuracy in predicting faults. Additionally, the Adjusted R-squared

values for the proposed model remain consistently above 0.999, indicating a very strong fit to the actual data. In contrast, the Rawat and Saraf models display lower Adjusted R-squared values, with the Saraf model occasionally dropping below 0.97, reflecting a weaker fit.

Bias measures the systematic deviation of the predicted values from the actual faults. The proposed model maintains a low bias across all sprints, indicating that its predictions are consistently close to the actual values. For instance, in Sprint 1, the bias of the proposed model is − 0.026, while the Rawat model exhibits a higher bias of − 1.0913, and the Saraf model shows a positive bias of 1.0137. This pattern persists across other sprints, highlighting the superior accuracy of the proposed model.

The variation metric further supports this observation. The proposed model demonstrates lower variation values, especially in early sprints, which implies greater stability in prediction. The Rawat and Saraf models show higher variation, indicating a less consistent prediction performance. For example, in Sprint 1, the variation for the proposed model is 1.6287, whereas the Rawat model shows 3.9566 and the Saraf model exhibits 12.0628.

The Akaike Information Criterion (AIC) is used to assess the quality of each model, with lower AIC values indicating a better model fit. The proposed model consistently shows the lowest AIC values across all six sprints. For Sprint 1, the proposed model's AIC is 133.8636, compared to 227.4361 for the Rawat model and 309.7431 for the Saraf model. This trend continues throughout all sprints, demonstrating that the proposed model provides accurate predictions with greater model efficiency (Table 1).

The results show that the proposed model's predictions are consistently the closest to the actual fault counts across all six sprints. For instance, in Sprint 1, the proposed model predicts 108.7868 faults, which is nearly identical to the actual 108 faults. In contrast, the Rawat and Saraf models underpredict with 96.27403 and 74.82382 faults, respectively. This pattern of higher accuracy for the proposed model persists across subsequent sprints. The proposed model's prediction of faults is much closer to the actual faults compared to the Rawat model and the Saraf model, which significantly underestimate the faults.

To visually demonstrate the superiority of the model, actual faults versus predicted faults have been plotted for each sprint, comparing it with the predictions of the three existing models as shown in Fig. 2. Additionally, fault prediction at the end time of each sprint is given in Table 2 to showcase the model's accuracy over existing approaches.

The proposed model performs better as compared to existing models by incorporating temporal dependencies using previous sprint data and utilizing a Bi-Directional LSTM with a logarithmic encoding. It effectively captured fault relationships, learned patterns, and dependencies, leading to improved fault prediction accuracy. Overall, the proposed model offers a comprehensive representation of faults in agile software development, demonstrating superior performance across multiple sprints. It stands as a valuable tool for enhancing software reliability in real-world applications.

Table 1 Error comparison

Error	MSE	Adjusted R-squared	Bias	Variation	AIC
Sprint 1	2.577588	0.999377	− 0.026	1.6287	133.8636
	16.45469	0.995409	− 1.0913	3.9566	227.4361
	142.9003	0.96406	1.0137	12.0628	309.7431
Sprint 2	6.563602	0.999382	0.5567	2.5523	114.6007
	31.51504	0.995634	1.7021	5.441	190.6503
	100.2255	0.98622	− 2.4838	9.864	225.1301
Sprint 3	9.08685	0.999471	− 0.3412	3.043	163.2685
	29.55817	0.998018	1.6366	5.2561	232.2568
	308.9351	0.979459	− 3.8025	17.397	318.7724
Sprint 4	4.488384	0.999338	0.5194	2.0981	100.8034
	10.98392	0.997582	1.1866	3.1493	153.795
	33.99081	0.992589	− 2.5677	5.3269	186.2768
Sprint 5	10.86894	0.999222	0.4226	3.3237	163.0183
	22.49441	0.998127	0.537	4.7792	215.9338
	37.94549	0.996813	1.3751	6.0897	235.065
Sprint 6	6.251744	0.998609	− 0.1363	2.5262	202.8214
	8.965589	0.997811	0.5361	2.9771	242.9347
	333.5274	0.954944	9.8137	15.5649	388.1098

4 Discussion and Future Work

The analysis demonstrates that the proposed model, utilizing historical sprint data and a neural network approach, consistently outperforms the Rawat and Saraf models across multiple metrics. Its ability to maintain low MSE values and high Adjusted R-squared values highlights its robustness in predicting software faults accurately. The low bias and variation values confirm that the model provides stable and unbiased predictions, even in early and later sprints.

The lower AIC values of the proposed model indicate not only a better fit to the data but also an efficient use of model parameters. By leveraging previous sprint data effectively, the proposed model adapts well to the agile development environment and shows promise as a reliable tool for real-time fault prediction in multi-sprint software projects. The neural network approach proves advantageous in the context of agile development, where frequent, short sprints require quick adaptation of predictive models. The use of cumulative failure sequences from past sprints helps overcome data limitations typical in early sprint phases. By refining the model with new data as it becomes available, the proposed approach outperforms static models that rely solely on initial data sets. While the proposed model shows strong performance, the current model does not account for the correction of faults during sprints, which

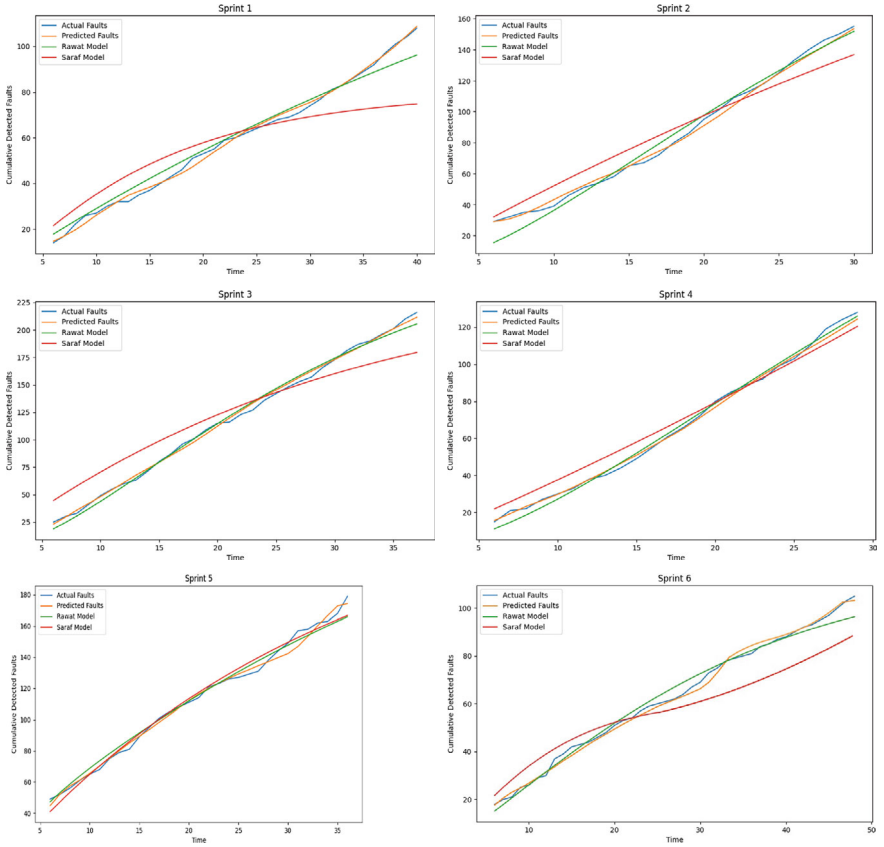


Fig. 2 Goodness of fit test

Table 2 End time fault prediction

Sprint	Actual faults	Proposed model	Rawat model	Saraf model
Sprint 1	108	108.7868	96.27403	74.82382
Sprint 2	155	153.5015	151.763	136.7732
Sprint 3	216	211.5918	205.4018	179.5026
Sprint 4	128	124.4027	121.9794	120.3831
Sprint 5	179	174.3546	165.8343	166.7214
Sprint 6	105	103.2505	96.41277	60.93631

limits its ability to dynamically track software reliability. Also, the model treats all faults equally, not distinguishing based on severity. Therefore, in future work we will try to integrate fault correction data and fault severity, helping teams prioritize and allocate resources efficiently, improving the model's ability to reflect the software's

real-time reliability during sprints. The model will be validated on various datasets to enhance its adaptability and generalizability across different agile projects.

5 Conclusion

This paper presents a LSTM-based model for predicting software reliability in agile environments. By leveraging historical sprint data and utilizing a Bi-LSTM architecture with a sliding window approach, the proposed model adapts effectively to the dynamic nature of agile development. Compared to existing models such as the Rawat and Saraf models, the proposed approach demonstrates superior accuracy, stability, and adaptability, as validated through the JIRA dataset. The model consistently achieves lower MSE values, higher Adjusted R-squared values, and better AIC scores, indicating its robustness in fault prediction. Overall, the proposed model shows promising result for improving software reliability in agile settings, offering development teams a reliable and adaptable method for anticipating and managing software faults throughout the sprint cycles.

References

1. Koteska, B., Mishev, A.: Agile software testing technologies in a large scale project. Balk. Conf. Inform. (2012)
2. Malik, A.: Software Testing Process in Agile Development (2008)
3. Li, Q., Pham, H.: Software reliability modeling incorporating fault detection and fault correction processes with testing coverage and fault amount dependency. Mathematics **10**(1) (2022). https://doi.org/10.3390/math10010060
4. Jia, L., Yang, B., Guo, S., Park, D.H.: Software reliability modeling considering fault correction process. In: IEICE Transactions on Information and Systems, vol. E93-D, no. 1, pp. 185–188 (2010). https://doi.org/10.1587/transinf.E93.D.185
5. Kapur, P.K., et al.: A survey of software reliability factor. IEEE Trans. Reliab. **1**(3), 799–817 (2011). https://doi.org/10.9790/0661-1215055
6. Ohba, M.: Inflection S-Shaped Software Reliability Growth Model, pp. 144–162. Springer, Berlin, Heidelberg (1984). https://doi.org/10.1007/978-3-642-45587-2_10
7. Zhu, M., Pham, H.: A generalized multiple environmental factors software reliability model with stochastic fault detection process. Ann. Oper. Res. **311**(1), 525–546 (2022). https://doi.org/10.1007/S10479-020-03732-3/FIGURES/10
8. Singhal, S., Kapur, P.K., Kumar, V., Panwar, S.: Stochastic debugging based reliability growth models for open source software project. Ann. Oper. Res. 1–39 (2023). https://doi.org/10.1007/S10479-023-05240-6/TABLES/9
9. Karunanithi, N., Malaiya, Y.K., Whitley, D.: Prediction of software reliability using neural networks. Proceedings—International Symposium on Software Reliability Engineering ISSRE, pp. 124–130, 1991. https://doi.org/10.1109/ISSRE.1991.145366
10. Hu, Q.P., Dai, Y.S., Xie, M., Ng, S.H.: Early software reliability prediction with extended ANN model. Proceedings—International Computer Software & Applications Conference, vol. 2, pp. 234–239 (2006). https://doi.org/10.1109/COMPSAC.2006.130

11. Bisi, M., Goyal, N.K.: Prediction of software inter-failure times using artificial neural network and particle swarm optimisation models. Int. J. Softw. Eng. Technol. Appl. **1**(2/3/4), 222 (2015). https://doi.org/10.1504/IJSETA.2015.075629
12. Wang, J., Zhang, C.: Software reliability prediction using a deep learning model based on the RNN encoder–decoder. Reliab. Eng. Syst. Saf. **170**, 73–82 (2018). https://doi.org/10.1016/J.RESS.2017.10.019
13. Samal, U., Kumar, A.: A neural network approach for software reliability prediction. Int. J. Reliab. Qual. Saf. Eng. **31**(3), 2450009 (2024). https://doi.org/10.1142/S0218539324500098
14. Montgomery, L., Lüders, C., Maalej, P.D.W.: The Public Jira Dataset (2022). https://doi.org/10.5281/ZENODO.5901804
15. Rawat, S., Goyal, N., Ram, M.: Software reliability growth modeling for agile software development. Int. J. Appl. Math. Comput. Sci. **27**(4), 777–783 (2017). https://doi.org/10.1515/AMCS-2017-0054
16. Saraf, I., Iqbal, J., Shrivastava, A.K., Khurshid, S.: Modelling reliability growth for multi-version open source software considering varied testing and debugging factors. Qual. Reliab. Eng. Int. **38**(4), 1814–1825 (2022). https://doi.org/10.1002/QRE.3048

A Temporal Decomposition-Based LSTM Model for Forecasting Reservoir Inflow Using Real-Time Data



Kshitij Tandon and Subhamoy Sen

Abstract Accurate forecasting of reservoir inflow is essential for efficient water resource management and operations. Considering the nonlinear and nonstationary characteristics of real-world hydrological data, this paper introduces a novel model (EMD-LSTM) that employs empirical mode decomposition (EMD) in conjunction with long short-term memory (LSTM) networks for predicting daily reservoir inflow up to 10 days lead time. The accuracy of the model is assessed using Nash–Sutcliffe efficiency (NSE), root mean square error (RMSE) and mean absolute error (MAE). To assess its effectiveness, the proposed EMD-LSTM model is compared with standard artificial neural networks (ANN) and LSTM models for 3-day, 7-day, and 10-day lead times. Daily inflow data recorded using water level recorders from 2013 to 2022 for the Bhakra reservoir, located on the Sutlej River in Himachal Pradesh, India, serves as the case study. The EMD-LSTM model shows promising results, achieving an NSE of up to 0.94 in the validation phase for the 10-day forecast, outperforming the ANN and LSTM models. These findings suggest that the EMD-LSTM model offers valuable insights for decision-making and can enhance the safety and efficiency of reservoir operations.

Keywords Bhakra reservoir · Real-time forecasting · ANN · LSTM · EMD-LSTM

1 Introduction

Forecasting reservoir inflow is essential for the efficient management and operation of reservoirs. The accuracy of these forecasts significantly impacts reservoir operations, power generation, flood control, and the optimal distribution of water

K. Tandon · S. Sen (✉)
i4S Laboratory, Indian Institute of Technology Mandi, Mandi, Himachal Pradesh, India
e-mail: subhamoy@iitmandi.ac.in

K. Tandon
e-mail: ktandon36@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_31

resources within a basin. The real hydrological data shows nonlinearity and non-stationarity due to the consequences of human actions and natural factors posing a challenge in forecasting the inflow. To deal with such challenges, various forecasting methods have been developed over time prop [1, 2] adopting various methodologies ranging from first principal-based to data-based time series modeling or more recently machine learning (ML) algorithms.

Time series models based on data including the auto-regressive moving average (ARMA), [3] have been proposed to forecast floods with a forecast lead-time of up to three days. These models, however, do not consider nonlinearity and/or non-stationarity in the data. To deal with such problems, AI prediction models through ML approaches have been widely used. ANN models have found widespread application for reservoir inflow forecasting as they can handle the non-linearity associated with the reservoir inflow data [4]. In these ML-based predictor models, difficulties do originate from the selection process of the model parameters since their impact on the prediction precision is most significant. To handle time series horological data, Recurrent neural networks (RNN) have been widely used especially when the model has to be estimated using long sequence data. RNN, however, poses the problem of vanishing gradients and the effectiveness of RNNs in time series forecasting does not improve much while the required computational demand is at times substantial. Long short-term memory network (LSTM), an improvised form of RNN, solves the challenges of RNN in training of long sequence data and problem of vanishing gradients by utilizing constant error propagation within special memory cells of the LSTM networks. Many studies have successfully applied LSTM models for groundwater level prediction and daily runoff prediction [5, 6].

The runoff process is characterized as a highly intricate and dynamic time series that follows patterns across multiple scales. The above-mentioned prediction methods, however, do not take into consideration the random nature, trend, and periodicity of the real hydrological data, thereby rendering the AI prediction models sometimes incompetent. To extract such characteristics from the data, extensive research on decomposition-based deep learning models has been employed in reservoir inflow forecasting [7]. Various signal processing techniques, including wavelet analysis and empirical mode decomposition (EMD) [2, 8, 9] have emerged over time, integrating signal decomposition with time series analysis. To improve model performance, several decomposition techniques including the Fourier transform, EMD, and wavelet transforms have been integrated with regression models based on neural networks in this approach [10, 11].

In this paper, EMD is coupled with LSTM to enhance the accuracy of the forecast by decomposing the inflow time series into several intrinsic mode functions (IMFs). These IMFs are then provided as input to the LSTM model. This model is further compared with the ANN and LSTM model to check for performance and accuracy. The following sections outline the remainder of the paper: Sect. 2 illustrates the proposed method, providing a detailed explanation of the ANN, LSTM, and EMD-LSTM models. Section 3 details the study area and the dataset utilized. Section 4 illustrates the results and their interpretation along with the managerial implications. Finally, Sect. 5 concludes with key insights of the study.

2 Methodology

2.1 Artificial Neural Network (ANN)

ANN model is a supervised machine learning algorithm wherein the underlying correlation between training data (as input) and pertinent target data or labels (as output) are modeled using several layers of interconnected neurons. The model represents a complex and nonlinear process that enables identifying the suitable labels to the input once trained extensively using a rich archive of training data. In this study, the inflow along with its two lagged values are provided as input to the model, and the inflow at the desired lead time (3 days, 7 days, 10 days) is considered as output. The input data is normalized between 0 and 1 in order to improve the simulation performance. The model is structured with an input layer comprising three nodes, two hidden layers, each containing 32 neurons, and a single output node. ReLU activation function is used for the ANN model. Following, the network architecture is defined through a rigorous trial and error process prior to training it using a part of the available time series data.

2.2 Long Short-Term Memory Model (LSTM)

The traditional or vanilla Recurrent neural network (RNN) is a class of neural networks that uses recursive approaches to model sequential data. The output of the network at time step t is influenced by both the input at that time step and the recurrent inputs from earlier time steps. However, time series forecasting problems demand long-term dependencies for which the RNN is often rendered incapable due to the much-mentioned problem of vanishing and exploding gradients. Hochreiter et al. [12] proposed the Long short-term memory model (LSTM) equipped with memory blocks to deal with vanishing gradients by memorizing the network parameters for a long duration. This article employs this approach for modeling the time series. The model consists of one hidden layer with twelve nodes. The inflow with its two lagged values is provided as an input sequence to the model and ReLU activation function is employed in the model.

2.3 Empirical Mode Decomposition (EMD)

EMD proposed is a signal decomposition method specifically designed for time series data that is both non-linear and non-stationary. The core concept of EMD is to decompose the time series into a finite set of Intrinsic Mode Functions (IMFs) with one residual component. A key advantage of the EMD technique is that the resulting components are time series data that oscillate around zero, making them suitable for

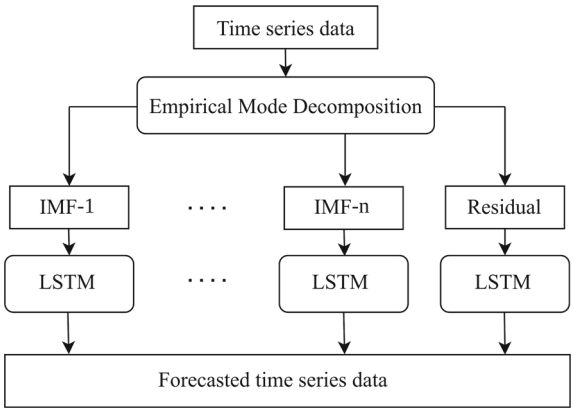


Fig. 1 Flowchart describing the proposed EMD-LSTM model

effective analysis. Also, the decomposed IMFs and the residue can be reassembled to realize the original time series without any information loss. By principle, IMFs should satisfy the following two criteria,

1. The total number of zeros and extremes in the complete set of Intrinsic Mode Functions (IMFs) should be equal.
2. The mean value of the envelope, determined by connecting local maxima and minima, is zero.

The original sequence is decomposed into a set of intrinsic mode functions (IMFs) with one residual component as follows:

$$x_k = \sum_{j=1}^{n} c_j(t) + r_n(t) \tag{1}$$

where, x_k is the original time series, $\sum_{j=1}^{n} c_j(t)$ are the IMFs and $r_n(t)$ is the residual.

In the proposed model, EMD is combined with LSTM wherein instead of raw data, the highly correlated IMFs are assumed as input with a presumption that these IMFs store the majority of the information required for the prediction and the noise or random nature in the data can be judiciously kept out of the modeling by removing the corresponding IMFs. The flow chart describing the proposed methodology is shown in Fig. 1.

2.4 Model Performance Evaluation

The time series model once trained needs to be validated for its prediction ability which is undertaken through several error and statistical indices, discussed in the following.

Nash–Sutcliffe efficiency (NSE)

NSE estimates the goodness of fit. It is evaluated by the following equation:

$$NSE = 1 - \frac{\sum_{i=1}^{N} \left(y_i^{act} - y_i^{pred}\right)^2}{\sum_{i=1}^{N} \left(y_i^{act} - y_{act}^{avg}\right)^2} \tag{2}$$

where, y_i^{pred}, y_i^{act} are the predicted and actual observations of inflow, N denotes the number of samples, y_{act}^{avg} is the mean value of measured sample.

Mean absolute error (MAE)

MAE depicts the mean of the absolute difference between the observed value y_i^{act} and the predicted value y_i^{pred}. It is given as:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| y_i^{act} - y_i^{pred} \right| \tag{3}$$

Root mean square error (RMSE)

RMSE assesses a model's predictive performance. A smaller RMSE indicates higher accuracy in the prediction model. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i^{act} - y_i^{pred}\right)^2} \tag{4}$$

3 Study Area and Data Description

The Bhakra Nangal dam is the second-highest concrete gravity dam in Asia located across the river Satluj in the state of Himachal Pradesh, India. The upstream catchment area of the dam is 56,980 km^2. Rainfall within the catchment area fluctuates across the basin, with an annual average of 875 mm. The Govindsagar reservoir has a total capacity of 9876 m^3 and a live storage capacity of 7814 m^3 above a dead storage level of 445.62 m. The area covered by the reservoir is 168.35 km^2 when full. The total runoff is 16,441 m^3 for a mean year. River Satluj, which originates from the Himalayas, forms a major source of water in the region. Due to climate change and rainfall pattern change, the magnitude of the streamflow is of concern for water management. The river flows are high from June to September because of the monsoon rains and melting of snow. Eventually, accurate forecasting of the inflow

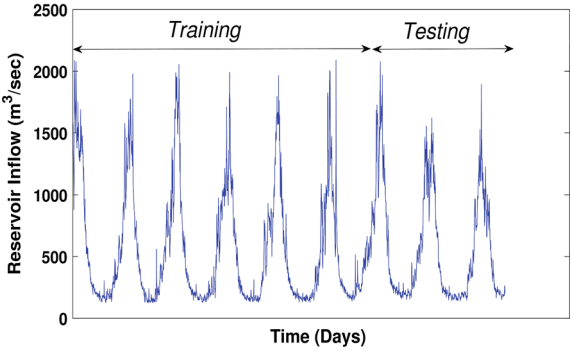


Fig. 2 Reservoir inflow data of Bhakra reservoir from 2013 to 2022

for such a huge reservoir is imperative for the reservoir operation, flood control and also ensures the structural safety of the dam at the same time.

The real time observed daily inflow recorded using water level recording devices for the period 2013–2022 is available on the website of Bhakra Beas Management Board (BBMB). A univariate time series forecasting is performed to forecast the inflow for different lead times (3, 5 and 7 days) with the objective of flood forecasting, reservoir safety, and health assessment. The daily inflow data is split into training (calibration) and testing (validation) with a 70:30 ratio each respectively as shown in Fig. 2.

4 Results and Discussion

The reservoir inflow time series was decomposed into nine IMF components and a residual trend using EMD, with the components arranged in decreasing order of frequency. The final component, the residue, captures the overall trend of the series. For brevity, all IMFs are not displayed here. To understand the physical significance of the extracted IMFs, a cross-correlation analysis was conducted between each IMF and the original inflow data. Table 1 illustrates that IMFs 2, 3, 4, 5, and 6 exhibit a significant positive correlation with the original series.

In order to reduce the impact of noise (or random nature present in the data) on the forecast and to maintain economy in computation, the selection of the vital IMF components is essential. The selection is done based on the maximum positive correlation of the concerned IMF with the original time series data. Figure 3 depicts the variation of correlation coefficients with the number of IMFs considered. The correlation increases with the increase in number of IMFs (arranged in descending correlation coefficient from Table 1). The curve shows an increasing correlation up to six IMF's and becomes constant thereafter allowing the identification of an elbow to decide the practical numbers of IMFs to be considered. Accordingly, six IMFs are considered for our proposed EMD-LSTM model. Figure 4 shows the graphical representation of the six IMFs employed for the EMD-LSTM model.

Table 1 Correlation of IMFs with original data

IMF	Cross-correlation coefficient
1	0.0285
2	0.1731
3	0.1691
4	0.4301
5	0.8199
6	0.1129
7	0.0829
8	0.0484
9	0.0757
Residual	0.0742

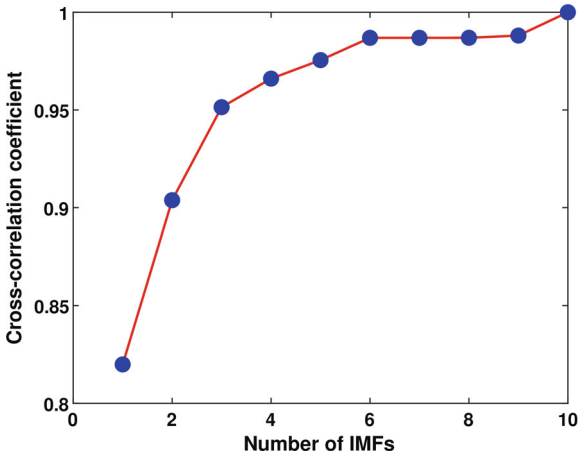


Fig. 3 Variation in cross-correlation coefficient with the number of IMFs

Figures 5, 6 and 7 shows the forecast plots along with the scatter plots of the actual inflow and predicted inflow from the three models (ANN, LSTM and EMD-LSTM) for three lead times (3, 7 and 10 days). The model's forecasting accuracy exhibits an inverse relationship with lead time. The comparison results reveal that the EMD-LSTM model provides relatively better performance and indicates superior model fit to observed data compared to LSTM and ANN models. Despite its limitations, the ANN model can extract underlying physical principles from observed data for most of the parts, yet it does fail to replicate the complete physical process for all data ranges. Eventually, ANN model fails to represent the variability for higher lead times. Broadly, the scatter plot depicts that EMD-LSTM model can predict inflow much better than its alternatives.

Table 2 depicts three performance indices (NSE, RMSE and MAE) for the performance evaluation of the three competing models for inflow prediction. The NSE values range from 0.84 to 0.92 for ANN model, 0.93 to 0.96 for the LSTM model, and

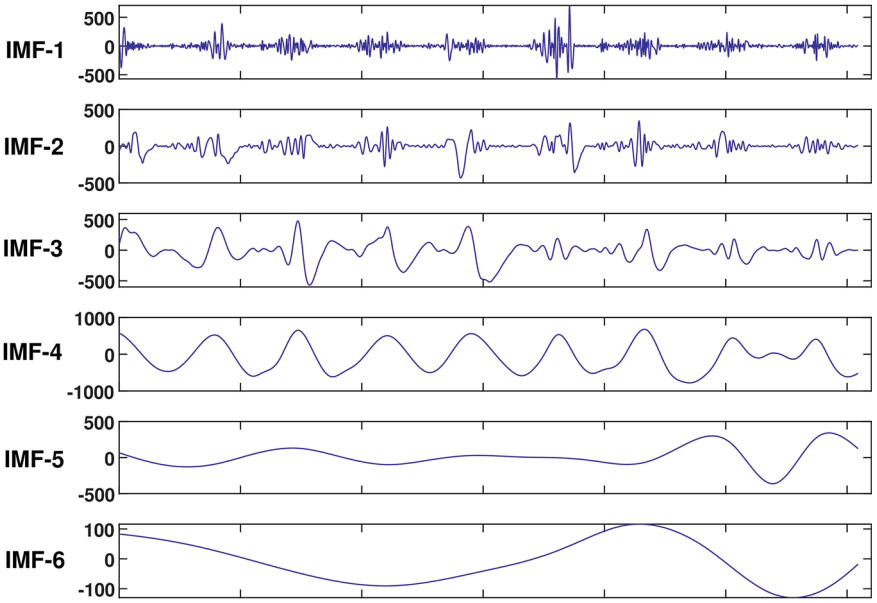


Fig. 4 The intrinsic mode functions (IMFs) employed for prediction in EMD-LSTM model

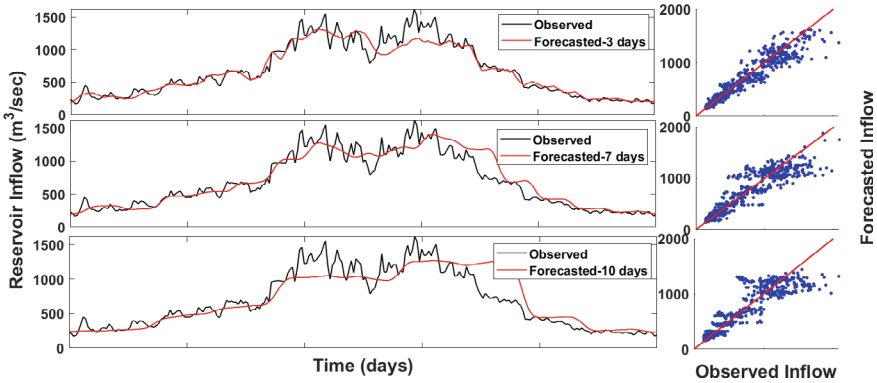


Fig. 5 Reservoir inflow forecast for 3 days, 7 days and 10 days ahead respectively by ANN model

0.94 to 0.98 for EMD-LSTM model for 3 days, 7 days, and 10 days lead times respectively. The RMSE increased significantly with the increase in the lead time for all candidate models. Nevertheless, EMD-LSTM yielded quite low RMSE compared to its alternatives (Fig. 8 and Table 2). The MAE results further signify that the proposed model predicts more accurately and the other alternatives have more errors.

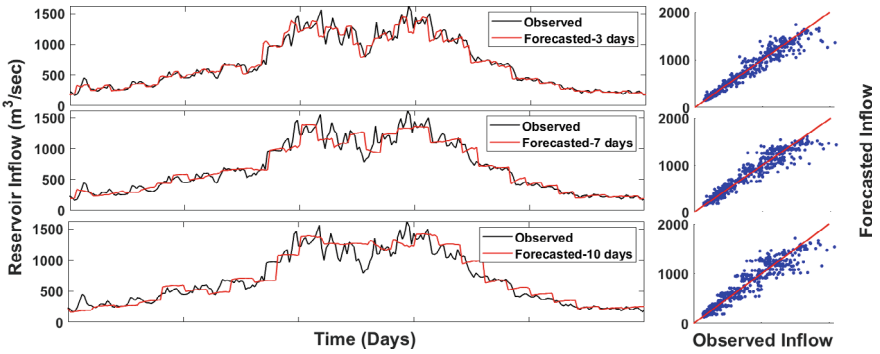


Fig. 6 Reservoir inflow forecast for 3 days, 7 days and 10 days ahead respectively by LSTM model

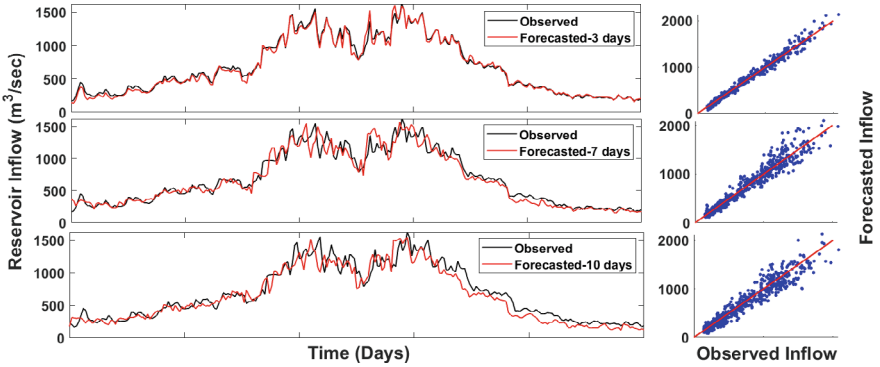


Fig. 7 Reservoir inflow forecast for 3 days, 7 days and 10 days ahead respectively by EMD-LSTM model

Table 2 Model performance evaluation for 3 days, 7 days and 10 days ahead lead time respectively

	NSE			RMSE (m^3/sec)			MAE %		
Lead time (days)	3	7	10	3	7	10	3	7	10
ANN	0.92	0.87	0.84	116.63	155.59	172.38	68.75	93.94	106.57
LSTM	0.96	0.94	0.93	87.86	101.96	110.33	45.62	59.71	66.58
EMD-LSTM	0.98	0.95	0.94	47.80	91.96	97.06	31.98	59.18	73.21

The results are consistent with findings from Zuo et al. [5] and Alizadeh et al. [6], which also demonstrated that LSTM-based models significantly outperform traditional methods for streamflow prediction. In this study, the proposed EMD-LSTM model integrates empirical mode decomposition (EMD) to reduce noise and enhance forecast accuracy, particularly for longer lead times as shown by NSE values reaching 0.94 for the 10-day forecast. To optimize computational efficiency and further reduce

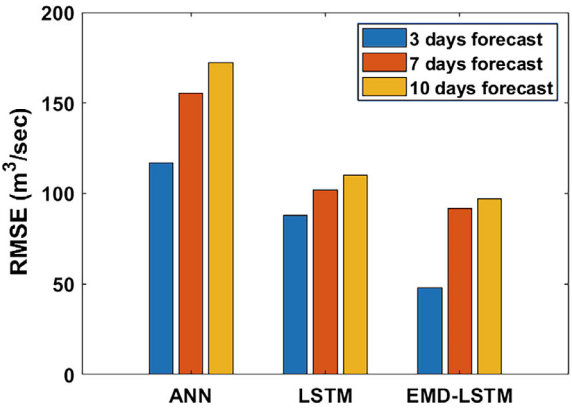


Fig. 8 Comparison of root mean square error for ANN, LSTM and EMD-LSTM model respectively

noise, key intrinsic mode functions (IMFs) were selected based on a correlation analysis. Moreover, as the lead time increases, the model's accuracy declines as evidenced by the performance metrics: NSE decreases, while RMSE and MAE increase. This trend aligns with previous findings by Le et al. [13] and Yuan et al. [14].

4.1 Managerial Implications

Reservoir inflow forecasting has wide-ranging managerial implications that optimize dam and water resource operations. The proposed EMD-LSTM model with its ability to accurately predict inflow up to 10 days in advance provides water resource managers with a powerful tool to anticipate high inflows. Accurate forecasts support flood control, allowing timely water releases to prevent downstream flooding and infrastructure damage. It also aids in water supply management by ensuring reliable allocation for agriculture, industry, and municipalities, especially during dry periods. For hydropower generation, inflow forecasting helps maximize electricity production while maintaining reservoir sustainability. In drought management, forecasts guide rationing and water storage to mitigate scarcity. It also supports environmental management by maintaining ecological flows for aquatic ecosystems. Operationally, inflow predictions assist with infrastructure maintenance, scheduling repairs during low inflow periods to minimize disruptions. In multi-reservoir systems, forecasts ensure coordinated water storage and release, balancing supply across the network. Furthermore, it contributes to financial planning by enhancing revenue forecasting in sectors like agriculture and energy leading to better budget and resource allocation. These insights can guide the development of more reliable early warning systems for water management authorities.

5 Conclusions

By coupling EMD and LSTM networks, the proposed model significantly improves forecasting accuracy and prediction performance. It effectively addresses the non-stationarity and noise present in the original time series data. For shorter lead times, all models perform well, delivering high accuracy and low error. However, for longer lead times, the proposed model outperforms others by a significant margin. The ANN model struggles to capture the underlying trend in the time series for longer lead times. In contrast, the LSTM model is well-suited for both model fitting and forecasting tasks. Additionally, EMD proves valuable by reducing noise and extracting the inherent trend and periodicity of the time series, resulting in improved prediction outcomes. Therefore, the proposed model offers a robust solution for real-time forecasting of reservoir capacity using water level recorder data, which can greatly enhance the efficiency of reservoir operations and maintenance.

References

1. Ramaswamy, V., Saleh, F.: Ensemble based forecasting and optimization framework to optimize releases from water supply reservoirs for flood control. Water Resour. Manag. **34**(3), 989–1004 (2020)
2. Bai, Y., Chen, Z., Xie, J., Li, C.: Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. J. Hydrol. **532**, 193–206 (2016)
3. Mohammadi, K., Eslami, H., Kahawita, R.: Parameter estimation of an ARMA model for river flow forecasting using goal programming. J. Hydrol. **331**(1–2), 293–299 (2006)
4. Wu, J.S., Han, J., Annambhotla, S., Bryant, S.: Artificial neural networks for forecasting watershed runoff and stream flows. J. Hydrol. Eng. **10**(3), 216–222 (2005)
5. Zuo, G., Luo, J., Wang, N., Lian, Y., He, X.: Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. J. Hydrol. **585**, 124776 (2020)
6. Alizadeh, B., Bafti, A.G., Kamangir, H., Zhang, Y., Wright, D.B., Franz, K.J.: A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. J. Hydrol. **601**, 126526 (2021)
7. Tang, L., Wang, S., He, K., Wang, S.: A novel mode-characteristic-based decomposition ensemble model for nuclear energy consumption forecasting. Ann. Oper. Res. **234**(1), 111–132 (2015)
8. Kisi, O., Latifoğlu, L., Latifoğlu, F.: Investigation of empirical mode decom- position in forecasting of hydrological time series. Water Resour. Manag. **28**(12), 4045–4057 (2014)
9. Wang, Y.-Y., Wang, W.-C., Xu, D.-M., Zhao, Y.-W., Zang, H.-F.: A compound approach for ten-day runoff prediction by coupling wavelet denoising, attention mechanism, and LSTM based on GPU parallel acceleration technology. Earth Sci. Inf. **17**(2), 1281–1299 (2024)
10. Maheswaran, R., Khosa, R.: Wavelets-based non-linear model for real-time daily flow forecasting in Krishna River. J. Hydroinf. **15**(3), 1022–1041 (2013)
11. Xu, D.-M., Liao, A.-D., Wang, W., Tian, W.-C., Zang, H.-F.: Improved monthly runoff time series prediction using the CABES-LSTM mixture model based on CEEMDAN-VMD decomposition. J. Hydroinf. **26**(1), 255–283 (2024)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

13. Le, X.-H., Ho, H.V., Lee, G., Jung, S.: Application of long short-term memory (LSTM) neural network for flood forecasting. Water **11**(7), 1387 (2019)
14. Yuan, R., Cai, S., Liao, W., Lei, X., Zhang, Y., Yin, Z., Ding, G., Wang, J., Xu, Y.: Daily runoff forecasting using ensemble empirical mode decomposition and long short-term memory. Front. Earth Sci. **9**, 621780 (2021)

Weak Key Analysis of HQC



Ajay S. Rodge and Bhupendra Singh

Abstract The Hamming Quasi-Cyclic (HQC) cryptosystem is a prominent candidate in post-quantum cryptography, leveraging the hardness of decoding quasi-cyclic codes to ensure security against quantum adversaries. Despite its robustness, the security of HQC can be compromised by the presence of weak keys—keys that, due to structural anomalies or insufficient randomness, significantly reduce the cryptosystem's resistance to attacks. This research presents a comprehensive analysis of weak keys within the HQC framework, emphasizing their role in enabling decryption failure attacks. We investigate how deviations in key properties, such as abnormal Hamming weights, patterned bit sequences, and improper quasi-cyclic structures, can lead to increased decryption failure rates (DFR). Through a combination of theoretical analysis and experimental validation, we demonstrate how these weak keys can be exploited to facilitate key recovery via decryption failures. Our study also explores the impact of weak keys on Information Set Decoding (ISD) attacks, revealing how structural weaknesses in the keys can dramatically reduce the complexity of such attacks. To mitigate these vulnerabilities, we propose a set of detection and prevention strategies aimed at ensuring the generation of strong keys in HQC. These include rigorous Hamming weight checks, enhanced randomness validation, and structural pattern detection during key generation. Our findings underscore the critical importance of robust key management in maintaining the security of HQC and similar post-quantum cryptosystems. This research contributes to the broader field of cryptographic security by providing practical insights and methodologies for strengthening post-quantum encryption schemes against advanced cryptanalytic attacks.

Keywords Cryptanalysis · HQC · Weak key analysis · Security · Post quantum cryptography

A. S. Rodge (✉)
Computer Science and Data Science Department, GHRCEM, Pune, India
e-mail: Ajayrodge713@gmail.com

B. Singh
Defence Research and Development Organization-DRDO, Bangalore, India
e-mail: Bhupendra.cair@gov.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_32

1 Weak Key Identification

In a cryptographic systems, the security of encryption and decryption processes fundamentally depends on the robustness of the keys involved. A weak key is one that, due to its specific structure or properties, fails to provide the necessary level of security, making the cryptosystem vulnerable to various forms of attack. The presence of weak keys in a cryptosystem can significantly undermine its security, enabling attackers to break encryption, recover plaintext, or even extract the secret key.

Weak Key Analysis is the systematic study and identification of such keys to ensure they do not compromise the cryptosystem. In the context of post-quantum cryptographic schemes like HQC (Hamming Quasi-Cyclic), where security is designed to withstand attacks from quantum computers, weak keys can be particularly dangerous. These keys might result from low Hamming weight, lack of randomness, structural patterns, or improper cyclic properties, all of which can be exploited to reduce the complexity of cryptanalytic attacks.

The objective of a Weak Key Analysis Scheme is to identify, analyze, and mitigate the impact of weak keys within a cryptosystem. This scheme involves several key steps, including detecting keys with abnormal properties, testing their randomness and structural integrity, and ensuring that they adhere to the required security standards. By implementing such a scheme, it is possible to significantly reduce the likelihood of cryptographic failure due to weak keys, thereby enhancing the overall robustness of the cryptosystem.

Creating a detailed scheme for weak key analysis in the HQC (Hamming Quasi-Cyclic) cryptosystem involves outlining a systematic approach to detect, analyze, and mitigate weak keys during the key generation and verification process. Below is a structured scheme that can be implemented to ensure the security of the HQC cryptosystem:

1.1 Key Generation Process

Initialization. Secure Randomness Source: Initialize a cryptographically secure random number generator (CSPRNG) that will be used to generate the secret vectors and polynomials. Ensure that the CSPRNG is seeded with high-entropy data.

Parameter Selection: Choose the cryptographic parameters for the HQC scheme, such as the degree of the polynomials, the length of the vectors, and the target Hamming weight w for the secret vectors.

Secret Key Generation. Vector Generation: Generate the secret vectors $\mathbf{x}$ and $\mathbf{y}$ as sparse binary vectors with exactly w non-zero entries. This can be done by randomly selecting w positions out of the total length and setting them to 1, with the rest set to 0.

Hamming Weight Check: Immediately check the Hamming weight of **x** and **y** to ensure they meet the required threshold w. If the Hamming weight is incorrect, discard the vectors and regenerate.

Public Key Generation. Matrix Construction: Construct the public key matrix **H** based on the secret vectors **x** and **y**. The matrix should be quasi-cyclic, meaning that each row is a cyclic shift of the previous row. Cyclic Property Verification: Verify the quasi-cyclic nature of the matrix **H**. Ensure that the cyclic shifts are consistent and that the matrix does not exhibit any unexpected patterns or symmetries.

2 Example of a Weak Key

2.1 Setup

Assume we are working with an HQC scheme where the secret keys **x** and **y** are binary vectors of length 256, with an expected Hamming weight of $w = 40$. This means that each vector should have exactly 40 non-zero entries (1 s), and the rest should be zeros.

2.2 Weak Secret Key Vectors

Let's consider the following secret vectors **x** and **y**:

Weak Secret Vector $\boldsymbol{x}$:

$$\boldsymbol{x} = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \ldots, 0, 1)$$

Hamming Weight: The Hamming weight of **x** is only 2, which is far below the required 40.

This makes **x** a weak key because it drastically reduces the search space for an attacker, making brute-force attacks or decoding attacks significantly easier.

Weak Secret Vector y:

$$\mathbf{y} = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, \ldots, 1)$$

Hamming Weight: The Hamming weight of **y** is 256, which is the maximum possible for this vector. This is also a weak key because a high Hamming weight can create predictable patterns in the public key, making it easier for an attacker to reverse-engineer the private key.

2.3 Public Key Construction

Assume that the public key matrix **H** is generated using these weak secret vectors. Since **x** has only two non-zero entries, the corresponding rows in **H** will have very few non-zero elements, which could make certain rows identical or very similar. This lack of randomness in the public key matrix could expose it to structural attacks.

Low Hamming Weight (x). With only two non-zero entries, the vector **x** significantly reduces the complexity of the decoding problem. An attacker can perform a brute-force search over all possible combinations of two positions to identify **x**, which is computationally trivial compared to the expected search space.

High Hamming Weight (*y*). A vector with a high Hamming weight (in this case, all 1 s) may introduce predictability into the public key. The corresponding public key matrix might have rows that are highly similar or even identical, reducing the effective entropy of the public key and making it easier for an attacker to infer the private key.

Cyclic Structure Weakness. The quasi-cyclic nature of the public key matrix might also be compromised if the weak keys result in shorter or repetitive cycles, further reducing the scheme's security.

2.4 Mitigation

To avoid generating such weak keys. Enforce Hamming Weight Constraints: During key generation, ensure that the vectors **x** and **y** have a Hamming weight close to the expected value $w = 40$. Introduce Randomness: Ensure that the vectors are generated using a secure and random process that avoids patterns like all 1 s or all 0 s. Post-Generation Checks: Implement checks to identify and discard weak keys before they are used in cryptographic operations.

3 Definition of Weak Keys in HQC

3.1 Weak Keys in the HQC Scheme Typically Fall into One or More of the Following Categories

Low Hamming Weight: Secret keys **x** and **y** with a Hamming weight significantly lower than the expected w. These keys reduce the search space for an attacker, making brute-force or decoding attacks easier.

High Hamming Weight: Secret keys with a Hamming weight close to or equal to the vector length n. Such keys can create predictable patterns in the public key matrix, potentially exposing it to algebraic attacks.

Structured Polynomials: Keys where the vectors $\mathbf{x}$ and $\mathbf{y}$ have patterns, such as all ones in certain positions or repetitive sequences. These patterns can make the public key matrix $\mathbf{H}$ less random and more susceptible to attacks.

Quasi-Cyclic Structure Weakness: Keys that lead to a public key matrix $\mathbf{H}$ with short cycles or unintended symmetries, which can be exploited by attackers to reduce the effective key space.

3.2 Specific Weak Keys and Range

Given that HQC keys are binary vectors of length n with a specific Hamming weight w, weak keys can be identified.

Low Hamming Weight Keys. Example Weak Key (Low Hamming Weight):
Suppose $n = 256$ and $w = 40$ is the expected Hamming weight.
A weak key example would be $\mathbf{x}$ with only 2 non-zero bits:

$$\mathbf{x} = (1, 0, 0, 0, \ldots, 0, 1)$$

The range of weak keys due to low Hamming weight can be approximated askeys with Hamming weight $w < 20$.

Range of Weak Keys (Low Hamming Weight):

Lower Bound: $w < 20$ (below half of the expected weight).
Upper Bound: $w = 39$ (just below the expected weight).

High Hamming Weight Keys. Example Weak Key (High Hamming Weight):

A weak key example would be $\mathbf{y}$ with almost all bits set to 1:

$$\mathbf{y} = (1, 1, 1, \ldots, 1, 0, 0, 0)$$

The range of weak keys due to high Hamming weight can be approximated as keys with Hamming weight $w > 200$.

Range of Weak Keys (High Hamming Weight):

Lower Bound: $w = 41$ (just above the expected weight).
Upper Bound: $w > 230$ (approaching the vector length n).

Structured Polynomial Keys. Example Weak Key (Structured):

A weak key might have a repetitive pattern:

$$\mathbf{x} = (1, 0, 1, 0, 1, 0, \ldots, 1, 0)$$

Such patterns reduce randomness and can be exploited. The range of such weak keys is harder to quantify exactly, but any key exhibiting non-random, repeating patterns would fall into this category.

Range of Weak Keys (Structured):

Detection: Keys where the distribution deviates significantly from a uniform distribution or exhibits clear repetitive patterns.

Quasi-cyclic Structure Weaknesses. Example Weak Key (Cyclic Weakness):

A weak key that generates a public key matrix $\mathbf{H}$ where a short cycle appears, such as:

$$\text{If } \mathbf{H}_k = \mathbf{H}_{k+32} \text{ for some } k$$

where 32 is much smaller than n.

Range of Weak Keys (Cyclic):

Detection: Any key that results in shorter cycles than expected or exhibits unintended symmetries in the quasi-cyclic matrix.

4 Summary of Ranges

Low Hamming Weight Keys: Typically, $w < 20$ (out of 256).
High Hamming Weight Keys: Typically, $w > 230$ (out of 256).
Structured Polynomial Keys: Any key with clear patterns, such as repetitive sequences.
Cyclic Structure Weakness: Keys leading to short cycles in the quasi-cyclic matrix, shorter than expected.

Detecting weak keys in the HQC (Hamming Quasi-Cyclic) cryptosystem involves implementing a series of checks and tests to identify keys that do not meet the required security properties. These checks focus on ensuring that the keys exhibit sufficient randomness, have appropriate Hamming weights, and do not introduce structural weaknesses into the public key matrix.

5 Detection of Weak Keys

5.1 Hamming Weight Check

The Hamming weight of a vector is the number of non-zero elements (1 s) in the vector. The HQC scheme expects the secret vectors **x** and **y** to have a specific Hamming weight w. Deviations from this expected weight can indicate weak keys [1].

Compute the Hamming weight of the secret vectors **x** and **y**:

$$w_x = \sum_{i=1}^{n} x_i, w_y = \sum_{i=1}^{n} y_i$$

Check if the weights are within the acceptable range:

$$w_{\min} \leq w_x \leq w_{\max}, w_{\min} \leq w_y \leq w_{\max}$$

where $w_{\min}$ and $w_{\max}$ are thresholds defining the acceptable range around the expected Hamming weight w.

If w_x or w_y is outside the range, flag the key as weak and regenerate the vectors.

5.2 Randomness Test

Randomness is crucial for ensuring that the keys do not exhibit patterns that could be exploited by an attacker. Various statistical tests can be used to check for randomness in the key vectors.

Chi-Square Test: This test compares the observed frequency of 1 s and 0 s in the vectors **x** and **y** to the expected frequency based on a uniform distribution.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency and E_i is the expected frequency for each bin.

Monobit Test: This test checks if the number of 1 s in the vector is approximately half of the total length.

$$S = \left| \sum_{i=1}^{n} x_i - \frac{n}{2} \right|$$

If S is significantly large, flag the key as weak.

Runs Test: This test checks for the frequency and length of runs of consecutive identical bits in the vector.

$$\text{Run length} = \max(\text{length of consecutive 1 s or 0 s})$$

If the run length is too high, flag the key as weak.

5.3 Structured Pattern Detection

Structured patterns in the vectors **x** and **y** can weaken the key. Detecting such patterns involves looking for repetitive sequences or regular intervals of 1 s and 0 s.

Fourier Transform: Apply a Fourier transform to detect periodicities in the sequence.

$$F(\mathbf{x}) = \sum_{k=1}^{n} x_k \mathrm{e}^{-2\pi ik/n}$$

If significant peaks are observed at specific frequencies, indicating periodic patterns, flag the key as weak.

Pattern Matching: Scan for known repetitive patterns (e.g., alternating 1 s and 0 s).

$$\text{Pattern found} \rightarrow \text{flag the key as weak}$$

5.4 Quasi-cyclic Structure Verification

In HQC, the public key is based on a quasi-cyclic matrix. Keys that result in shorter-than-expected cycles or exhibit unintended symmetries in the matrix are weak.

Cycle Length Verification: Ensure that the cyclic shifts in the matrix **H** are of the expected length [2].

$$\forall k, \ \text{Cycle length of } \mathbf{H}_k = n$$

If a shorter cycle is detected, flag the key as weak.

Symmetry Check: Ensure that the matrix **H** does not exhibit unintended symmetries.

$$\mathbf{H}_k \neq \mathbf{H}_{k+m} \text{ for } m \ll n$$

If symmetries are detested fact the key as weak.

6 Mathematical Notation Scheme for Weak Key Analysis in HQC

6.1 Introduction

Let n be the length of the binary vectors used in the HQC cryptosystem. The secret keys $\mathbf{x}$ and $\mathbf{y}$ are binary vectors with an expected Hamming weight w [3]. The public key matrix $\mathbf{H}$ is derived from these secret vectors and has a quasi-cyclic structure.

The goal of this scheme is to ensure that the keys $\mathbf{x}$ and $\mathbf{y}$ do not exhibit weaknesses that could compromise the security of the cryptosystem. Weak keys can be characterized by abnormal Hamming weights, insufficient randomness, structured patterns [4], or improper quasi-cyclic properties.

6.2 Key Properties to Analyse

Hamming Weight. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be the secret key vectors [5].

Hamming weight $w(\mathbf{x})$ of vector $\mathbf{x}$:

$$w(\mathbf{x}) = \sum_{i=1}^{n} x_i$$

Hamming weight $w(\mathbf{y})$ of vector $\mathbf{y}$:

$$w(\mathbf{y}) = \sum_{i=1}^{n} y_i$$

Expected Hamming weight w_{expected} :

$$w_{\text{expected}} \approx w$$

Acceptable range for Hamming weight:

$$w_{\min} \leq w(\mathbf{x}) \leq w_{\max}, w_{\min} \leq w(\mathbf{y}) \leq w_{\max}$$

Randomness. Randomness of $\mathbf{x}$ and $\mathbf{y}$ should be tested to ensure uniform distribution of 1 s and 0 s.

Chi-Square Test χ^2 for uniformity:

$$\chi^2(\mathbf{x}) = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j}, \chi^2(\mathbf{y}) = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j}$$

where O_j is the observed frequency and E_j is the expected frequency for each bin.

Monobit Test S for balance of 1 s and 0 s:

$$S(\mathbf{x}) = \left|\sum_{i=1}^{n} x_i - \frac{n}{2}\right|, S(\mathbf{y}) = \left|\sum_{i=1}^{n} y_i - \frac{n}{2}\right|$$

Runs Test for the frequency and length of consecutive bits:

$$R(\mathbf{x}) = \max(\text{length of consecutive 1 s or 0 s in } \mathbf{x})$$

$$R(\mathbf{y}) = \max(\text{ length of consecutive 1 s or 0 s in } \mathbf{y})$$

Structural Patterns. Pattern Detection in vectors **x** and **y** should identify repetitive or regular sequences that reduce entropy.

Fourier Transform $F(\mathbf{x})$ for detecting periodicities:

$$F(\mathbf{x}) = \sum_{k=1}^{n} x_k \cdot \mathrm{e}^{-2\pi ik/n}$$
$$F(\mathbf{y}) = \sum_{k=1}^{n} y_k \cdot \mathrm{e}^{-2\pi ik/n}$$

Pattern matching for regular intervals:

$$\text{Pattern}(\mathbf{x}) = \text{True if } \exists \text{ regular sequence in } \mathbf{x}$$

$$\text{Pattern}(\mathbf{y}) = \text{True if } \exists \text{ regular sequence in } \mathbf{y}.$$

Quasi-cyclic Structure. Quasi-Cyclic Matrix **H** derived from vectors **x** and **y** should maintain proper cyclic properties without short cycles or symmetries.

Matrix Construction H:

$$\mathbf{H} = \begin{pmatrix} \text{CircShift}(\mathbf{x}, 0) & \text{CircShift}(\mathbf{y}, 0) \\ \text{CircShift}(\mathbf{x}, 1) & \text{CircShift}(\mathbf{y}, 1) \\ \vdots & \vdots \\ \text{CircShift}(\mathbf{x}, n-1) & \text{CircShift}(\mathbf{y}, n-1) \end{pmatrix}$$

where $\text{CircShift}(\mathbf{v}, k)$ denotes the vector **v** cyclically shifted by k positions.

Cycle length verification:

$$\forall k, \mathbf{H}_{k+1} = \text{CircShift}(\mathbf{H}_k, 1)$$

Symmetry check:

$$\mathbf{H}_k \neq \mathbf{H}_{k+m} \text{ for any } m < n$$

6.3 Detection of Weak Keys

Hamming Weight Analysis

1. Calculate the Hamming weight of vectors **x** and **y**:

$$w(\mathbf{x}) = \sum_{i=1}^{n} x_i, w(\mathbf{y}) = \sum_{i=1}^{n} y_i$$

2. Check if the weights are within the acceptable range:

$$w_{\min} \leq w(\mathbf{x}) \leq w_{\max}, w_{\min} \leq w(\mathbf{y}) \leq w_{\max}$$

3. Flag as weak if $w(\mathbf{x})$ or $w(\mathbf{y})$ falls outsid ↓, is range.

Randomness Testing. *Perform the Chi-Square Test on* **x** *and* **y**.

$$\chi^2(\mathbf{x}) = \sum_{j=1}^{k} \frac{(O_j(\mathbf{x}) - E_j(\mathbf{x}))^2}{E_j(\mathbf{x})}$$
$$\chi^2(\mathbf{y}) = \sum_{j=1}^{k} \frac{(O_j(\mathbf{y}) - E_j(\mathbf{y}))^2}{E_j(\mathbf{y})}$$

Conduct the Monobit Test.

$$S(\mathbf{x}) = \left| \sum_{i=1}^{n} x_i - \frac{n}{2} \right|, S(\mathbf{y}) = \left| \sum_{i=1}^{n} y_i - \frac{n}{2} \right|$$

Run the Runs Test to check for consecutive bits.

$$R(\mathbf{x}) = \max(\text{length of consecutive 1 s or 0 s in } \mathbf{x})$$

$$R(\mathbf{y}) = \max(\text{length of consecutive 1 s or 0 s in } \mathbf{y})$$

Flag as weak if any test fails to indicate sufficient randomness.

Structural Pattern Detection

Apply the Fourier Transform to **x** and **y** to detect periodic patterns [6]:

$$F(\mathbf{x}) = \sum_{k=1}^{n} x_k \cdot e^{-2\pi ik/n}$$
$$F(\mathbf{y}) = \sum_{k=1}^{n} y_k \cdot e^{-2\pi ik/n}$$

Check for patterns using pattern matching techniques.
Flag as weak if any regular patterns are detected [7].
Quasi-Cyclic Structure Verification.
Verify the cyclic structure of matrix **H**:

$$\forall k, \mathbf{H}_{k+1} = \text{CircShift}(\mathbf{H}_k, 1)$$

Check for short cycles or unintended symmetries:

$$\mathbf{H}_k \neq \mathbf{H}_{k+m} \text{ for any } m < n$$

Flag as weak if improper cyclic properties are detected.

Mitigation of Weak Keys

Regenerate Keys

Regenerate the secret vectors **x** and **y** using a cryptographically secure random number generator (CSPRNG) if flagged as weak.

Ensure the new vectors meet the Hamming weight, randomness, pattern, and cyclic structure criteria [8]:

$$w_{\min} \leq w(\mathbf{x}) \leq w_{\max}, w_{\min} \leq w(\mathbf{y}) \leq w_{\max}$$

Adjust Hamming Weight

1. If the Hamming weight is slightly off, adjust by flipping bits:
   If $w(\mathbf{x}) < w_{\min}$, flip random 0 s to 1 s.
   If $w(\mathbf{x}) > w_{\max}$, flip random 1 s to 0 s [9].
2. Recalculate the Hamming weight after adjustments:

$$w(\mathbf{x}) = \sum_{i=1}^{n} x_i, w(\mathbf{y}) = \sum_{i=1}^{n} y_i$$

Introduce Random

1. Apply random bit flips to **x** and **y** to disrupt patterns:

$$\mathbf{x}' = \mathbf{x} \oplus \mathbf{r}, \mathbf{y}' = \mathbf{y} \oplus \mathbf{r}$$

Re-test for patterns and randomness

Enhance Cyclic Structure.

1. Regenerate the matrix $\mathbf{H}$ if improper cyclic structures are found.
2. Ensure the new matrix maintains the proper cyclic behavior:

$$\forall k, \mathbf{H}_{k+1} = \text{CircShift}(\mathbf{H}_k, 1)$$

References

1. Sendrier, N.: Decoding one out of many. In: International Workshop on Post Quantum Cryptography, p. 51 67. Springer, Berlin (2011). https://eprint.iacr.org/2011/367.pdf
2. Aguilar-Melchor, C., Blazy, O., Deneuville, J.-C., Gaborit, P., Zémor, G.: Efficient encryption from random quasi-cyclic codes. IEEE Trans. Actions Inf. Theor. **64**(5), 39273943 (2018)
3. Berlekamp, E.R., McEliece, R.J., van Tilborg, H.C.A.: On the inherent intractability of certain coding problems. IEEE Trans. Inf. Theor. **24**(3), 384386 (1978). http://authors.library.caltech.edu/5607/1/BERieeetit78.pdf
4. Berlekamp, E.: Algebraic coding theory. World Scientific (1968)
5. Dumer, I.: On minimum distance decoding of linear codes. In: Proceedings 5th Joint Soviet-Swedish International Workshop on Information Theory, p. 5052 (1991). https://www.researchgate.net/publication/296573348_On_minimum_distance_decoding_of_linear_codes
6. Gao, S., Mateer, T.: Additive fast Fourier transforms over finite fields. IEEE Trans. Inf. Theor. **56**(12), 62656272 (2010)
7. Lin, S., Costello, DJ (2004) Error control coding, vol. 2. Prentice Hall Englewood Cliffs (2004)
8. Mishra, N., Rodge, A., Deo, M., Agalwe, P., Bansode, N.: Road surveillance using real-time detection and alert system. In: 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), pp. 797–802. Salem, India (2024). https://doi.org/10.1109/ICPCSN62568.2024.00134
9. Sendrier, N.: Secure Sampling of Constant Weight Words Application to BIKE. Cryptology ePrint Archive, Report 2021/1631 (2021). https://eprint.iacr.org/2021/1631
10. Goldwasser, S., Micali, S.: Probabilistic encryption. J. Comput. Syst. Sci. **28**(2), 270299 (1984)
11. Guo, Q., Johansson, T., Löndahl, C.: A new algorithm for solving ring-LPN with a reducible polynomial. IEEE Trans. Inf. Theor. **61**(11), 6204–6212 (2015). https://arxiv.org/abs/1409.0472
12. Prange, E.: The use of information sets in decoding cyclic codes. IRE Trans. Inf. Theor. **8**(5), 59 (1962). http://ieeexplore.ieee.org/document/1057777/

A Survey on Hamming Quasi Cyclic (HQC)



Ajay S. Rodge and Bhupendra Singh

Abstract This paper presents an in-depth exploration of Hamming Quasi-Cyclic (HQC) codes, an algorithmic framework that has garnered significant attention within the domain of post-quantum cryptography. Acknowledged as a third-round finalist in the NIST competition, this cryptographic approach demonstrates promising capabilities in addressing the imminent challenges posed by quantum computing to contemporary cryptographic protocols. The paper begins by elucidating the foundational principles underlying HQC codes, tracing their evolution from traditional cyclic codes to the more specialized domain of quasi-cyclic structures. It subsequently delves into the distinctive characteristics and cryptographic properties exhibited by HQC Hamming Quasi-Cyclic codes, emphasizing their suitability for cryptographic applications in the post quantum era. A detailed examination of the construction methodologies employed in HQC code generation follows, encompassing algebraic, combinatorial, and iterative techniques. These methodologies are scrutinized not only for their efficacy in producing robust and secure codes but also for their compatibility with the stringent requirements of post quantum cryptographic frameworks. In conclusion, this survey paper serves as a comprehensive resource for researchers and practitioners seeking to comprehend the cryptographic significance of HQC Hamming Quasi-Cyclic codes. By elucidating their role as third-round finalists in the NIST competition and their potential impact on post-quantum cryptographic paradigms, this paper aims to catalyze further advancements in the field of quantum-resistant cryptography.

Keywords HQC · Post quantum cryptography · NIST · Reed–Soloman codes · Reed–Muller codes

A. S. Rodge (✉)
Computer Science and Data Science Department, GHRCEM, Pune, India
e-mail: Ajayrodge713@gmail.com

B. Singh
Defence Research and Development Organization-DRDO, Bangalore, India
e-mail: Bhupendra.cair@gov.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_33

1 Introduction

The Hamming Quasi-Cyclic (HQC) cryptographic scheme has become a significant focus within the realm of post-quantum cryptography. Designed to withstand the potential threats posed by quantum computers, HQC offers a robust framework for secure digital communications. Over the years, the HQC scheme has undergone numerous updates to enhance its security, performance, and compliance with evolving cryptographic standards. This survey aims to provide a comprehensive overview of these updates, tracing the evolution of HQC and highlighting the key modifications that have been made to address various challenges and vulnerabilities. Since its inception, HQC has seen several pivotal changes, such as the transition from the BCH-Repetition decoder to the RMRS decoder, improvements in the security proofs and definitions, and the incorporation of counter-measures against specific attacks. These updates not only reflect the ongoing efforts to strengthen the HQC scheme but also demonstrate the dynamic nature of cryptographic research and the need for continuous improvement in response to new findings and technological advancements. This survey is structured to systematically explore the history of updates to the HQC scheme. By documenting the rationale and impact of each update, we aim to provide a detailed and transparent account of HQC development. This will serve as a valuable resource for researchers, practitioners, and anyone interested in the field of post-quantum cryptography, offering insights into the current state of HQC and the journey it has undergone to achieve its present form. The subsequent sections will delve into the specifics of each update, starting from the earliest modifications to the most recent advancements. Through this detailed examination, we hope to shed light on the continuous efforts and innovations that underpin the HQC cryptographic scheme, ensuring its robustness and reliability in the face of evolving cryptographic challenges.

2 Scheme

2.1 HQC Public Key Encryption (HQC.PKE)

A decodable $[n, k]$ code C is derived by a generator matrix $G \in F^{k \times n}$. It enables efficient correction of up to δ errors using the decoding algorithm C. Decode().

A random double-circulant $[2n, n]$ code is characterized by its parity-check matrix $(1, h)$, where h is a randomly chosen element, forming the structural foundation of the code.

The composition of scheme is four polynomial-time algorithms:

1. Setup $\left(1^{\lambda}\right)$:

$$\text{paramaters} = (n, k, \delta, w, w_r, w_e)$$

where n, k, δ define the code structure, and w, w_r, w_e are parameters for the weight distributions used in the scheme.

Key Generation (KeyGen)

- Samples $h \leftarrow \mathbb{R}$.
- Generates the generator matrix $G \in \mathbb{F}^{k \times n}$ for C.
- Samples $(x, y) \leftarrow \mathbb{R}^2$ such that $\omega(x) = \omega(y) = w$.
- Sets $s = x + hy$.
- Sets the public key $\text{pk} = (h, s)$ and private key $\text{sk} = (x, y)$.
- Returns (pk, sk).

Encryption Process (Encrypt). Randomly generate $e \leftarrow R$ and $r = (r_1, r_2) \leftarrow R^2$, ensuring that the weight constraints $\omega(e) = w_e$ and $\omega(r_1) = \omega(r_2) = w_r$ are satisfied.

1. Compute:
   - $u = r_1 + hr_2$
   - $v = mG + s \cdot r_2 + e$

   The ciphertext is returned as $c = (u, v)$.

Decryption (Decrypt)

Returns C. Decode $(v - u \cdot y)$.

Correctness. The scheme's correctness is ensured by the decoding capability of the code C. In particular, if C. Decode successfully recovers $v - u \cdot y$, the decryption process correctly retrieves the original message, ensuring:

$$\text{Decrypt}(\text{sk}, \text{Encrypt}(\text{pk}, m)) = m.$$

The decoding succeeds whenever the conditions for C. Decode's error correction threshold are met.

$$\begin{aligned}
&\omega(s \cdot r_2 - u \cdot y + e) \leq \delta \\
&\omega((x + h \cdot y) \cdot r_2 - (r_1 + h \cdot r_2) \cdot y + e) \leq \delta \\
&\omega(x \cdot r_2 - r_1 \cdot y + e) \leq \delta
\end{aligned}$$

A maximum limit for the likelihood of decryption failure is established through examining the distribution of the vector $e' = x \cdot r_2 - r_1 \cdot y + e$.

2.2 HQC Key Encapsulation Mechanism

HQC.KEM is derived from HQC.PKE and includes additional steps to achieve IND-CCA2 security. The KEM-DEM version uses hash functions G, H, and K (typically SHA512).

Setup(1^{λ}): Similar to the HQC.PKE setup but with the symmetric key length k (typically 256).

Key Generation (KeyGen):

Encapsulate (pk):

1. Generate $m \in \mathbb{F}^k$ to act as a seed for the shared key.
2. Use m to derive randomness $G(m)$.
3. Compute the ciphertext $c = (u, v) = \text{Encrypt}(pk, m)$.
4. Symmetric key K as $K = K(m, c)$.
5. Calculate $d = H(m)$ and output (c, d).

Decapsulate (sk, c, d):

1. Recover m by decrypting c with Decrypt(sk, c).
2. Recompute the ciphertext $c' = \text{Encrypt}(pk, m)$.
3. Verify $c' = c$ and $d = H(m)$; abort if either check fails.
4. Derive symmetric key $K = K(m, c)$.

3 Example

During the last round of the NIST, HQC (Hamming Quasi-Cyclic) undergoes assessment with revised parameters and security guarantees. The following illustrates HQC public key encryption (PKE) in accordance with the fourth round's guidelines and protocols:

3.1 Parameters and Setup

1. Code Length $\boldsymbol{n}$.
2. Message Length k.
3. Error Vector Length t.
4. Field: Binary field $\mathbb{F}_2$.

Let's use hypothetical values for simplicity:

- $n = 10$.
- $k = 5$.
- $t = 3$.

Key Generation

1. Secret Key:

- Choose random binary vectors **x** and **y** of length $\boldsymbol{n}$:

$$x = [1, 0, 1, 0, 1, 0, 1, 0, 1, 0]$$
$$y = [0, 1, 0, 1, 0, 1, 0, 1, 0, 1]$$

2. Public Key:

- Generate a quasi-cyclic code generator matrix $\boldsymbol{H}$:

$$H = \begin{bmatrix} 1\,0\,0\,1\,0\,1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,1 \\ 0\,0\,1\,0\,0\,1\,0\,1\,0\,0 \\ 1\,0\,0\,1\,0\,1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,1 \end{bmatrix}$$

- Choose a random error vector **e** of length $\boldsymbol{n}$:

$$e = [1, 1, 0, 0, 1, 0, 1, 0, 1, 1]$$

- Compute $\mathbf{h} = \boldsymbol{H} \cdot \mathbf{x} + \mathbf{e}$:

$$H \cdot \mathbf{x} = \begin{bmatrix} 1\,0\,0\,1\,0\,1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,1 \\ 0\,0\,1\,0\,0\,1\,0\,1\,0\,0 \\ 1\,0\,0\,1\,0\,1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 1 \\ 2 \\ 3 \\ 1 \end{bmatrix} \text{mod2} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \mathbf{h} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

So, the public key $pk = (H, \mathbf{h})$ and the secret key $sk = (\mathbf{x}, \mathbf{y})$.

Encryption

1. Message $\mathbf{m} = [1, 0, 1, 0, 1]$.
2. Encode message using public key:

$$\mathbf{c} = \boldsymbol{H} \cdot \mathbf{m} = \begin{bmatrix} 1\,0\,0\,1\,0\,1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,1 \\ 0\,0\,1\,0\,0\,1\,0\,1\,0\,0 \\ 1\,0\,0\,1\,0\,1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1\,0\,0\,1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \\ 2 \end{bmatrix} \text{mod2} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

3. Add random error vector $\mathbf{e}' = [1, 1, 0, 1, 1]$:

$$\mathbf{ct} = \mathbf{c} + \mathbf{e}' = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The ciphertext is ct = [1, 1, 1, 1, 1].

Decryption

1. Received ciphertext ct = [1, 1, 1, 1, 1]
2. Decode using the secret key:

$$\mathbf{m}' = \mathbf{ct} - \mathbf{e}' = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

3. Retrieve the original message:

$$\mathbf{m} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

4 First Round

Introduction to HQC (Hamming Quasi-Cyclic). The HQC (Hamming Quasi-Cyclic) scheme is a post-quantum cryptographic algorithm that aims to provide security resistant to quantum attacks. It is a code-based public key encryption scheme leveraging the hardness of decoding random quasi-cyclic codes. The security of HQC is based on the hardness of the Quasi-Cyclic Syndrome Decoding (QCSD) problem, which remains difficult for both classical and quantum computers.

Components of HQC

1. Public Key (pk):

   A quasi-cyclic code generator matrix H.
   A vector $\mathbf{h}$ derived from the secret key and an error vector.

2. Secret Key (sk):

   A pair of vectors $\mathbf{x}$ and $\mathbf{y}$.

Scheme Description Key Generation

1. Choose secret key:

   Select two random binary vectors $\mathbf{x}$ and $\mathbf{y}$ of length $\boldsymbol{n}$.

2. Compute public key:

   Generate a quasi-cyclic code generator matrix H.
   Choose a random error vector $\mathbf{e}$.
   Compute $\mathbf{h} = \boldsymbol{H} \cdot \mathbf{x} + \mathbf{e}$.

3. Output:

   Public key $pk = (H, \mathbf{h})$.
   Secret key $sk = (\mathbf{x}, \mathbf{y})$.

Encryption

1. Message encoding:

   Convert the message $\mathbf{m}$ into a binary vector of length k.

2. Compute ciphertext:

   Encode the message using the public key matrix H.
   Add a random error vector $\mathbf{e}'$ to the encoded message.

5 Second Round

5.1 Theoretical Improvements

Improvement 1. A refined analysis of the error distribution model leads to a lower Decryption Failure Rate (DFR). Public Key Size Reduction: This optimization reduces the public key size by 3%, bringing it to 3024 bytes for 128-bit security.

Improvement 2. A novel decoding approach integrates Reed-Muller and Reed-Solomon codes. Impact on Key Size: This technique reduces the public key size by 17% , lowering it to 2607 bytes for 128-bit security A novel decoding approach integrates Reed-Muller and Reed-Solomon codes.

New Parameters. A new parameter set, HQC-RMRS, has been introduced for 128-bit, 192-bit, and 256bit security levels.

5.2 Modifications in Round 1 and Round 2

Scheme Modification. The adoption of tensor product codes (BCH the and repetition) removes the requirement for the code length to be prime.

- The new code length is defined as $n_1 \times n_2$, where:
- n_1 is the BCH code length.
- n_2 is the repetition code length.
- To prevent algebraic attacks via polynomial factorization, a primitive prime n slightly larger than $n_1 \times n_2$ is selected.
- Redundant bits are truncated to optimize performance.

Parameter Refinement. Parameters with a Decoding Failure Rate (DFR) exceeding 2^{-128} were discarded to meet the required security threshold.

6 Third Round

6.1 Decoder Update

Scheme Modification. The scheme was updated to incorporate tensor product of the codes, specifically combining BCH and repetition codes, eliminating the requirement for the code length to be a prime number. The updated code length is defined as $n_1 \times n_2$, where n_1 represents the BCH code length and n_2 represents the repetition code length. To counter algebraic attacks based on polynomial factorization, a primitive prime n

slightly larger than $n_1 \times n_2$ was chosen. Any excess bits introduced by this adjustment were truncated, and the security proof was revised to accommodate these changes.

Parameter Refinement. During parameter selection, configurations with a Decoding Failure Rate (DFR) exceeding 2^{-128} were excluded to ensure robust security. This refinement minimizes decryption errors to an exceptionally low probability, aligning the scheme with stringent security standards. By optimizing parameters to achieve this negligible DFR, the scheme's reliability is reinforced without compromising performance.

Multi-Ciphertext Attack Mitigation. Researchers achieved a significant advancement in the Reed-Muller decoder's theoretical lower limit, bringing it closer to optimal performance. This enhancement reduced the theoretical DFR limit, allowing for further optimization of the scheme's parameters. These improvements bolster resistance to multi-ciphertext attacks, enhancing overall security.

$$\theta = \text{SHAKE256} - 512(m \| pk \| \text{ salt })$$

Counter-Measure to Timing Attack. Countermeasure Against Timing Attack: To address a timing attack highlighted in the reference, the implementation has been updated to use Barrett reduction instead of the modulo operator. This modification enhances security by minimizing vulnerability to timing-based side-channel attacks.

IND-CPA Security Proof Update

The HQC scheme now incorporates the HHK transform with implicit rejection strengthening its security framework.
A rigorous IND-CCA2 security proof has been developed within the HHK framework to formally validate this modification.
The implementation has been updated accordingly with Known Answer Tests KATs adjusted to ensure consistency with the revised scheme.
Security Analysis of Nonuniform Sampling
An in-depth analysis was conducted to assess the impact of nonuniform sampling of small weight vectors.
Findings indicate that while the sampling deviates slightly from perfect uniformity, it has a negligible effect on the IND-CCA2 security of the HQC scheme.

7 Fourth Round

7.1 Security Enhancements

Integration of HHK Transform and IND-CCA2 Security Proof.
The HHK transform with implicit rejection has been integrated into the HQC scheme.

A formal IND-CCA2 security proof was developed within the HHK framework to validate this modification.
The implementation has been updated accordingly, and Known Answer Tests (KATs) have been adjusted to align with the revised scheme.

7.2 Security Analysis for Nonuniform Sampling

A detailed analysis was conducted to assess the impact of nonuniform sampling of small weight vectors. The findings confirm that sampling in a nonuniform but nearly uniform manner has negligible impact on the IND-CCA2 security of the HQC scheme.

References

1. Aguilar-Melchor, C., Blazy, O., Deneuville, J.-C., Gaborit, P., Zémor, G.: Effcient encryption from random quasi-cyclic codes. IEEE Trans. Actions Inf. Theor. **64**(5), 39273943 (2018)
2. Applebaum, B., Ishai, Y., Kushilevitz, E.: Cryptography with constant input locality. In: Menezes, A. (ed.) CRYPTO 2007, vol. 4622 of LNCS, p. 92110. Springer, Heidelberg (2007)
3. Aragon, N., Gaborit, P., Zémor, G.: HQC-RMRS, an instantiation of the HQC encryption framework with a more e cient auxiliary error-correcting code. https://arxiv.org/abs/2005.10741
4. Becker, A., Joux, A., May, A., Meurer, A.: Decoding random binary linear codes in 2n/20: How 1 + 1 = 0 improves information set decoding. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012, vol. 7237 of LNCS, p 520536. Springer, Heidelberg (2012)
5. Berlekamp, E.R., McEliece, R.J., van Tilborg, H.C.A.: On the in herent intractability of certain coding problems. IEEE Trans. Inf. Theor **24**(3), 384386 (1978). http://authors.library.caltech.edu/5607/1/BERieeetit78.pdf
6. Bernstein, D.J.: Grover vs. McEliece. In: Post-Quantum Cryptography, p. 7380. Springer, Berlin (2010). https://cr.yp.to/codes/grovercode-20091123.pdf
7. Bernstein, D.J., Lange, T., Peters, C.: Attacking and defending the mceliece cryptosystem. In: Post-Quantum Cryptography, p. 3146. Springer, Berlin (2008). https://cr.yp.to/codes/mceliece-20080807.pdf
8. Torres, R.C., Sendrier, N.: Analysis of information set decoding for a sub-linear error weight. In: Takagi, T. (ed.) Post-Quantum Cryptography—7th International Workshop, PQCrypto 2016, Fukuoka, Japan, Feb 24–26, 2016, Proceedings, vol. 9606 of Lecture Notes in Computer Science, p. 144161. Springer, Berlin (2016). https://hal.inria.fr/hal-01244886
9. Dumer, I.: On minimum distance decoding of linear codes. In: proceedings of the 5th joint Soviet-Swedish International Workshop on Information Theory, p. 5052 (1991). https://www.researchgate.net/publication/296573348_On_minimum_distance_decoding_of_linear_codes
10. Finiasz, M., Sendrier, N.: Security bounds for the design of code-based cryptosystems. In: Matsui, M. (ed.) ASIACRYPT 2009, vol. 5912 of LNCS, p. 88105. Springer, Heidelberg (2009)
11. Gaborit, P.: Shorter keys for code based cryptography. In: Proceedings of the 2005 International Workshop on Coding and Cryptography (WCC 2005), p. 8191 (2005). http://www.unilim.fr/pages_perso/philippe.gaborit/shortIC.ps
12. Gaborit, P., Girault, M.: Lightweight code-based identi cation and sig nature. In: 2007 IEEE International Symposium on Information Theory, p. 191–195. IEEE, USA (2007). https://www.unilim.fr/pages_perso/philippe.gaborit/isit_short_rev.pdf

13. Gao, S., Mateer, T.: Additive fast fourier transforms over nite elds. IEEE Trans. Inf. Theor. **56**(12), 62656272 (2010)
14. Gligoroski, D.: PQC forum, official comment on bike submission. NIST PQC forum (2017). https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/round-1/official-comments/BIKE-official-comment.pdf
15. Goldwasser, S., Micali, S.: Probabilistic encryption. J. Comput. Syst. Sci. **28**(2), 270299 (1984)
16. Guo, Q., Johansson, T., Löndahl, C.: A new algorithm for solving ring-LPN with a reducible polynomial. IEEE Trans. Inf. Theor **61**(11), 6204–6212 (2015). https://arxiv.org/abs/1409.0472
17. Hofheinz, D., Hövelmanns, K., Kiltz, E.: A modular analysis of the Fujisaki-Okamoto transformation. In: Theory of Cryptography Conference, p. 341–371. Springer, Berlin (2017)
18. Human, W.C., Pless, V.: Fundamentals of error-correcting Cambridge university press (2010). https://www.amazon.fr/Fundamentals-Error-Correcting-Codes-Cary-Huffman/dp/0521131707
19. Lin, S., Costello, D.J: Error control coding, vol. 2. Prentice Hall Englewood Cliffs (2004)
20. Liu, Z., Pan, Y.: PQC forum, official comment on HQC submission. NIST PQC forum (2018). https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/round-1/official-comments/HQC-official-comment.pdf
21. Liu, Z., Pan, Y., Xie, T.: Breaking the hardness assumption and IND-CPA security of HQC submitted to NIST PQC project. In: International Conference on Cryptology and Network Security, p. 344356. Springer, Berlin (2018)
22. Liu, Z., Pan, Y., Xie, T.: Breaking the hardness assumption and IND-CPA security of HQC submitted to NIST PQC project. In: International Conference on Cryptology and Network Security, p. 344356. Springer, Berlin (2018)
23. MacWilliams, F.J., Sloane, N.J.A.: The theory of error correcting codes, vol. 16. Elsevier, Netherlands (1977)
24. May, A., Meurer, A., Thomae, E.: Decoding random linear codes in $\tilde{O}(20.054n)$. In Asiacrypt, vol. 7073, p. 107124. Springer, Berlin (2011). https://doi.org/10.1007/978-3-642-25385-0_6
25. May, A., Ozerov, I.: On computing nearest neighbors with applications to decoding of binary linear codes. In: EUROCRYPT, vol. 1, p. 203228 (2015). http://www.cits.rub.de/imperia/md/content/may/paper/codes.pdf
26. Misoczki, R., Tillich, J.-P., Sendrier, N., Barreto, P.: MDPC McEliece: new McEliece variants from moderate density parity-check codes. In: Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on, p. 20692073. IEEE, USA (2013). https://eprint.iacr.org/2012/409.pdf
27. Prange, E.: The use of information sets in decoding cyclic codes. IRE Trans. Inf. Theor. **8**(5), 59 (1962). http://ieeexplore.ieee.org/document/1057777/
28. Schröder, R.L., Gast, S., Guo, Q.: Divide and surrender: Exploiting Variable Division Instruction Timing in HQC Key Recovery Attacks. Cryptology ePrint Archive, Paper 2024/299 (2024). https://eprint.iacr.org/2024/299
29. Sendrier, N.: Decoding one out of many. In: International Workshop on Post Quantum Cryptography, p. 51–67. Springer, Berlin (2011). https://eprint.iacr.org/2011/367.pdf
30. Sendrier, N.: Secure Sampling of Constant Weight Words Application to BIKE. Cryptology ePrint Archive, Report 2021/1631 (2021). https://eprint.iacr.org/2021/1631
31. Shoup, V.: NTL: A Library for Doing Number Theory (2001). http://www.shoup.net/ntl
32. Berlekamp, E.: Algebraic Coding Theory. World Scientific (1968)
33. Mishra, N., Rodge, A., Deo, M., Agalwe, P., Bansode, N.: Road surveillance using real-time detection and alert system. In: 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, pp. 797–802 (2024). https://doi.org/10.1109/ICPCSN62568.2024.00134
34. Stern, J.: A method for nding codewords of small weight. In: International Colloquium on Coding Theory and Applications, p. 106113. Springer, Berlin (1988). https://doi.org/10.1007/BFb0019850

Enhancing Phishing Detection: The Impact of Visual Warning Designs and User Engagement in Email Clients



Sandeep Malipeddi

Abstract Phishing attacks remain a significant cybersecurity threat, exploiting user vulnerabilities through deceptive email tactics. This study investigates the effectiveness of various visual warning designs in email clients, focusing on user engagement and behavioral responses. Using eye tracking technology and qualitative feedback, I explored the impact of warning placement, clarity, and interactivity on user attention and trust. Findings reveal that context-rich warnings integrated directly within email content enhance user engagement, prompt immediate responses, and serve an educational function, empowering users to recognize future threats independently. However, challenges such as alert fatigue, ecological validity, and participant diversity highlight areas for further research. Expanding studies to real-world settings and incorporating diverse user groups can improve the robustness and applicability of findings. The study underscores the potential of advanced technologies, including artificial intelligence and adaptive phishing detection algorithms, to create more personalized and effective warning systems. As cyber threats evolve, this research emphasizes the importance of continuously refining phishing warnings to protect digital interactions and enhance overall cybersecurity resilience.

Keywords Phishing attacks · EMail security · User management · Visual warning design · Eye tracking · Cybersecurity awareness

1 Introduction

In today's digital age, the internet is integral to daily activities such as communication, transactions, and accessing information. Yet, this increased dependency brings heightened security risks, especially from phishing attacks. Phishing, a prevalent cyber threat, involves deceiving individuals into revealing sensitive information by masquerading as a trustworthy entity in digital communications. There is a vast

S. Malipeddi (✉)
Hood College, Frederick, MD, USA
e-mail: sandeepmalipedii8@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_34

amount of phishing vectors, such as SMS, social media or email, with the latter being the most common medium for phishing attacks [1]. These attacks are used to direct victims to fraudulent websites where they are coerced into providing personal and financial details, allowing attackers unauthorized access to accounts [2].

The impact of phishing extends beyond individual victims to organizations worldwide, causing significant financial losses and eroding trust in digital systems. In 2023, the Anti-Phishing Working Group (AWPG) observed approximately 4.9 Million phishing attacks [3]. Despite the deployment of various, anti-phishing technologies, the human factor remains a critical vulnerability. The effectiveness of phishing warnings, therefore, is paramount in the cybersecurity defense arsenal. These warnings, if designed and implemented effectively, can play a crucial role in educating users and averting the success of phishing attacks.

In response to the escalating threat of phishing attacks, numerous efforts have been undertaken by cybersecurity experts and organizations to mitigate the risks. These include the development of algorithms for threat detection [4] and the implementation of user awareness programs [5]. Users' ability to detect and respond to phishing attempts is critical, prompting a shift toward developing more user-centered security solutions. These solutions emphasize visual warnings that act as direct interventions, alerting users to potential threats. However, the effectiveness of such warnings hinges on several factors that this study aims to explore and understand.

For my study, I developed mock-up visual warnings to simulate real alerts in an email client and expanded their functionality with a custom plugin for Mozilla Thunderbird, a widely-used open-source email client chosen for its adaptability. This plugin enhances the visual presentation of phishing warnings, utilizing common design principles to capture users' attention more effectively.

We created a unique set of emails incorporating both regular communications and mock phishing scenarios to simulate a realistic email environment. Using eye tracking technology and detailed user feedback, I explored how users perceive, understand, and react to various visual warning strategies. This investigation offers insights into the design of more effective phishing warnings and highlights the importance of user interface design in cybersecurity. By focusing on user centric design elements, my study aims to improve how visual cues are used to strengthen defenses against phishing attacks. My research is structured around the following research questions: user interface design in cybersecurity. By focusing on user centric design elements, my study aims to improve how visual cues are used to strengthen defenses against phishing attacks. My research is structured around the following research questions:

- RQ1: How do users' perceptions differ in response to various phishing warnings in email interfaces, and which type is perceived as the most effective? *Here* "*effective*" *refers to the ability of the warnings to increase user awareness, prompt appropriate user actions to mitigate risk (e.g., not clicking on malicious links), and enhance recognition of phishing threats in future encounters.*
  - How do users perceive the helpfulness of each warning type?
  - Which warning design do users prefer and why?

Table 1 Acronyms used in this paper

Acronym	Definition
AOI	Area of Interest
QDA	Qualitative Data Analysis
ATI-S	Affinity for Technology Interaction Short Scale
TOI	Time of Interest

- RQ2: What common gaze patterns emerge among users when scanning email content, and how can these patterns help with the strategic placement of phishing warnings to maximize user attention?

To address these research question I have laid out the following objectives for my study:

1. *Evaluating Visual Warning Strategies*: I aim to examine the effectiveness of various visual elements (e.g. banner placement, animations) in warning designs in terms of their immediate recognizability and understandability (Table 1).
2. *Analyzing Attention Distribution*: Using eye tracking technology, I plan to identify which visual elements attract users' attention and how these elements influence perception and behaviour.
3. *Assessing User Reactions*: Through interviews with participants, I intend to capture their subjective evaluations and opinions regarding the different warning types. This includes exploring user trust and their perception on the warnings helpfulness.
4. *Establishing Best Practices*: Based on the collected data and user feedback, I seek to identify design principles that maximize the effectiveness of phishing warnings.

In summary, my research reveals that effectively designed phishing warnings can significantly influence users' abilities to recognize and respond to phishing attempts. Eye tracking data confirmed that warnings integrated within the email body are noticed and acted upon more swiftly than those placed peripherally or appearing with a delay. Qualitative feedback underscored the importance of clear, contextual information within these warnings, enhancing users' understanding of potential threats.

The implications of these results are substantial, advocating for a shift in how cybersecurity defenses are structured. As phishing tactics evolve, the need for warnings that seamlessly integrate into users' workflows and effectively communicate risks becomes increasingly critical. By focusing on the development of user friendly and informative warning systems, I can better equip individuals to combat phishing threats, potentially reducing the success rate of such attacks across various platforms.

2 Background and Related Work

In the dynamic field of cybersecurity, understanding the multifaceted nature of phishing attacks is essential for developing effective defensive strategies. This section provides an overview of the key areas of research that inform my study on phishing warnings. By exploring the psychological underpinnings of social engineering, the evolution of phishing detection technologies, and the critical role of user education, I aim to contextualize my research within the broader discourse on combating phishing. Each of these areas offers vital insights that directly influence the design and implementation of phishing warnings, ensuring they are not only technically sound but also psychologically effective.

3 Phishing Definition and Techniques

Phishing is a deceptive practice where cybercriminals use fraudulent communication, typically via email, to trick individuals into revealing personal information, such as passwords and credit card numbers. This cyber threat exploits the trust of users through cleverly disguised messages that mimic legitimate sources. Understanding the array of phishing techniques is essential for developing effective countermeasures and for training users to recognize and avoid these threats.

As detailed by Alkhalil et al. [6], phishing attacks employ a combination of technical and psychological tactics to extract sensitive information from victims. These attacks are not uniform but vary significantly in their approach and execution. Alkhalil et al. identify several prevalent phishing techniques:

- *Email Phishing*: The primary method where attackers send fraudulent emails mimicking legitimate sources to steal personal information. Visual warnings for this technique should clearly differentiate genuine communications from phishing attempts, for instance by highlighting suspicious elements directly within the email interface.
- *Spear Phishing and Whaling*: Highly targeted forms that use personalized information to increase the attack's credibility.
- *Vishing (Voice Phishing) and SMiShing (SMS Phishing)*: Phishing via phone calls and SMS that exploit the user's trust in mobile communications.
- *Website Forgery*: Creating counterfeit websites that replicate genuine ones to capture personal data.
- *Social Media Phishing*: Using social platforms to trick users into revealing sensitive data through deceptive posts or messages.

This classification underscores the importance of tailored defensive strategies that address the specific vectors and tactics of each phishing type, enhancing both technological defenses and user training programs.

4 Technical Detection Strategies

Phishing detection technologies are essential tools, designed to identify and mitigate phishing attempts across various digital platforms. Among these technologies, email client spam filters, with phishing emails being a subset of spam mails, represent a fundamental line of defense, filtering out potential phishing emails based on specific criteria such as sender reputation [7] or message content [8]. In their paper, Bhomwick et al. [8] delve into the evolution of machine learning techniques in the realm of email spam filtering. They chart the journey from initial heuristic-based approaches to the adoption of sophisticated machine learning models, capable of dynamically adapting to the continually evolving strategies employed by spammers. The paper not only highlights the effectiveness of these techniques but also points towards future directions for research in improving spam detection mechanisms.

Another approach in phishing detection is explored by Siddiq et al. [9], utilizing deep learning techniques. Their study, employing neural networks, showcases the potential of deep learning in accurately identifying phishing websites. This highlights deep learning's pivotal role in cybersecurity, especially in distinguishing malicious content, and emphasizes its contribution to advancing phishing detection systems.

While technological solutions are crucial, their effectiveness ultimately depends on the users' ability to understand and respond to the warnings they generate. This interdependency brings my research into focus, aiming to optimize visual warnings to ensure they not only capture attention but also convey urgency effectively, prompting appropriate user actions in the face of potential threats.

5 User Education and Training

Understanding the mechanisms behind phishing attacks and the reasons users fall prey to them is essential in developing effective countermeasures. A key vulnerability is the general lack of cybersecurity awareness among users, rendering them susceptible to manipulation by attackers. In their paper Dhamija et al. [10] highlight this issue, pointing out the critical need for enhanced user understanding to mitigate vulnerability to phishing shemes.

To adress this gap, various training methods have been explored, aiming to improve user ability to recognize and resist phishing attempts. Research by Sheng et al. [11] showcases an innovative approach in phishing education through the 'Anti-Phishing Phil' game. This research demonstrates that integrating educational content with interactive gameplay can significantly increase the efficacy of cybersecurity training. Such engaging methods not only educate but also actively involve users in the learning process, making the lessons more memorable and practical.

Further emphasizing the importance of user education, Kumaraguru et al. [12] investigated the impact of training on users' abilites to distinguish between legitimate and fraudulent websites. Their findings underscore the effectiveness of well-crafted

educational programs in empowering users against phishing threats. Additionally, Wash's article [13] on how IT experts detect phishing emails provides a unique perspective on the cognitive strategies employed by experienced professionals. By interviewing IT experts, Wash explains a three-stage process these professionals use: sensemaking, suspicion, and decision-making, when evaluating the legitimacy of an email. This approach underscores the depth of expertise and vigilance required to effectively identify phishing attempts, high-lighting the potential benefits of training programs.

Integrating these insights, my study leverages the principles of effective education and cognitive engagement in designing visual warnings that are not only noticeable but also informative. By enhancing user understanding of what phishing attempts look like, I aim to elevate the baseline of cybersecurity awareness.

6 Social Engineering in Phishing

Social engineering is described as an attack type where the assailant exploits human vulnerabilities through social interaction to compromise cybersecurity goals like confidentiality, integrity, and availability of cyber elements. Unlike traditional attacks that may require overcoming technical barriers, social engineering attacks manipulate human psychology to bypass security measures. It's particularly challenging to defend against because it preys on innate human traits, which can often be turned into security vulnerabilities by skilled attackers [14].

Highlighting persuasive email techniques, Koddebusch [15] focuses on the specific persuasive techniques employed in phishing emails, such as invoking authority or urgency, which effectively manipulate users' decision-making processes. This manipulation illustrates how cyber attackers exploit psychological triggers to induce actions that users might otherwise consider risky or unnecessary. Koddebusch's findings provide empirical evidence that these persuasion strategies significantly increase the success rate of phishing attacks, underscoring the critical need for awareness and training that can help individuals recognize and resist such tactics.

In their research, Abdrabou et al. [16] investigate how phishing emails affect user behavior by tracking their gaze and mouse movements during an email sorting task. The study aims to understand if these physiological and behavioral cues can serve as indicators of exposure to phishing attacks, which are often designed to manipulate users into revealing sensitive information. Their findings suggest that subtle changes in behavior, such as eye movements or mouse dynamics, could potentially be used to identify when users are interacting with phishing content.

Distler's study [17] examines the impact of various contextual factors on individuals' responses to phishing within an organizational setting. Using an in-situ deception methodology with 14 participants in their real work environments, the study reveals that contexts significantly influence the likelihood of falling for phishing emails. Distler's findings underscore the importance of contextualized training and

proactive organizational policies to enhance cybersecurity. The research also underscores the role of language and writing style in the detection of phishing mails, revealing that unexpected variations in these elements can alert recipients to potential phishing attempts.

Building on these findings, my study introduces and evaluates targeted phishing warnings that focus on anomalies in greeting patterns and email signatures—common areas where deviations from normative language use might occur. By flagging unusual language patterns that recipients have learned to associate with phishing these warnings aim to enhance the detection capabilities of email users, potentially increasing their vigilance and reducing the likelihood of successful phishing attacks.

7 Methodology

7.1 Research Design

Our study adopts a mixed-methods approach, integrating eye tracking data and qualitative interviews to examine how users perceive and interact with phishing warnings in email interfaces. The quantitative component makes use of eye tracking to objectively measure user attention and gaze patterns across various warning designs. The qualitative segment, through semi-structured interviews, investigates participant perspectives, enriching the data with personal insights and preferences. This dual-faceted approach aims to reveal not only immediate visual engagements but also deeper user experiences.

7.2 Research Setting

Our research was conducted in a controlled lab environment at the University of the Bundeswehr. Upon arrival, participants were welcomed and escorted to a dedicated workstation, which was arranged to facilitate the study's requirements. This workstation was equipped with a computer system that had the email client, Mozilla Thunderbird, preloaded and configured for immediate use. An eye tracking device was connected to the computer to record and analyze participants' gaze patterns as they interacted with the email client. The setting was optimized to mimic a typical user experience while ensuring precise data capture for the study's analysis.

7.3 Recruitment of Participants

For the recruitment of participants for this study, I utilized multiple communication channels within the informatics community of the University. The study was advertised in group chats on popular platforms such as Discord, Whatsapp and Slack. Additionally, I utilized the LMU newsletter to reach a broader audience within the university setting. There were no specific prerequisites for participants, allowing for a wide demographic range. The study was open to all individuals aged 18 and above. Participants received a 10 EUR compensation for their participation.

8 Material

8.1 Technical Equipment and Tools

Central to this study is the integration of eye tracking technology, specifically utilizing the Tobii Pro Spark eye tracker in conjunction with the Tobii Pro Lab software. This setup plays a crucial role in objectively capturing where participants' attention is focused as they interact with email content and phishing warnings. By recording gaze patterns I gain valuable insights into the reactions to different warning designs. The Tobii Pro Lab software allows me the seamless recording of the participants' screen alongside their gaze movements and patterns, simplifying a comprehensive analysis of the captured data.

Audio recording was conducted using a Google Pixel smartphone, which features built-in auto transcription software, providing me with a reliable method for capturing and transcribing participant feedback during the study. Each transcript was additionally manually reviewed to ensure the accuracy of the data and to filter out irrelevant or personal information.

8.2 Emails

To establish a genuine email environment for the study, I created a suite of 30 emails to mirror the typical inbox of a professional workspace. This collection included routine business correspondences, such as project updates, meeting invitations, and newsletters, mixed with 9 simulated phishing emails. The inclusion of one false negative, a phishing email not recognized as such, further emulates the realistic challenges users face. These simulations employed prevalent phishing strategies observed in actual attacks, such as soliciting confidential information, fraudulent alerts, and deceptive links to malicious websites (Fig. 1).

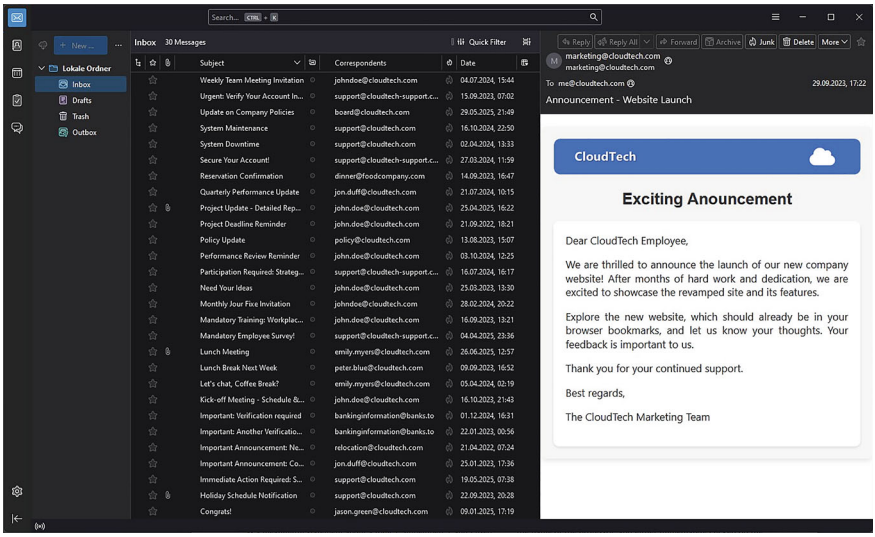


Fig. 1 Participants' environment in Mozilla Thunderbird

9 Procedure

9.1 Consent, Demographics and ATI-S

Before beginning with the study tasks, participants are required to sign a consent form that includes permissions for screen and audio recording as well as eye tracking. Following consent, they are asked to provide demographic information such as age, gender, educational background, and professional experience. Additionally, participants fill out the short version of the Affinity for Technology Interaction Scale (ATI-S).[1] This scale assesses individuals' comfort levels and predispositions towards technology, which helps to illuminate how these factors may influence their perceptions and interactions with phishing warnings (Fig. 2).

We opted to use the short version of the ATI scale due to its ability to efficiently measure technological affinity without imposing extensive questionnaire fatigue on participants and maintaining high engagement levels.

9.2 Briefing and Setup

Participants receive a briefing on the eye tracking process and a brief overview of Mozilla Thunderbird if they are not already familiar with its interface. The

[1] https://ati-scale.org/.

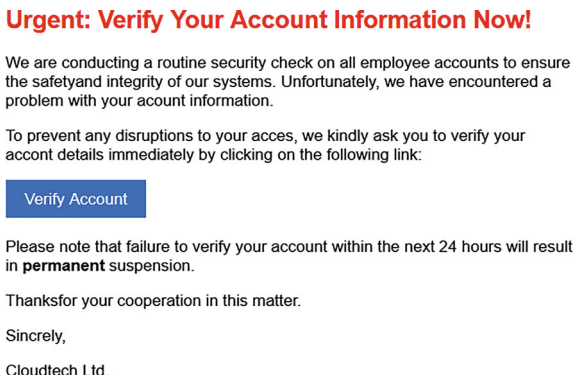


Fig. 2 An example for a mock phishing email

eye tracking calibration is conducted next to ensure accurate data capture. Audio recording is also enabled to capture the participants' verbal feedback throughout the study, facilitating later analysis of their qualitative responses. Participants are informed that they are welcome to ask questions at any time during the study to clarify any doubts or concerns they may have.

9.3 Phase One: Sorting Task and Collection of Eye tracking Data

Phase One serves as the quantitative backbone of my study, utilizing eye tracking technology to gather gaze data as participants engage with a series of emails. This phase is designed to simulate a real-world working environment to objectively assess how participants interact with my phishing warnings in a typical workflow.

Task Setup and Environment

Participants are cast in the role of an employee at the fictional company "CloudTech". This setup is intended to create a realistic email environment. Upon commencement they are provided with a folder containing a mix of emails, including routine communications from fictional colleagues and company announcements. Within these emails there are several phishing attempts, though participants are not made aware of these inclusions to avoid biasing their responses.

Instructions and Procedure

To simplify the email management task and ensure clarity, participants are instructed to discard any email into the trash folder that they suspect or deem unnecessary, without the need to further categorize or mark emails. This is designed to streamline the sorting process, allowing participants to focus more on the content and less on organizational tasks. During the sorting process, participants are encouraged to

verbalize their thoughts and decisions. This "think aloud" method is crucial for understanding the decision-making process behind their actions.

Observer Role and Time Allocation

The observer's role during this phase is to listen and monitor the explanations provided by the participants, intervening minimally (e.g. when questions arise) to avoid influencing their natural behavior. Each participant is allocated approximately 10 min to complete the task.

9.4 Phase Two: Gathering Qualitative Feedback

Following the sorting task, participants undergo a detailed semi-structured interview. They are asked about their subjective perceptions, their strategies for identifying phishing attempts and their opinions on the clarity and effectiveness of different warning designs. This qualitative data enriches the study's findings by adding personal personal context to the observed behaviours and eye tracking metrics. The interview questions cover several key areas:

- General Perception: Inquiry into how participants perceived the warnings overall, including any elements that particularly stood out to them and the influence of the warnings on their decision making process.
- Effectiveness: Evaluation of the warnings' effectiveness presented during the study, identifying any warnings perceived as particularly helpful or less so. This includes a comparision between the different warnings.
- Design: Assessement of the warnings' design aspects, such as color, placement, and animations, and their impact on the users' attention.
- Personal Behaviour: Discussion around any new information noticed in the warnings that hadn't been considered before and typical behaviours when encountering phishing emails.
- Feedback: Suggestions for improving future iterations of the study and an opportunity for participants to raise any further questions or comments (Fig. 3).

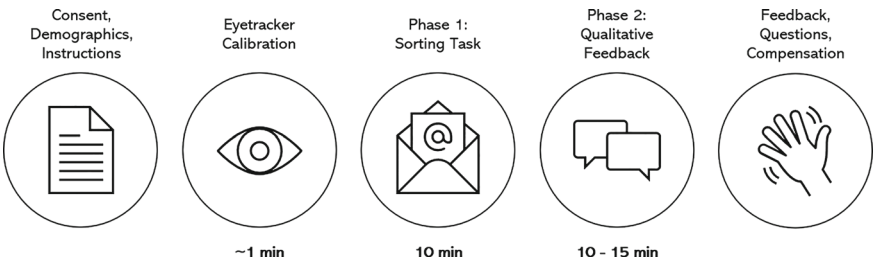


Fig. 3 Summary of the study procedure

The full study protocol and questioning framework is provided in appendix A and B.

10 Data Analysis

10.1 Qualitative Data Analysis

To enhance my understanding of participants' perceptions regarding different phishing warning designs, I employed a Qualitative Data Analysis (QDA) approach. Following established methodologies outlined in literature [18], my method involved a hybrid coding strategy, combining both deductive and inductive techniques. Initially I established a set of codes based on my research questions and theoretical framework (deductive). As I analyzed the data, I expanded and refined this set to include new themes and patterns that emerged from participant feedback (inductive). This approach ensured that my analysis remained both structured and adaptive to the insights yielded by the study.

To ease up the exploration of qualitative data, I utilized MAXQDA2024. This software provided robust tools for organizing, coding, and analyzing the interview data, making the complex process of qualitative analysis more manageable and systematic.

Our coding framework captures the key dimensions of my research. Each main category and sub-category focuses on different aspects of how users interact with and perceive the visual and functional elements of phishing warnings. A detailed list of codes with explanations is provided in appendix C.

10.2 Eye tracking Data Analysis

To evaluate the eye tracking data, I formulated the following hypotheses:

H1 Delayed slide-in animations of warnings will disrupt typical email reading patterns and draw more attention.

H2 Warnings placed on the side of the email interface will be noticed later compared to those within the email body or at the top.

H3 Content integrated directly into the email body (e.g. Greeting or Signature Warning) will hold user attention for longer periods.

To analyze the eye tracking data, I used the Tobii Pro Lab software, which provided many tools for visualizing and interpreting gaze patterns. My analysis involved several key tasks (Fig. 4):

1. **Defining Areas of Interest (AOI)**: I defined specific AOIs within the email interface. This allowed me to quantify metrics such as fixation durations and

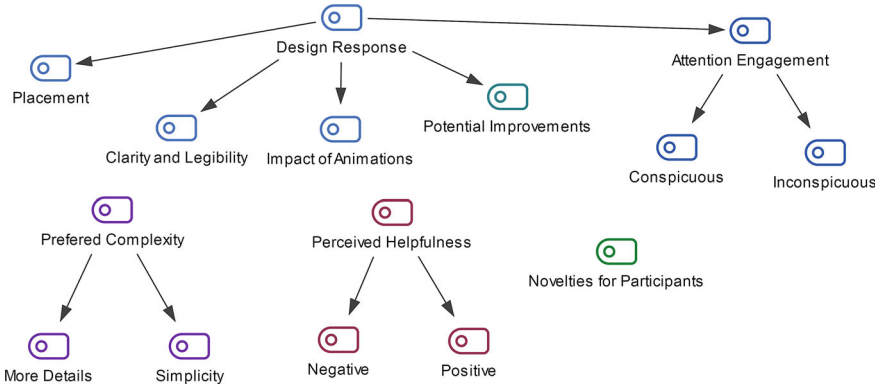


Fig. 4 Graphic illustration of my QDA codes, generated in MAXQDA 2024

frequencies for specific areas. The AOIs included different parts of the email where warnings were placed (see Fig. 5).

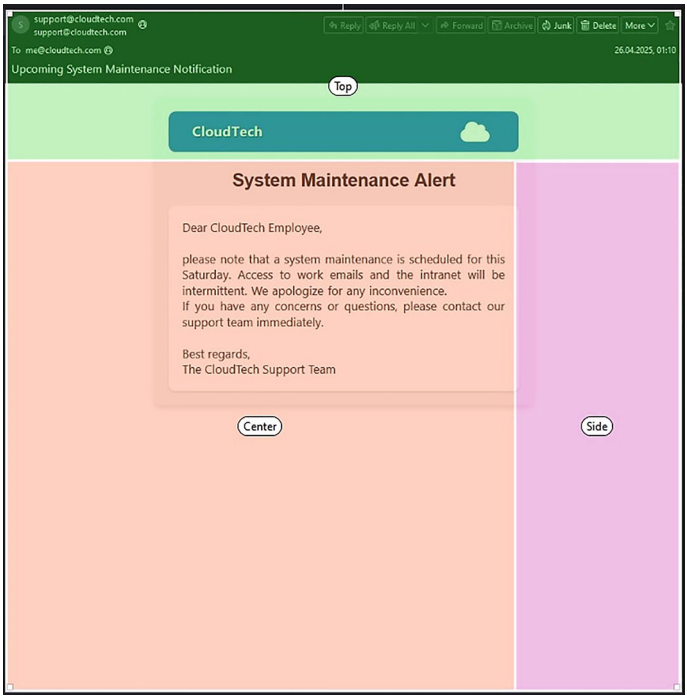


Fig. 5 AOI zones in Tobii Pro Lab

2. **Establishing Times of Interest (TOI)**: To focus on critical moments of interaction, I defined TOIs. These were specific instances when participants opened emails containing warnings.
3. **Visualization and Gaze Path Analysis**: Using Tobii Pro Lab's visualization tools, I oberserved and analyzed gaze points. This was crucial for Hypothesis 1, as it allowed me to see how animations affected reading patterns.

To align with my hypotheses, I collected the following eye tracking metrics:

1. **Time to First Fixation**: This metric measures the time it takes for participants to fixate on a specific AOI after the email is presented [19]. It helps me understand how quickly each warning design draws attention compared to others.
2. **Average Fixation Duration**: By averaging the duration of all fixations within each AOI, I can assess the overall engagement level with different components of the email interface. Fixation durations range from 150-300 ms, with higher durations meaning a higher level of engagement [19].

These methods and metrics allowed me to test my hypotheses and evaluate the effectiveness of different phishing warning designs in capturing user attention.

11 Design and Implementation

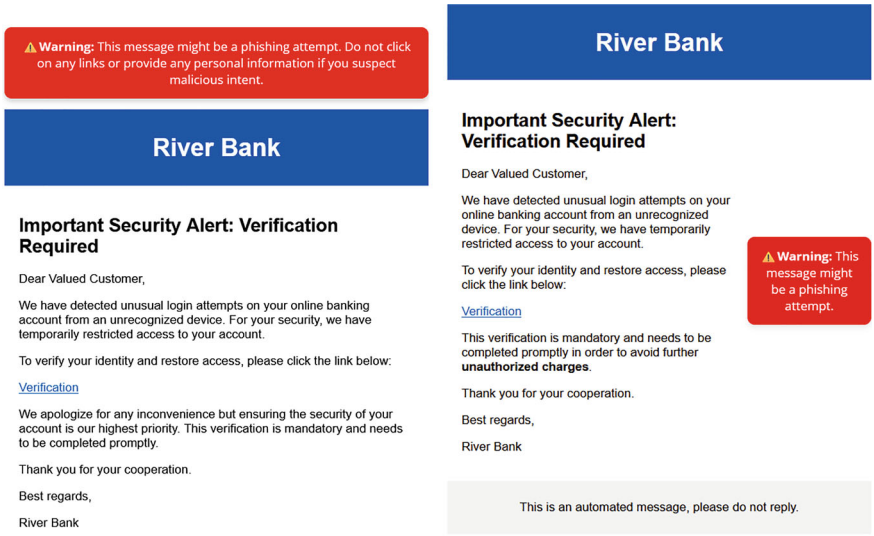
11.1 Warning Types

For my study I designed four primary types of phishing warnings, each having two different variations. These variations aim to explore different aspects of user interaction and attention.

Group 1: General Alerts Generic Banner: A broad, non-specific warning that does not target any specific content within the email but serves as a general alert to the possibility of phishing.

- Variant 1: The banner slides in from the top. This placement tests the immediate visibility and impact of warnings when users first open an email.
- Variant 2: The banner is positioned on the right side of the email body. This alternative aims to assess how peripheral placement affects user noticeability and response compared to more traditional top placement.

Group 2: Context-Aware Warnings Tooltip Warning on Link Hover: This warning appears when a user hovers over a link, providing immediate contextual information about the potential danger of the link, making it context aware as it is specifically related to the actions the user is about to take (i.e., clicking a link). A tooltip appears, warning the user and preventing them from clicking the link. After a 3 s countdown, the tooltip moves to the side, making the link clickable at the users own risk.



(a) Variant 1 (b) Variant 2

Fig. 6 Generic banner

This approach aligns closely with the research conducted by Petelka et al. [20], which emphasizes the effectiveness of proximity and user interaction with link-focused phishing warnings in reducing susceptibility to phishing attacks. In this thesis, the tooltip warning activates when a user hovers over a link, immediately providing contextual information and temporarily blocking access to the link, thereby forcing the user to recognize and process the warning before proceeding. This method mirrors Petelka et al.'s findings that such forced interaction can significantly heighten user caution and awareness. My study further explores the impact of varying levels of informational detail in the tooltips, assessing how these variations influence user perception (Fig. 6).

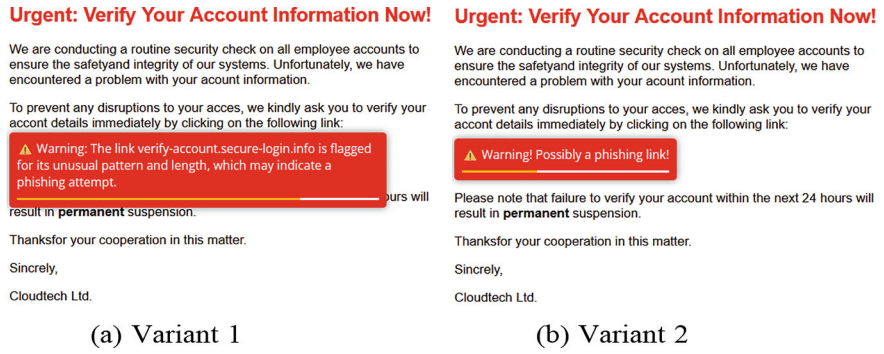
- Variant 1: Includes more detailed information within the banner, such as the actual URL. This version tests whether providing additional context within the warning influences user decisions about clicking on links.
- Variant 2: Presents less information, focusing on a simple cautionary message. The aim here is to evaluate the effectiveness of minimalistic warnings in deterring harmful interactions.

Group 3: Content-Specific Warnings Signature-Specific Warning: This warning is triggered by anomalies in the email's signature block which may indicate that the message is not from a legitimate source. It focuses on deviations from standard or expected signature formats commonly seen in regular correspondence with the sender.

- Variant 1: Accompanied by proof from past emails, offering a direct comparision to highlight discrepancies. This tests the baseline effectiveness of alerting users to potential signature anomalies with additional context.
- Variant 2: Presents less information, focusing on a simple cautionary message next to the signature (Figs. 7 and 8).

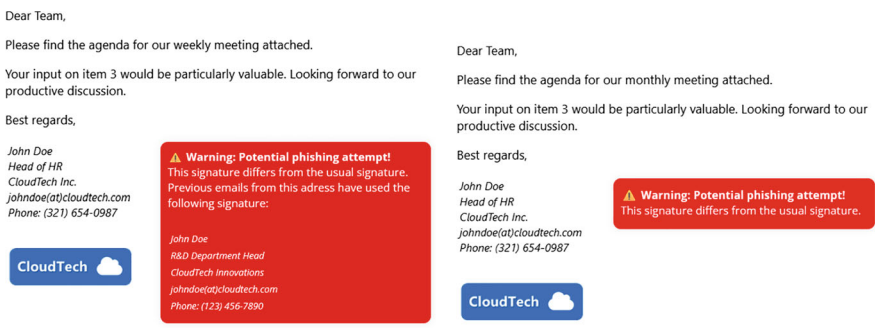
Greeting-Specific Warning: Similar to the previously described signature warning, this alert focuses on irregularities in the email's greeting. It is designed to catch attention when the greeting does not match the usual formats, suggesting that the email could be a phishing attempt.

- Variant 1: Includes "proof" from past emails, similar to the signature warning variation. Additionally this warning banner is interactive, as the user can click on the corresponding email, opening it in a new tab. This is designed to see if direct comparisons with previous legitimate emails help users to better understand why the email might be a phishing attempt.



(a) Variant 1 (b) Variant 2

Fig. 7 Tooltip warning on link hover



(a) Variant 1 (b) Variant 2

Fig. 8 Signature-specific warning

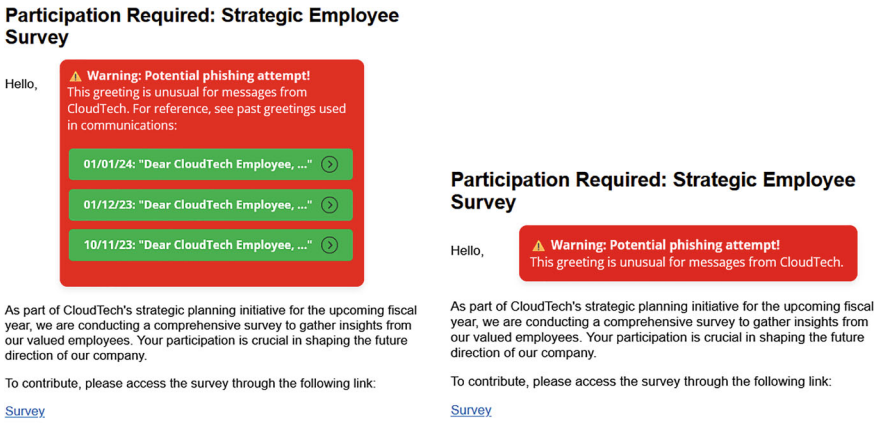


Fig. 9 Greeting-specific warning

- Variant 2: Without additional context, focusing on a simple cautionary message next to the greeting. This variation assesses users' ability to detect phishing without the aid of explicit historical comparisons (Fig. 9).

12 Visual Design Strategy and Rationale

We followed a visual design strategy for the phishing warnings developed in my study, aiming to merge aesthetics with functionality to create an effective warning mechanism. Building upon established research [21], my design approach was crafted to enhance the salience, recognizability, and comprehensibility of warnings. I focused on several key elements identified in the literature as critical for maximizing the effectiveness of visual warnings: color usage, typography, and the inclusion of universally recognized symbols. These elements were chosen based on their proven ability to capture attention quickly and convey urgency effectively, ensuring that users could not only notice but also understand the warnings promptly and act accordingly.

12.1 Color Scheme and Typography

The choice of a vivid shade of red (hex: #ff4d4d) for the background of my warnings leverages its psychological impact. This decision is rooted in research indicating that red can trigger heightened alertness and caution [22, 29]. The addition of a contrasting yellow triangle, serving as a pictorial symbol universally recognized for caution, further enhances the visibility and impact of the warnings. This use of

universally understood symbols is supported by literature emphasizing the importance of immediate comprehension in warning design [21]. Rounded corners were introduced to the warning design to make the visual alerts appear more modern and less aggressive, maintaining user comfort without sacrificing visibility. This design choice aligns with contemporary design trends that favor user-friendly interfaces.

The font choice, Segoe UI, was selected for its clarity and the decision to use bold fonts aims to further enhance readability.

12.2 Animation and Timing

Each warning features a sliding animation that initiates one second after the email is opened. This deliberate delay and the dynamic entrance of the warnings were designed to distinguish them clearly from the static email content. By introducing the warnings as distinct, dynamic elements, I aimed to disrupt the user's initial scan of the email content, drawing their focus directly to the alert and reinforcing the perception of the warnings as separate and urgent. Research shows that animations can significantly enhance the allocation of attentional resources, making such dynamic elements more effective at capturing and holding viewer attention [23]. This strategic use of animations is therefore critical in ensuring that the warnings are not only noticed but also perceived as immediate and pressing alerts.

13 Iterative Design Adjustments

Over the course of the study, feedback from participants highlighted several areas for improvement in my warning designs. This led to an iterative design process, where modifications were made to enhance clarity and legibility. The new warnings were implemented from the twelfth participant onwards. This allowed me to compare the effectiveness of the original and modified designs directly within the same study framework. Here are the key adjustments made:

- **Color and Typography**: I intensified the shade of red used for the warning backgrounds to a deeper, more vivid tone (hex: #f00c0c). The font was changed to 'Open Sans' and I also used more bold fonts. This resulted in a generally more legibile design.
- **Animation Timing**: I removed the delay in the animations, allowing warnings to appear immediately as the email is accessed.
- **Greeting-Specific Warning**: The buttons that link to previous emails for comparison have been updated to more clearly indicate that they are interactive.

Figures 5, 6, 7, 8 and 9 in Sect. 11.1 already incorporate these changes. An overview of my first iteration of the warning types is provided in appendix **??**.

14 Technical Background

14.1 Email Client

Our experimental setup was anchored in Mozilla Thunderbird, selected for its adaptability in customizing email functions via plugins. I utilized a portable installation of Thunderbird,[2] which facilitated an easy setup and ensured portability between setups. Custom emails were directly placed into the local folder of Thunderbird, eliminating the need for participants to log into an actual email account, thereby simplifying the process and enhancing security. This approach also simplified the process of resetting the setup after each participant, ensuring consistency and efficiency across sessions.

14.2 Emails and Custom Plugin

To faciliate the integration of phishing warnings into the emails, the warnings were directly embedded into the HTML structure of each email. This integration was aimed at replicating how an email client's security system might naturally react to detecting potential threats.

Additionally, a custom plugin was developed to enhance the functionality of the phishing warnings embedded within the email content. Utilizing Mozilla Thunderbird's WebExtensions API[3] and JavaScript code, this plugin was engineered to introduce dynamic visual cues, such as animations and interactive elements. The core functionality of the plugin centers around its capability to inject JavaScript code directly into the HTML of each email. This injection process is crucial as it allows for real-time modification of the email content, enabling animations and interactive features that provide additional context.

Check Email Subject

As I only want my plugin to work for a predefined set of emails, the script first checks the subject line of the email that is currently viewed. If it detects one of my predefined phishing emails, the plugin dynamically injects the necessary scripts into the email's HTML content. This method ensures that the interactive elements and animations are only activated when relevant.

Event Listeners

For my hover-triggered tooltips on suspicious links I added an event listener to the hyperlink, which listens for mouse events and triggers the appearance of the tooltip when necessary.

---

[2] https://www.thunderbird-mail.de/lexicon/entry/44-portable-thunderbird/?l=2.

[3] https://webextension-api.thunderbird.net/.

Animations

For animations I used the CSS @keyframes at-rule.[4] Doing this via the plugin was strictly necessary as Mozilla Thunderbird does not natively support animations embedded in the HTML code of emails.

Interactive Elements

This plugin also makes the functionality of variant 1 of my greeting-specific warning (9a) possible. Within the HTML of the email, I embedded a list that represents past emails. The plugin expands each list item with an interactive capability: clicking any list entry opens a new tab displaying the content of the corresponding past email. Additionally the plugin handles the styling of the buttons. It applies a green background color to indicate interactivity and sets the cursor to a pointer, visually cueing users that the entries are clickable.

Mock Phishing Page

We implemented a feature in the plugin to handle cases when users click on links within my mock phishing emails. Upon doing so, my plugin redirects them to a mock phishing page, designed to simulate the potential consequences of clicking on a phishing link. In terms of the technical implementation, the feature operates by attaching an event listener to each hyperlink within the mock phishing emails. When a link is clicked, rather than following the original URL, the event listener intercepts this action and instead directs the user to my predefined HTML page. This workaround was necessary as Mozilla Thunderbird restricts plugins from directly embedding links in emails for security purposes (Fig. 10).

14.3 Python Script to Shuffle Emails

Additionally, a Python script was utilized to shuffle the date of each email. As the emails were sorted by their corresponding dates, this allowed me to randomize the order of the emails for each participant. This step was critical to minimize any sequence bias and ensure that each participant's exposure to the emails was unique. The full source code of the plugin, the Python script, and the email samples are provided in my public.

[4] https://developer.mozilla.org/en-US/docs/Web/CSS/@keyframes.

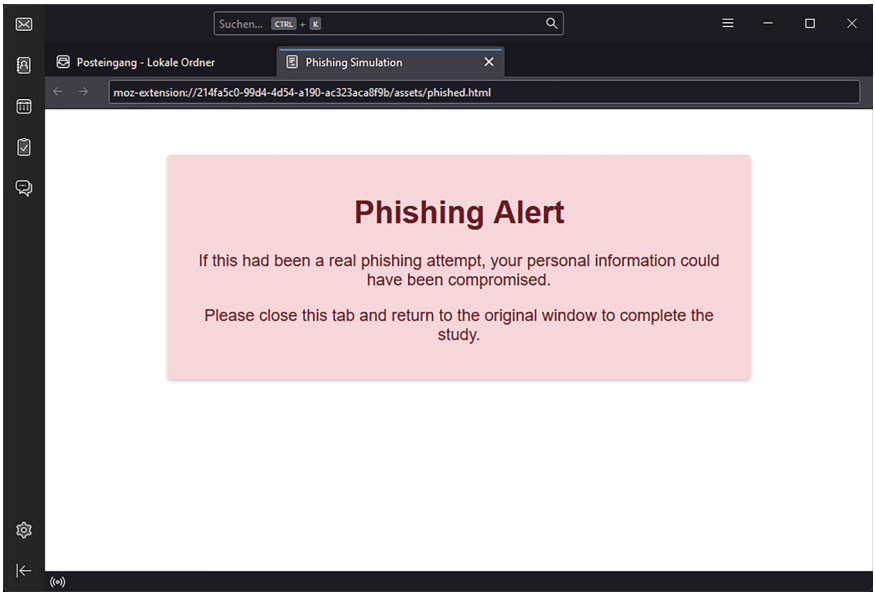


Fig. 10 Mock phishing page when participants click on links

15 Results

15.1 Questionnaire Data

Demographics The demographic data from my study primarily shows a young participant base with 81.25% aged between 20 and 29, primarily students (87.5%). The gender distribution was majority male (62.50%) with the remainder female. Most participants had at least some university education, with 56.25% holding an undergraduate degree and a smaller number having graduate degrees or a high school diploma. This generally younger demographic was anticipated, as the recruitment strategies employed were focused on platforms commonly frequented by younger audiences, such as university chat groups on Discord, WhatsApp, and Slack, as well as through the LMU newsletter (Table 2).

ATI-S Scale A significant aspect of the participant profile is their technological affinity, as indicated by the Affinity for Technology Interaction Short Scale (ATI-S) values. This data revealed a pronounced tendency towards technology, with 12 participants scoring high or very high in technology affinity. This implies that the majority of participants were not only familiar with but also comfortable using technology, which is a crucial consideration in the context of phishing awareness and the effectiveness of warnings (Fig. 11).

Table 2 Demographic data collected in the survey

Demographics	Total (N = 16)	%
Age		
20–29	13	81.25
30–39	2	12.50
40 +	1	6.25
Gender		
Male	10	62.50
Female	6	37.50
Non-binary	0	0.00
Prefer not to tell	0	0.00
Education		
Doctorate degree	1	6.25
Graduate degree	2	12.50
Undergraduate degree	9	56.25
High school diploma	4	25.00
Secondary education	0	0.00
Other	0	0.0
Profession		
Student	13	87.5
Research Assistant	1	6.25
Retiree	1	6.25

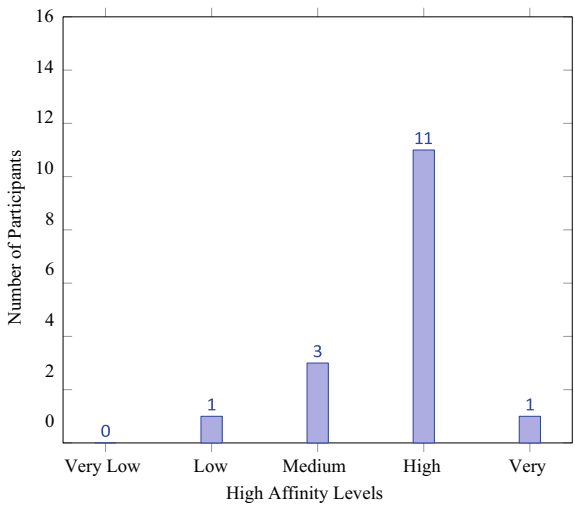


Fig. 11 Distribution of affinity for technology interaction among participants

16 Interview Results

Design Response

Clarity and Legibility: Participants generally found the color red effective in drawing attention, with one noting "*The red is just very catchy*" (P9), highlighting the visual impact of vibrant colors. The legibility of warnings was often a concern, as evidenced by comments such as "*It's the white on red what makes it difficult to read at times.*" (P9), indicating difficulties with text contrast.

Impact of Animations: Animations were noted for their effective attention grabbing capabilities. For instance, one participant mentioned, "*When the text field moves in like this, your eyes automatically move there.*" (P2), showing how animations guide viewers' focus. A participant remarked on the possible negative impact of frequent animations, stating, "*Moving warnings are more noticeable than ones that just appear on the screen. But if you put moving warnings on all phishing mails, it might be overwhelming and tire the eyes and mind. A balance is necessary. Maybe being too flashy, could even lead to desensitization over time.*" (P5) Some participants also noted that the animations were too slow to appear, "*It would be better if the warnings appeared immediately upon opening the email instead of having a delay*" (P5), which could lead to critical oversights.

Placement: Participants commonly observed that warnings placed on the side of the email were less noticeable, often residing in their peripheral vision. As one participant expressed, "*I actually found this one a bit inconspicuous because you have the email body here and the warning there. It reminds me of those advertisements on websites placed on the side of the page.*" (P1) Conversely, warnings integrated directly into the email body proved to be more conspicuous. Participants reported that these warnings grabbed their attention immediately upon opening the email, as they were in the direct line of sight while reading the email content.

Perceived Helpfulness

Context-Aware Warnings (Tooltip on Hover Over Link): This type of warning was generally seen as highly effective because it actively prevents clicks on potentially malicious links. A participant mentioned, "*I believe the one covering the link was the most effective because it immediately prevented interaction.*" (P6).

Content-Specific Warnings (Signature and Greeting Warnings): This type of warning was appreciated for its specificity, noting that these kind of details are easily overlooked in phishing emails: "*I really liked the signature warnings. That could be something that might not catch the user's attention at all.*" (P5). A common point of criticism was, that this kind of warning is often not applicable, as not every email, especially in a non-business context, ends with a signature or uses a consistent greeting.

General Alerts: This type of warning received mixed feedback due to their positioning and the lack of detailed information they provided. Participants noted its visibility

issues and expressed a preference for warnings that include explanatory details about the alert: "*Those that just flew in and had standard text might be the weakest.*" (P6).

Preferred Level of Detail

A key theme that emerged from the interviews was a clear division in preference for either more detailed warnings or simpler alerts. Participants who favored more details appreciated the additional context and explanations provided in the warnings, which helped them understand why an email was flagged as suspicious. One participant stated, "*I preferred the longer ones, because it teaches me how to recognize future phishing attempts.*" (P11), emphasizing the educational aspect. Another point mentioned was the fact that a longer explanation might make it easier and more helpful for non tech-savy people. The Greeting Warning with interactive buttons presented a unique challenge during the study. While intended to enhance user engagement by providing direct context and proof, many participants either did not realize the buttons were clickable or hesitated to use them. This hesitation stemmed from a fear that interacting with the buttons might inadvertently lead them into a phishing trap. One participant expressed, "*I wouldn't like to click on those buttons if I would get such a warning. I don't really know if I can trust it.*" (P16).

Conversely, some participants (P5, P6) preferred simple warnings, valuing the straightforwardness and immediacy of the message without additional information. This preference indicates that for some users, efficiency and speed in email processing are prioritized over detailed educational content.

Novelties for Participants

The qualitative data analysis underscored that many participants found the greeting and signature-specific warnings to be novel features. These aspects of the phishing warnings introduced new elements that participants had not previously encountered or considered significant in identifying phishing emails. As one participant reflected, "*I have not seen such tailored warnings in phishing emails, which recognize a specific pattern and tell you why it's a phishing attempt. That is very new to me.*" (P13).

Refining Warning Design: Incorporating Feedback

As outlined in Sect. 13, modifications were made to the phishing warning designs beginning with participant 12. These changes were positively received, as indicated by subsequent participant feedback. Post-adjustment, there were no further complaints regarding the readability of the text or the timing of the animations, except for a single comment about the red tone being potentially too strong (P15). Participants appreciated the quicker responsiveness and greater clarity of the warnings, which effectively addressed the initial concerns raised about legibility and the timing of animation effects.

17 Eye Tracking Results

During the data analysis phase, it was noted that the eye tracking data for one participant (P14) was significantly lower in sample size compared to others. This anomaly could be attributed to improper seating position or a flawed calibration process. To maintain the integrity and reliability of the analysis, this participant's data was excluded, resulting in a total of 15 eye tracking samples being considered for the final evaluation.

H1: **Delayed Animation's Influence on Reading Pattern**

Eye tracking data indicated no significant disruption in reading patterns due to delayed animations in phishing warnings. Observing the gaze patterns showed that the onset timing of animations, whether immediate or delayed, did not alter the attention flow significantly. Users exhibited similar engagement with immediate and delayed warning presentations, suggesting a uniform approach to processing these visual cues.

H2: **Visibility of Side-Placed Warnings**

The eye tracking analysis showed that warnings positioned on the side of the email interface captured attention quickly but consistently later than other warnings. The focus of gaze primarily remained on the other areas of the emails, with peripheral warnings drawing attention only after the central and top content had been reviewed. This observation was consistent across the study's participants, supporting the hypothesis that side-placed warnings are less immediately noticeable.

Eye tracking data (Fig. 12) supports this finding, showing that the median time taken for participants to first fixate on the Side AOI was approximately 2 s. To put this in contrast I compared this to the Top AOI data, where I recorded a notably quicker median time to first fixation of approximately 1.2 s.

H3: **Sustained Engagement of Body-Integrated Warnings**

Data strongly supported the hypothesis that warnings integrated directly into the email body were engaged with more thoroughly by users than those placed at the side. The analysis of average fixation durations (Fig. 13) revealed that engagement times for the Center AOI were generally longer across participants. However, there were notable exceptions for participants 7 and 13, where the fixation duration on the Center AOI was shorter compared to the other participants.

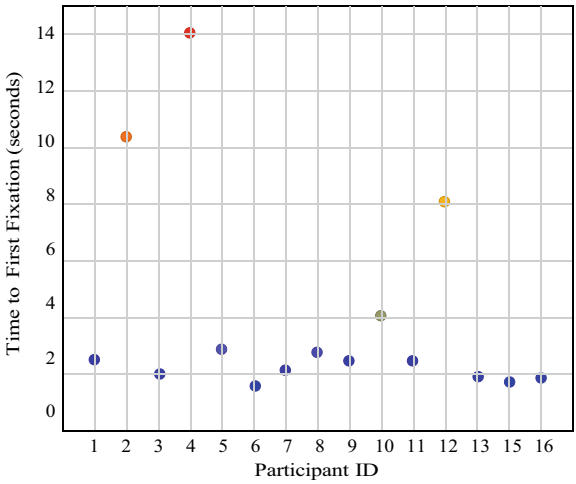


Fig. 12 Time to first fixation for the side AOI

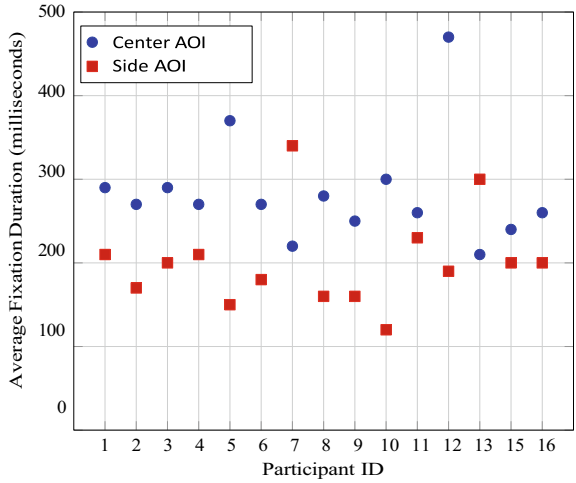


Fig. 13 Comparison of the average fixation duration for center and side AOIs

18 Discussion

18.1 Interpretation of Qualitative Feedback

Emerging Themes and Best Practices The qualitative feedback analysis illuminates several emerging themes that contribute to the effectiveness of phishing warnings. Reflecting on these, I can delineate best practices for designing phishing warning systems that align with user expectations.

Attention-Grabbing without being Overwhelming

A recurring theme is the delicate balance between capturing users' attention and avoiding sensory overload. Warnings that are too subtle may go unnoticed, while those that are too aggressive could be dismissed as annoyances. Best practices suggest employing moderate animations and placing warnings within the users' immediate visual field. For example, one participant noted, "*When the text field moves in like this, your eyes automatically move there.*" (P2) indicating the value of animated cues. However, animations should not be excessively flashy or complex, as they risk desensitizing users over time. Supporting this, a study on the effects of animation in online contexts found that while animations increase visibility and engagement, they must be balanced to avoid overshadowing important content or overwhelming the viewer [23].

Clarity and Contextual Information

Clarity in warning messages is essential. Participants expressed a preference for warnings with a clear and legible typography set against a contrasting background, improving readability. The use of the color red was frequently mentioned as effective due to its association with caution and urgency.

Moreover, providing contextual information within the warning, such as the reason an email was flagged and highlighting subtle language cues, such as the greeting, can enhance users' understanding and prompt more informed action. As one user aptly put it, "*I preferred the longer ones, because it teaches me how to recognize phishing attempts.*" (P11), highlighting the educative function.

This emphasis on context-rich warnings aligns with finding by Buono et al. [24], who emphasize that explantory warnings enchance user understanding and reduce risky intreactions. Such strategies are crucial for improving user understanding and ensuring that warnings lead to informed actions and thereby enhancing the overall effectiveness of phishing warnings.

Trust and Interactivity

The trustworthiness of warnings is key; users must feel confident interacting with them. This includes a hesitance to engage with clickable elements within the warnings due to fear of intensifying a potential threat. Best practices would therefore involve clearly distinguishing warning elements from potential phishing mechanisms and ensuring that interactive components, when used, are immediately recognizable as safe and official parts of the email client interface.

Incorporating User Feedback for Iterative Improvement

The inclusion of user feedback into the design process is not only beneficial but essential for the development of effective phishing warnings. Adjustments made in response to participant insights, such as enhancing color contrast to improve legibility, exemplify how iterative design based on user experience can significantly help in designing effective security measures. As highlighted in the article by Grobler et al. [25] actively engaging users in refining security solutions helps tailor these systems

to better meet their needs and expectations, thereby increasing their efficacy and user adoption rates. Best practices would therefore advocate for continuous user testing and refinement, ensuring that warnings adapt over time to align with evolving user behaviors and phishing tactics. This proactive approach ensures that security measures remain robust, user-friendly, and effective in real-world scenarios.

Summary: Users' Perceptions of Phishing Warnings To address RQ1 regarding user perceptions and the effectiveness of various phishing warnings in email interfaces, I analyzed qualitative feedback from participants. The effectiveness was assessed based on the ability of warnings to enhance user awareness, encourage appropriate actions to mitigate risks, such as not clicking on malicious links, and improve recognition of phishing threats in future encounters.

Perceived Helpfulness and Effectiveness of Warning Types

1. **Context-Aware Warnings (Tooltip on Hover Over Link)**: Participants found context-aware warnings extremely effective. These warnings prevent direct interaction with malicious links, thereby immediately mitigating risk. This direct intervention makes them highly valuable in protecting users in real-time.
2. **Content-Specific Warnings (Signature and Greeting Warnings)**: These warnings are appreciated for their specificity and relevance, particularly in environments where consistency is expected, such as professional emails. They provide clear and relevant context by highlighting anomalies in expected email patterns, offering a direct comparison with typical content. Their educational value is significant as they enhance users' ability to identify subtle signs of phishing attempts, which might otherwise be overlooked.
3. **Generic Banners**: Generic banners received the least favorable feedback. Their effectiveness was often questioned due to poor placement and lack of detailed information, which led to them being easily ignored. Especially variant 2 (side placed banner) was criticized for being too inconsipicous. Participants compared its noticeability to peripheral advertisements on websites, which are often disregarded due to desensitization. This feedback highlights the importance of not only the placement of warning banners but also the inclusion of specific reasons why an email might be flagged as potentially dangerous, enhancing user understanding and response to phishing threats.

Preferred Warning Design

From the feedback, it's clear that users prefer warnings that are not only visible but also informative. Warnings that blend seamlessly into the user's workflow in the email interface, such as those integrated within the primary content, are particularly effective. These warnings grab attention quickly, provide essential information, and are directly interacted with, enhancing both immediate and long-term phishing threat recognition (Table 3).

Table 3 Overview of warning types and user feedback

Warning type	Key feature	User feedback
Context-aware	Actively prevents clicking on links	Very positive; users felt it was very effective and helpful, appreciating the proactive nature of the warning
Content-specific	Highlights anomalies in the language	Positive; users appreciated their specificity and educational value, noting they felt more informed about phishing tactics
General	Broad, non-specific warning	Less favorable; often overlooked, reported by users as too generic and lacking actionable details

19 Interpretation of Eye Tracking Results

19.1 Impact of Delayed Animations

Our hypothesis that delayed animations would disrupt the typical email reading pattern and draw more attention was not supported by the data. The implications suggest that while animations serve as useful attention-grabbing tools, their effectiveness does not necessarily benefit from a delayed presentation. Aligning with the qualitative interview results, it appears that ensuring the visibility of warnings immediately upon email viewing is more crucial for capturing attention and facilitating prompt user reactions.

However, it is important to note, that the evaluation of this hypothesis may be limited by the methodology used. Only 4 participants tested the new warnings with no animation delay, which may not provide a sufficiently robust dataset to draw definitive conclusions.

19.2 Effective Warning Placement

The data supports that warnings placed within the central viewing area of an email are engaged with more extensively. This is evidenced by generally longer average fixation durations on the Center AOI. These findings underscore the efficacy of aligning warning placements with natural reading behaviors, thereby enhancing the visibility and impact of security warnings.

However, the analysis revealed notable deviations for participants 7 and 13, who exhibited shorter fixation durations on the Center AOI compared to other participants. This variability in user interaction suggests that while centrally-placed warnings are generally effective, their impact can vary significantly among users due to individual differences in reading habits or the visual salience of the warnings.

The extended engagement with central warnings highlights their effectiveness in maintaining user attention, which is crucial for ensuring that warnings are not only noticed but also sufficiently processed. This is vital for the effective communication of security-related information. Based on these insights, designers are encouraged to position crucial information and alerts within the central visual field to maximize their impact. Furthermore, understanding the reasons behind the shorter fixation times for specific participants could inform more tailored design approaches, ensuring that warnings are effective across a diverse user base.

In contrast, warnings positioned on the periphery of the email interface, such as on the sides, were consistently noticed later and engaged with less extensively. Although my data indicates only a slight delay in detection, this brief period can be critical, particularly during a hectic workday when users may be under pressure or multi-tasking. This oversight provides a window during which users could inadvertently interact with deceptive links or other malicious content. Moreover, the presence of outliers where the time to first fixation was exceedingly long highlights the potential severity of this issue.

Furthermore, the average fixation duration on these side-placed warnings was also notably lower than that for central warnings, suggesting that even when the warning is noticed, the engagement is less intensive. This lack of deep engagement suggests that the warnings may not be registering effectively with users. This observation aligns with participant feedback, where many reported not noticing the side-placed warnings at all, despite eye tracking data confirming visual detection.

This phenomenon, supported by the comments of two participants (P1, P9), reinforces the argument that peripheral warnings may be less effective due to a learned tendency to disregard non-central elements. The reduced fixation durations further signify that users engage only superficially with these warnings when they do notice them, which could seriously undermine their effectiveness in security-sensitive environments.

19.3 Summary: Gaze Patterns and Warning Placement

RQ2 explores how common gaze patterns among users can inform the strategic placement of phishing warnings. The analysis confirms a predominant focus on the central area of the email interface, a natural result of typical reading behaviors. This focal tendency significantly benefits the placement of warnings directly within this central area, ensuring they are both quickly noticed and more actively engaged with.

These insights provide a robust foundation for future email interface designs, suggesting that to maximize the impact of phishing warnings, designers should prioritize central visual placement of critical security cues. This approach will significantly improve defense mechanisms against phishing attacks, ensuring that warnings are not only visible but also integrated into the flow of natural user interactions with email content.

20 Limitations

20.1 Participant Diversity and Sample Size

The majority of my participants were drawn from a similar demographic group, predominantly consisting of university students and young adults within the same age range and academic environment. Moreover, a significant portion of participants demonstrated a high Affinity for Technological Interaction (ATI), as illustrated in Fig. 11. This heightened technological comfort and skill among my sample may not accurately represent the broader population's varying levels of technological proficiency. The lack of diversity in my sample, both demographically and in terms of technological affinity, could limit the generalizability of my findings across a broader population and potentially introduce biases in the evaluation of my phishing warnings. For example, older adults might interpret and respond to visual cues differently due to variations in technological fluency and cognitive processing speeds, which may differ markedly from the responses observed in younger, more technologically adept participants [26].

Additionally, the iterative design adjustment made based on participant feedback later in the study highlighted another significant limitation related to sample size. After implementing changes to the warning designs, it was planned to test these modifications with eight additional participants. Unfortunately, only four additional participants were able to participate in the study. This smaller than-expected sample size of participants might not provide a sufficiently robust dataset to draw definitive conclusions about the effectiveness of the redesigned warnings.

20.2 Alert Fatigue

An important aspect, which I did not explore in this study, is the potential for alert fatigue, a phenomenon where users become desensitized to warnings due to frequent exposure. This desensitization can result in reduced effectiveness of the warnings over time [27].

While my study focused on the immediate effectiveness of different phishing warning designs, it did not account for the long-term impact of repeated exposure to these warnings. To mitigate the effects of alert fatigue, future research could explore strategies such as varying the design and presentation of warnings.

Additionally, longitudinal studies could assess the impact of repeated exposure to phishing warnings over an extended period and provide valuable insights into how to design more resilient warning systems.

20.3 Study Setup and Ecological Validity

The use of a controlled lab environment to conduct this study poses significant limitations regarding ecological validity. Participants were aware that their actions were being observed, potentially influencing their behavior—a phenomenon known as the Hawthorne effect [22]. This awareness can alter natural responses and may not accurately reflect their behavior in less controlled environments. Furthermore, the artificial nature of the study environment can significantly impact participants' risk perception [28]. Since the risk to personal data is nonexistent in a simulated phishing attack, participants' reactions to warnings and their decision-making processes might not truly reflect their actions in a real threat scenario. This discrepancy can skew results related to actual risk perceptions and behaviors, potentially leading to overestimations of the effectiveness of warning systems in genuine contexts.

21 Conclusion

This thesis has extensively investigated the impact of various visual warning designs on user responses to phishing threats in email clients. The use of eye tracking technology and qualitative feedback has provided a deep understanding of how different warning strategies affect user behaviour and engagement.

The research conclusively demonstrates that integrating warnings directly within the email content markedly enhances both the immediacy of user reactions and their overall engagement with the warnings. These findings underline the importance of prominent, context-rich warnings that are easily visible and provide actionable information to users.

Furthermore, the study highlights the crucial role of educational content in warnings. Clear, informative warnings not only alert users to immediate threats but also educate them, enhancing their ability to identify similar threats independently in the future. This dual function of warnings is vital in the ongoing battle against phishing, as well-informed users are the first line of defense in cybersecurity.

However, the study also highlights significant opportunities for expansion in both scope and realism for future research. A crucial first step is broadening the participant pool to include a more diverse range of ages, professions, educational backgrounds, and increasing sample sizes to ensure findings are robust and generalizable. This will enhance my understanding of how different user groups perceive and react to phishing threats, supporting the development of more universally effective defensive measures. Additionally, to significantly boost ecological validity, future studies could integrate phishing detection research into real-world user environments. This advancement might entail the development and application of phishing detection algorithms capable of operating effectively within users' naturalistic settings. Such a move would not only provide more authentic data but also allow for a robust assessment of phishing warning effectiveness under typical usage conditions.

Looking forward, the rapid advancement of artificial intelligence and large language models opens new frontiers for phishing defense. These technologies could be leveraged to develop more adaptive, personalized phishing detection and prevention strategies, enhancing the precision and responsiveness of warnings. For instance, AI could analyze user behavior to tailor warnings to individual risk profiles or detect subtle phishing attempts that evade traditional detection methods. This proactive and technologically innovative approach is crucial in an era where cyber threats are continually evolving, thus ensuring the digital safety of users globally. As cybercriminals become more sophisticated, integrating cutting-edge technologies into cybersecurity measures will be essential for staying ahead of threats and safeguarding digital interactions.

References

1. Widup, S., Spitler, M., Hylender, D., Bassett, G.: 2018 verizon data breach investigations report. In: unknown (2018)
2. Wang, J., Herath, T.C., Chen, R., Vishwanath, A., Rao, H.: Research article phishing susceptibility: an investigation into the processing of a targeted spear phishing email. IEEE Trans. Prof. Commun. **55**, 345–362 (2012). https://doi.org/10.1109/TPC.2012.2208392
3. Group, A.P.W.: Phishing activity trends report 2nd quarter 2023. In: unknown (2023), https://docs.apwg.org/reports/apwg_trends_report_q2_2023.pdf
4. Thakur, K., Ali, M.L., Obaidat, M.A., Kamruzzaman, A.: A systematic review on deep-learning-based phishing email detection. Electronics **12**(21) (2023). https://doi.org/10.3390/electronics12214545, https://www.mdpi.com/2079-9292/12/21/4545
5. Jampen, D., Gu¨r, G., Sutter, T., Tellenbach, B.: Don't click: towards an effective anti-phishing training. A comparative literature review. Hum. Comp. Inf. Sci. **10**(1), 33 (2020). https://doi.org/10.1186/s13673-020-00237-7
6. Alkhalil, Z., Hewage, C., Nawaf, L., Khan, I.: Phishing attacks: a recent comprehensive study and a new anatomy. Front. Comput. Sci. **3** (2021). https://doi.org/10.3389/fcomp.2021.563060
7. Shi, W., Xie, M.: A reputation-based collaborative approach for spam filtering. 2013 AASRI Conference on Parallel and Distributed Computing and Systems. AASRI Proc **5**, 220–227 (2013). https://doi.org/10.1016/j.aasri.2013.10.082, https://www.sciencedirect.com/science/article/pii/S2212671613000838
8. Bhowmick, A., Hazarika, S.: Machine learning for e-mail spam filtering: review, techniques and trends. unknown (2016)
9. Siddiq, M.A.A., Arifuzzaman, M., Islam, M.S.: Phishing website detection using deep learning. In: Proceedings of the 2nd International Conference on Computing Advancements, p. 83–88. ICCA'22, Association for Computing Machinery, New York, USA (2022). https://doi.org/10.1145/3542954.3542967
10. Dhamija, R., Tygar, J.D., Hearst, M.: Why phishing works. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, p. 581–590. CHI'06, Association for Computing Machinery, New York, USA (2006). https://doi.org/10.1145/1124772.1124861
11. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E.: Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In: Proceedings of the 3rd Symposium on Usable Privacy and Security, p. 88–99. SOUPS'07, Association for Computing Machinery, New York, USA (2007). https://doi.org/10.1145/1280680.1280692
12. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L., Hong, J.: Teaching Johnny not to fall for phish. ACM Trans. Internet Techn. **10** (2010). https://doi.org/10.1145/1754393.1754396

13. Wash, R.: How experts detect phishing scam emails. Proc. ACM Hum. Comput. Interact. **4**(CSCW2) (2020). https://doi.org/10.1145/3415231
14. Wang, Z., Zhu, H., Liu, P., Sun, L.: Social engineering in cybersecurity: a domain ontology and knowledge graph application examples. Cybersecurity **4**(1), 31 (2021). https://doi.org/10.1186/s42400-021-00094-6
15. Koddebusch, M.: Exposing the phish: The effect of persuasion techniques in phishing e-mails. In: DG.O 2022: The 23rd Annual International Conference on Digital Government Research, p. 78–87. dg.o 2022, Association for Computing Machinery, New York, USA (2022). https://doi.org/10.1145/3543434.3543476
16. Abdrabou, Y., Dietz, F., Shams, A., Knierim, P., Abdelrahman, Y., Pfeuffer, K., Hassib, M., Alt, F.: Revealing the hidden effects of phishing emails: an analysis of eye and mouse movements in email sorting tasks (2023)
17. Distler, V.: The influence of context on response to spear-phishing attacks: an in-situ deception study. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI'23, Association for Computing Machinery, New York, USA (2023). https://doi.org/10.1145/3544548.3581170
18. Rädiker, S., Kuckartz, U.: Analyse qualitativer Daten mit MAXQDA. Springer Fachmedien Wiesbaden GmbH (2018). https://doi.org/10.1007/978-3-658-22095-2
19. Tullis, T., Albert, B.: Chapter 7 behavioral and physiological metrics. In: Tullis, T., Albert, B. (eds.) Measuring the User Experience, pp. 163–186. Interactive Technologies, Morgan Kaufmann, Boston (2013). https://doi.org/10.1016/B978-0-12-415781-1.00007-8
20. Petelka, J., Zou, Y., Schaub, F.: Put your warning where your link is: improving and evaluating email phishing warnings. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, p. 1–15. CHI'19, Association for Computing Machinery, New York, USA (2019). https://doi.org/10.1145/3290605.3300748
21. Wogalter, M.S., Conzola, V.C., Smith-Jackson, T.L.: Research-based guidelines for warning design and evaluation. Fundam. Rev. Appl. Ergon. **33**(3), 219–230 (2002). https://doi.org/10.1016/S0003-6870(02)00009-1, https://www.sciencedirect.com/science/article/pii/S0003687002000091
22. McCambridge, J., Witton, J., Elbourne, D.: Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J. Clin. Epidemiol. **67**, 267–277 (2013). https://doi.org/10.1016/j.jclinepi.2013.08.015
23. Cheung, M.Y., Hong, W., Thong, J.: Effects of animation on attentional resources of online consumers. J Assoc Inf Syst **18**, 605–632 (2017). https://doi.org/10.17705/1jais.00464
24. Buono, P., Desolda, G., Greco, F., Piccinno, A.: Let warnings interrupt the interaction and explain: designing and evaluating phishing email warnings. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. CHI EA'23, Association for Computing Machinery, New York, USA (2023). https://doi.org/10.1145/3544549.3585802
25. Grobler, M., Gaire, R., Nepal, S.: User, usage and usability: redefining human centric cyber security. Front. Big Data **4** (2021). https://doi.org/10.3389/fdata.2021.583723
26. Ebaid, D., Crewther, S.G.: Visual information processing in young and older adults. Front. Aging Neurosci. **11** (2019). https://doi.org/10.3389/FNAGI.2019.00116
27. Ban, T., Samuel, N., Takahashi, T., Inoue, D.: Combat security alert fatigue with AI-assisted techniques. In: Proceedings of the 14th Cyber Security Experimentation and Test Workshop, p. 9–16. CSET'21, Association for Computing Machinery, New York, USA (2021). https://doi.org/10.1145/3474718.3474723
28. Garfinkel, S.L., Lipford, H.R.: Usable security: History, themes, and challenges. In: Usable Security (2014). https://api.semanticscholar.org/CorpusID:39939011
29. Kuniecki, M., Pilarczyk, J., Wichary, S.: The color red attracts attention in an emotional context. An ERP study. Front. Hum. Neurosci. **9**, 212 (04 2015). https://doi.org/10.3389/fnhum.2015.00212

The Integration of Artificial Intelligence in the Context of Industry 4.0



Aleksey G. Tashkinov

Abstract The digital transformation of production systems and the advent of Industry 4.0 have led to significant changes in the way work in the field of artificial intelligence is organized. As companies increasingly rely on automated processes and digital tools, they are looking for new and more efficient ways to manage their systems while maintaining flexibility and modularity. This article proposes an interdisciplinary approach to identifying gaps in the implementation of digital technologies in industry, with an emphasis on artificial intelligence. The author offers a methodological framework based on the principles of artificial intelligence development and machine learning methods to eliminate these gaps that have been identified in the course of research. In particular, artificial intelligence consists of machines with intelligence, characterized by sufficient computing resources and algorithms that determine the learning process in order to gain knowledge, as well as increase productivity. In order to enhance the efficacy of production processes, the author puts forth a model that integrates the fundamentals of Industry 4.0 with advanced digital technologies, such as artificial intelligence, on an enterprise-wide scale. The outcome of this work is a successful implementation of a well-designed model for the digital platform in the of Industry 4.0.

Keywords Artificial intelligence · Digital transformation · Digital technologies · Industry 4.0

1 Introduction

The twenty-first century has seen a transformative change in the industrial landscape, driven largely by the rapid advancement of artificial intelligence (AI). This shift is often referred to as the "Fourth Industrial Revolution" or "Industry 4.0", which focuses on digital connectivity, automation, and intelligent decision-making.

A. G. Tashkinov (✉)
Perm National Research Polytechnic University, Perm, Russia
e-mail: alekss.perm@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_35

The ubiquity and exponential advancement of digital technologies have given rise to a constant state of flux in consumer preferences and market dynamics. This, in turn, has placed new demands on organizations, forcing them to adapt to the changing landscape. As a result, an increasing number of companies have begun to leverage digital tools in order to optimize customer experiences and enhance their offerings through data-driven insights. However, enterprises also face economic and geopolitical uncertainty arising from the rapidly evolving technological landscape, which may lead many companies to gravitate towards more conservative approaches. To illustrate this point, let us consider a scenario where additional data is collected following the commissioning of a plant. This data serves as a verification tool, providing precise information about the release and characteristics of products, which can be used for future development purposes. Industry 4.0 technologies are instrumental in enhancing and optimizing novel technologies, particularly through the use of machine learning. Businesses and non-governmental organizations devote substantial time, resources, and effort to developing and implementing AI in order to alleviate the adverse consequences of past events [1]. The term "Industry 4.0" denotes the integration of advanced technologies into manufacturing and business operations. These technologies enable seamless collaboration across geographical boundaries, time zones, and network connections. This is accomplished through the deployment of intelligent factories, which allow for rapid scaling up or down of production, ultimately leading to increased revenue for enterprises. The International Industrial Internet of Things (IIoT) serves as a bridge between user experience and operational efficiency, offering cost-effective solutions to businesses. It empowers enterprises to gather data from diverse sources, such as sensors and devices, enabling them to make data-driven decisions [2]. In view of the above, this paper aims to provide a concise overview of the rationale behind our research. To accomplish the objectives outlined in this investigation, we will thoroughly examine the literature pertaining to the topic of our study. Through this analysis, we aspire to discern the fundamental characteristics of advancements in the realm of Industry 4.0. Subsequently, we will discuss a systematic approach entailing the development of an appropriate business model utilizing artificial intelligence. Finally, we will conclude with a summary of the key findings.

2 Background

2.1 Development Concepts "Industry 4.0"

The business landscape is undergoing periods of rapid change due to significant digital transformations, caused by various factors such as global market integration and technological advancements, particularly digitalization [3]. The wide spread of ICT, digital platforms, integration of digital technologies into business processes predetermined the transformation of economic systems into dynamic network-based

systems with a flexible structure of internal relationships. These systems are able to quickly respond to change in the external environment and increased uncertainty, which allows them to adapt to the constant emergence of new technologies and changing demands in the market [4]. With thousands of sensors installed in the enterprise and software, it is possible to automatically monitor production and control it: monitor the quality of products by choosing optimal modes or monitor the condition of equipment to predict its wear and tear and replace parts in a timely manner, extending its life and preventing downtime due to accidents [5]. On the basis of data analysis, the load can be increased pointwise, thus increasing performance. Previously, statistical data collected manually or not collected at all were used for analysis. This makes it more difficult to access accurate information to assess the current state and forecast the future state of production. Against the background of the active growth of data flows, the negative role of the human factor increases, and the risk of making incorrect decisions increases because a person cannot process all the information collected. The specified requirements are applicable both to commercial firms and nonprofit organizations. All organization should respond quickly and adequately to changes in the situation of society and the market and make appropriate changes in their activities [6].

2.2 Artificial Intelligence Development in "Industry 4.0"

To understand the evolution and application of artificial intelligence (AI) in an industrial setting, it is essential to explore its historical background and benefits. In this section, we will begin by examining the history of AI development and then discuss its implementation in this context. This section will also focus on the impact of AI and potential challenges that may arise during its implementation. Artificial intelligence is a significant technological advancement that contributes to the Fourth Industrial Revolution due to its impact on digital systems and business strategies. The concept of "artificial intelligence" was introduced by John McCarthy at the first conference on the topic, held at Dartmouth College in 1956 [1]. As we have previously discussed, this technology is intersectoral in nature and allows for the unification of requirements for verifying the functionality of AI algorithms and coordinating the development of evaluation methods and tools. This helps to avoid duplication of effort and ensures that uniform standards are followed when creating test data sets. In the light of these considerations, we can now delve into the realm of artificial intelligence (AI) development and explore its multifaceted benefits. Let us define a term that is pertinent to this discussion. The development of technologies that enable automatons and robots to make decisions represents a significant advance. This allows machines to collect data from their surroundings, analyze it for decision-making, learn from past experiences, and tackle increasingly complex challenges. This technological breakthrough has the potential to revolutionize various aspects of human life.

3 Research Design

3.1 Optimizing the Implementation of Digital Technologies in the Context of Industry 4.0

Based on the importance of what was mentioned, the contribution of this study was to frame this perspective and discuss its role in developing the performance of the companies studied. The problem that the study focuses on is related to lean management and artificial intelligence, which includes all concepts related to the productive industries sector. There has become interest in applying artificial intelligence in contemporary institutions as a branch of information technology, which researches and develops smart technology, and through computer applications, enables intelligent behavior when performing tasks with the aim of reducing the administrative burden, providing better service and high quality in solving work problems, and facilitating business. By transforming management systems into electronic systems based on artificial intelligence. The first beginnings of the concept of artificial intelligence were among neuroscientists and psychologists. The principle upon which the science of artificial intelligence is based is the principle of processing information, regardless of its nature and size, in an automated manner or with human intervention in the processing to achieve specific goals, and artificial intelligence or machine intelligence can be considered a broad field for the intersection of many knowledge. Old and new sciences and technologies. The researcher believes that artificial intelligence is a memory derived from the human mind, and uses modern technologies to simulate the human mind and analyze and understand the relationships between things. The goal of this process is to enhance the efficiency, quality, and competitiveness of the company by utilizing advanced technologies and improving production processes. This includes analyzing and optimizing current production procedures, deploying artificial intelligence algorithms, implementing production management systems, and managing customer relationships effectively. By doing so, businesses can improve their overall performance and remain competitive in a rapidly evolving business environment. In the first part, we will describe the pilot project and its objectives, as well as explain how the introduction of modern technologies, including artificial intelligence (AI), allows businesses to take a holistic approach to problem-solving in the digital world. We will also discuss how these concepts and technologies relate to the first research question (RQ1). In the second part, we will present the results of the project and discuss how modern concepts such as lean digital production have enabled us to achieve our goals (RQ2). Our work aims to implement digital transformation projects that will allow the company to take advantage of digital tools and solve problems. We are also changing our approach to production processes and behavior, considering the recommendations provided by data-driven analytical tools. These technologies will allow us to design, create, and test new business models. Based on the research positions mentioned above and the discussion points discussed regarding our research, we would like to move on to the presentation of our results.

4 Results

4.1 Implementation of Lean and Digital Manufacturing Projects: Use of Artificial Intelligence

Artificial intelligence is being applied in a more professional way in many systems that we use in our daily lives as well as business systems. Thus, understanding the basic dimensions and sub-domains of each variable is appropriate. As a first result, we will describe the process of the company's work when switching to digital production. We will look at a well-known process within the enterprise. As the first result of our work, we would like to briefly describe the process of the company's transition to digital production. The process begins with collecting data from various sensors. This data is then processed using special algorithms and software, taking into account the technological processes and production features of the company. Additionally, the drives are turned on immediately to ensure the proper operation of the entire system. However, there are still some technological processes within the company that do not fit within this closed cycle. These processes may have limited access to technological data, and information may be collected manually. Analysis is also not fully carried out in these areas. Examples of such processes include monitoring the reliability of equipment, safety of processes and employees, and energy management. With the help of artificial intelligence, we have realized that it is essential to close this gap and automate these areas [7]. Let us consider a well-known process at the enterprise. At the beginning of the process, data are collected from sensors. Then, their complex processing by algorithms via software is carried out, taking into account processes and production. Furthermore, to ensure the correct operation of the entire circuit, the actuators are immediately activated. Enterprises have other technological processes that do not fit into the framework of such a closed cycle. The amount of process data can be limited, information collection can be carried out manually, and the analysis is practically not carried out. An example of such a process is monitoring the reliability of the equipment. Monitoring the security of processes, employees, and energy management. Thanks to cloud storage, there is almost unlimited potential for analyzing, storing, and processing this data using analytical tools. All this work aims to implement digital transformation projects. Thanks to digital tools, people can work remotely, even if the data they are working with is located far away. Secondly, we are changing the way we approach production processes and behaviors, taking into account data and recommendations from analytical tools. These technologies allow us to design, create, and evaluate new business models. Traditionally, employees in manufacturing companies have been involved in optimizing technological processes. They gather data independently and use software to process it or involve their own engineers to solve problems. However, in the past ten years, a number of remote locations have emerged, including test centers, landfills, mines, and oil and gas production platforms. This has led to the idea of a centralized control center for integrated or "digital" production. It has become clear that experts can be gathered in one central location where all data can be collected and shared,

instead of having them travel to different sites. However, this still happens within the information systems of a single company and is supported only by its employees. Due to the digital transformation and the industrial internet of things concept, the company has adopted a new approach to services. The presented system utilizes artificial intelligence (AI), which allows innovative products to learn, adapt, and continuously improve. Digital devices gather information from their environment and interact with each other to accelerate their development. The ability to transfer knowledge is impressive and remarkable, as it allows experts to share their knowledge on various topics and issues [8]. Therefore, this scientific contribution provides a systematic approach to identify the implications of lean production and Industry 4.0 methods on target dimensions that are highly important for the development of production system enterprises.

5 Conclusion

The study we conducted has identified gaps in the current knowledge of a particular field and allowed us to answer the questions we raised through scientific reasoning and practical experience. Our proposed methodological approach is based on scientific principles and comprises two main components: expanding theoretical knowledge in the field and developing methodological and practical techniques for implementing Industry 4.0. According to the author, future research on artificial intelligence in Industry 4.0 should be expanded to include a wider range of industries and regions, as well as more diverse applications. This will help to increase the generalizability of the results and ensure that the research is relevant to a wider audience.

References

1. Bagnoli, C., Albarelli, A., Biazzo, S., Biotto, G., Roberto, G., Maurizio, M., Massaro, M., Messina, M., Muraro, A., Troiano, L.: Digital business models for Industry 4.0 how innovation and technology shape the future of companies. In: Future of Business and Finance (2022). Springer Nature Switzerland AG. ISBN 978-3-030-97283-7(eBook). https://doi.org/10.1007/978-3-030-97284-4
2. Díaz-Chao, Á., Ficapal-Cusí, P., Torrent-Sellens, J.: Environmental assets, Industry 4.0 technologies and firm performance in Spain: a dynamic capabilities path to reward sustainability. J. Clean. Prod. **281**, 125264 (2021). https://doi.org/10.1016/j.jclepro.2020.125264
3. Frank, A.G., Dalenogare, L.S., Ayala, N.F.: Industry 4.0 technologies: implementation patterns in manufacturing companies. Int. J. Prod. Econ. **210**, 15–26 (2019)
4. Sinha, S., Lee, Y.M.: Challenges with developing and deploying AI models and applications in industrial systems. Discov. Artif. Intell. (2024). https://doi.org/10.1007/s44163-024-00151-2
5. Tashkinov, A.G.: Methodological approach to the development of digital transformation of an enterprise. J. Innov. Bus. Ind. **03**(03), 177–186 (2024). https://doi.org/10.61552/JIBI.2025.03.006

6. Dingli, A., Haddod, F., Klüver, C.: Artificial Intelligence in Industry 4.0. Springer Nature Switzerland AG (2021). https://doi.org/10.1007/978-3-030-61045-6
7. Tashkinov A.G.: The impact of lean manufacturing and Industry 4.0 on the efficient operation of an enterprise. Int. J. Manag. Human. **10**(11), 7–12 (2024). https://doi.org/10.35940/ijmh.f4515.10110724
8. Tashkinov, A.G.: The implementation of lean and digital management techniques using artificial intelligence in industrial settings. Discov. Artif. Intell. **4**(1), 1–23 (2024). https://doi.org/10.1007/s44163-024-00186-5

The Role of Artificial Intelligence in Sustainable Food Retailing: A Pathway to Reducing Waste and Enhancing Efficiency



Şirin Gizem Köse, Ipek Kazancoglu, and Yigit Kazancoglu

Abstract Food waste arises in the supply chain's distribution, marketing, and consumption phases. Retailers face problems related to food waste since it increases operating costs, causing savings and loss of profitability. Due to food waste's environmental, economic, and social impacts, food retailers increasingly incorporating artificial intelligence-based applications into their operations to increase efficiency and ensure sustainability. Demand forecasting, inventory management and optimization, dynamic pricing, supply chain, and order optimization are some of the main artificial intelligence applications utilized to address food waste. In line with this purpose, the paper aims to highlight the importance of artificial intelligence for reducing food waste in retailing. This study focuses on how food retailers increasingly leverage artificial intelligence to reduce food waste and improve sustainability across their operations. This research was conducted through case study and these cases provide applications of artificial intelligence to the food retail industry.

Keywords Artificial intelligence · Food retailing · Food waste · Food supply chain · Sustainable food retailing

1 Introduction

In retail, food waste occurs for different reasons, such as product expiration, inefficient coordination of supply chain and demand, improper quality of control, and overstocking [1, 2]. Food waste is generated especially from grain spillage from

Ş. G. Köse (✉)
Department of Business Administration, MEF University, İstanbul, Turkey
e-mail: koseg@mef.edu.tr

I. Kazancoglu
Department of Business Administration, Ege University, İzmir, Turkey

Y. Kazancoglu
Department of Logistics Management, Yasar University, İzmir, Turkey

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_36

sacks, poor quality products, rodent attacks, faulty packaging, poor infrastructure, and damage from transportation [3]. If retailers do not implement adequate quality control processes, they may return products that exceed one-third of their shelf life to manufacturers, causing an increase in food waste [4]. Retailers use the food waste hierarchy when [5]. Accordingly, reducing and preventing food waste are to reuse, recycle, and recover and to dispose and prevent food waste is reusing, recycling, and recreating and disposing of it as energy [6]. As reuse, food retailers deliver edible food to those in need, reducing food suitable for animal feed by donation. Foods are resold in discount stores, and secondary market platforms or exchange agreements are made with distributors [7]. In order to eliminate the uncertainty of the amount of food waste, retailers are taking advantage of digital technologies and marketing knowledge [6].

Food waste has an environmental, economic, and social impact. Economic impacts include the costs of production, transportation, storage and processing of food waste [8]. Food waste increases operating costs for retailers, causing savings and loss of profitability [9, 10]. Environmental impact: dumping food waste on land causes environmental pollution and the formation of hazardous gases [6]. Social impacts arise from inadequate resource allocation regarding food waste. Therefore, reducing food waste ensures the sustainability of the supply chain, operational efficiency, reduced costs, and increased profitability [11]. In order to reduce food waste in the retail sector, advanced demand forecasting, dynamic pricing and discounts, cold chain management, and advanced inventory management are required [12]. In particular, natural and organic foods represent a higher tendency for food waste since they do not contain additives or preservatives and have a short shelf life [11].

AI-supported demand forecasting increases the accuracy of demand forecasting and helps predict changes in consumer behavior more effectively through processing and analyzing vast amounts of complex data. Apart from historical sales data, AI takes into account social media, online reviews, and current market trends data and these data provides insights. Thus, AI tries to solve problems such as inventory, demand forecasting, stocking, order optimization and planning in the retail industry [13, 14]. According to McKinsey Digital, AI-powered forecasting reduces errors in the supply chain by 30–50%. This causes transportation and storage costs to decrease by 5–10% and supply chain management expenses to decrease by 25–40% [15]. It prevents food waste by ensuring the optimal amount of product is ordered through AI-based demand forecasting.

Limited studies have been conducted on AI technologies in reducing food waste for the retailing stage [6, 14, 16]. Production planning and demand forecasting based on an AI model was developed with data obtained from a small market offering natural and organic foods in Brazil. AI processing techniques increased production planning accuracy, reduced the store's food waste, and positively affected profitability [11].

This study aims to understand how retailers utilize artificial intelligence to reduce food waste and increase sustainability across their operations. The research conducted a multiple case studies examining different retailers' examples in reducing food waste using artificial intelligence technology. This paper is organized into four

parts. Second part outlines the literature review about food waste and the significance of AI for reducing food waste in retailing sector. Third part gives examples of retailers using AI for reducing food waste. The conclusion and implication is given in the last part.

2 Literature Review

2.1 Food Waste

Food waste is "the disposal of edible food at the retail and consumer level" [17]. In this respect, food waste occurs in the food supply chain's distribution, marketing and consumption stages [FSC]. As 13% of global food waste is generated at the retail stage [18], retailers play a key role in reducing food waste at every stage of the supply chain [19]. The retail sector gives market power to generate changes into farmers, agribusinesses, and food producers, and consumers within the FSC [6, 20].

Due to the limited predictability of consumer demand, food waste may also occur due to products not being sold on time and exceeding their expiration date [21]. Retailers are filling their shelves as full as possible to satisfy consumers' expectations for increased product variety; however, this situation causes food stocks. Food waste from overstocking occurs more easily, especially due to product groups with short shelf life [22]. Perishable foods, which make up approximately 50% of sales in grocery stores, are the highest waste rate.

Consumers who face half-empty shelves associate this situation with poor store management. For this reason, retailers order more food products [6] and must prevent overstock and out-of-stock to satisfy customers [23]. Inventory management and demand forecasting studies in retail stores are carried out by store personnel. However, the minimum order quantity may be too high or too low, based on human error and the staff's biases about the necessary level of stock and order. This can lead to high rates of food waste [10, 24]. Reducing waste will directly benefit retailers in reducing their costs. In this way, they will offer lower priced products to consumers. Additionally, it will increase profitability and corporate image [9]. This will enable the company to score higher in sustainability assessments and positively impact market value.

2.2 The Importance of AI for Reducing Food Waste in Retailing

In recent years, increasing competition, multichannel conditions, more selective and indecisive consumer preferences, and data's increasing speed and size have

made traditional forecasting methods increasingly inadequate. Traditional forecasting methods based on historical static data (e.g., past sales, order data, etc.) cannot understand changes in unpredictable dynamic market conditions [25]. Traditional methods do not consider changes in product life cycles, promotional activities, a viral social media campaign or the impact of competitors. Therefore, these methods may result in inaccurate forecasting that can affect the entire retail supply chain, causing out-of-stocks, overstock, decreased profitability, and negative effects on customer satisfaction and loyalty [26]. Retailers have begun to adopt technologies that support artificial intelligence (AI), especially to solve the food waste problem [27]. AI is using a machine's computational capabilities to imitate human intelligence [28]. AI and Machine Learning (ML), massive volumes of data, learn complex patterns and adapt to changing market conditions, thus being able to forecast demand more accurately [29]. AI-supported demand forecasting increases the accuracy of demand forecasting and helps predict changes in consumer behaviour more effectively through processing and analysing vast amounts of complex data. Apart from historical sales data, AI takes into account social media, online reviews, current market trends data and these data provides insights. Thus, AI tries to solve problems such as inventory, demand forecasting, stocking, order optimization and planning in the retail industry [13, 14]. According to McKinsey Digital, AI-powered forecasting reduces errors in the supply chain by 30–50%. This causes transportation and storage costs to decrease by 5–10% and supply chain management expenses to decrease by 25–40% [15]. It prevents food waste by ensuring the optimal amount of product is ordered through AI-based demand forecasting.

AI-based dynamic pricing is used to efficiently sell products approaching their expiration date at the most appropriate level and time. Finally, surplus food is effectively redistributed through AI-powered marketplaces [30]. AI is used in quality control and sourcing processes in retailing [31], thus eliminating inefficiencies in FSC [14]. Perishable foods have a significant impact on retailer revenues. A high market price for market products may result in a food surplus, while a low price indicates a low profit margin. While many retail companies in the market reduce the prices of their products whose expiration date is approaching, they adopt a single price strategy for fresh products [16]. Dynamic pricing strategy sets optimal product prices to maximize the seller's profit, reducing the generation of stock outs and food waste. Retailers use dynamic pricing models that discount products as they end their shelf life to eliminate low-selling or slow-growing products that often go to waste [32]. A dynamic pricing strategy based on food shelf life based on data obtained from Radio Frequency Identification [RFID] tags that measure product quality [33].

Dynamic pricing strategy uses real-time Internet of Things [IoT] sensor data at different sales season on food waste of a retailer [16]. Additionally, a retailer can independently determine the price of any product based on its freshness score, regardless of store central decision [34]. Based on this data, retailers will be able to update prices according to the quality characteristics of the products, and food waste will be minimized [35].

By taking into account real-time sales data, weather and seasonality, AI can predict the demand for each product more accurately by understanding and mapping the data

with new data added to the system. It can reduce food waste in stores, increase efficiency and increase sales based on more accurate demand forecasting [36]. Accurate demand forecasting is significant for inventory management and efficient purchasing. It provides agility in inventory management, allowing fewer out-of-stock products, more accurate shelf planning and greater customer satisfaction [14]. With the predictive ability provided by AI, various processes such as managing and optimizing the stock level within supply chains will be improved, resulting in cost reduction, time savings, inventory management, waste management and increased efficiency in the FSC [28, 37, 38]. Accurate predictive ability can also result in profitability, return on investment and provide businesses with energy and water consumption savings, adequate storage space, and reduce the operation's carbon footprint [11].

In utilizing AI, retailers need to access previous season historical data such as orders and inventory. However, some reasons make it difficult to benefit from AI in the retail industry. These challenges make the industry business priorities are different and there is no sufficient, accurate and consistent data regarding new product groups [36]. The traditional retail mindset to leverage AI solutions is that many products are offered to the consumer, regardless of how well the products sell. AI can increase sales by allocating products to shelf space based on sales data.

3 Artificial Intelligence Applications in Food Retail Sector

AI can be a powerful tool for food retailers to tackle waste and improve supply chain efficiency while contributing to a more sustainable business model. This section gives examples of retailers using AI for reducing food waste. The retailers included in the study are Walmart, Carrefour, Tesco and Aldi which are considered as largest and most famous food chain stores in the world in 2024 [39]. They are also recognized as companies that dominate the global food retailing industry [40]. Within the scope of the study, these companies were included in the study as the cases using artificial intelligence among the largest food retailers.

The case study analysis is used to construct the research, which looks at a variety of sources, including academic journals, industrial reports, and the official websites of food retailers. The ability to comprehend an intricate situation and provide an understandable description of elements and processes makes case studies the ideal tool for studying business situations in their micro- and macro-environment [41]. Case study approach enables a thorough understanding of the phenomenon being studied [42].

Walmart

Walmart's AI and ML-driven inventory management systems improve customer experience. This inventory management system ensures that Walmart customers can easily find what they want, increasing customer satisfaction. AI/ML models integrated into systems provide insights using historical data. The key strength of Walmart's use of AI/ML is its ability to exclude one-time abnormalities. However,

the always-learning system can optimize demand and reposition inventory to higher sales regions. With advanced learning systems, inputs are also used to help adapt to regional variations in needs, cultures and purchasing patterns [43].

In addition to using AI and ML techniques to predict when customers are most likely to shop and decide which items can be substituted for out-of-stock items, Walmart is using technology to shorten the time it takes to get fresh produce to customers. The company constantly monitors the selling amount of the products and follows this data throughout the entire supply chain [44].

Tesco

Tesco embraces the philosophy that the first step towards reducing is not creating waste. Redistributing products that are still consumable, redistributing leftover food as animal feed, and converting waste into energy are other steps followed. Tesco combats food waste by streamlining its operations using data and insights [45].

To make its warehouse operation more efficient, Tesco has opened highly automated urban fulfilment centres. Tesco also employs food rescue techniques to keep food from going to waste, AI-based technology that supports fresh inventory management of stores, and produce-coating technology to increase the shelf life of the items [46]. Tesco also establishes strong partnerships with its suppliers and works with them to ensure efficiency in food waste [47].

Carrefour

Carrefour always carries out technology management and digital transformation with a customer-oriented approach. In addition, the company prioritizes achieving productivity and competitive advantage by making its organizational structure agile to respond quickly to developments and in addition, giving importance to quality and sustainability in that the company attaches importance to. Carrefour uses technology extensively in food transformation. One of the company's main goals is to lead sustainable food consumption along with food production. For this purpose, by establishing close relationships and working together with producers, it supports them in making production more sustainable and at the same time maintaining food quality and making improvements [48].

Smart supply chain management is the key to avoiding inventory problems. All of these are included in the company's transformation plan [49]. In its journey to transform into a digital retail company, the company benefits from AI and technology in line with its sustainability goals. Innovative applications such as dynamic pricing are important in helping the company improve customer experience and increase efficiency [50]. Carrefour aims to all product pages with the help of generative AI. In this way, progress is made in defining products and providing more information to customers. In addition, generative AI has started to be used in the internal purchasing process. AI facilitates daily tasks such as preparing tender invitations and analysing bids. The company's digital and data culture makes it easier to take all these steps [51].

Aldi

Aldi is moving towards prioritizing the use of AI to manage its transportation operation better and strengthen this system Aldi also uses technology to optimize its supply chain by controlling freight rates, shipment handling and storage costs, and manual processes [52]. Furthermore, Aldi also uses AI in the fight against plastic waste. Used shopping bags and plastic packaging are gathered at collection points in front of the store, and the material flow is monitored with an AI-based system [53].

4 Conclusion

Food retailers utilize big data and data analytics techniques in predictive ML and AI to increase the predictive power of decision support systems and reduce food waste [6]. AI helps increase efficiency and sales and improves in-store operations with well-designed applications in replenishment and forecasting in the food supply chain [54].

Waste can occur if there is an ongoing promotion on the product due to unbalanced demand and sales exceed forecasts. AI quickly informs suppliers about this situation when sales exceed or fall short of the forecast to prevent this waste. This reveals that waste can be reduced by sharing information for both inventory and demand forecasts based on communication and coordination among supply chain members, and an optimized production with AI forecasts [21, 55]. Thus, when redundancy in the supply chain is prevented, less waste is occurred. Likewise, even if the demand forecast is accurate, implementing promotions and competitors simultaneously causes a product to be sold below the forecast and generate waste. AI can especially detect perishable products, so some of these products are sold cheaper in stores to prevent waste [56]. Thus, timely delivery reduces the risk of early ordering, especially perishable products, remaining in stock until their expiration date [14].

Artificial intelligence's supply chain efficiency to food retailers also contributes to sustainability and increases customer satisfaction. The case studies of food retailers included in the study also underline the benefits of AI to food retailers. In this context, the results of the case studies have indicated that artificial intelligence is used in different areas such as inventory management, demand forecasting, stocking, order-order optimization, supply chain risk management and planning, which are mentioned in different academic studies [11, 13, 14] to reduce food waste. As stated in the case-studies, AI used in quality control and internal purchasing processes in the retail industry [31], allowing fewer out-of-stock products, more accurate shelf planning, eliminating inefficiencies in the food supply chain [14]. Food retailers are increasingly investing in technologies that will assist them in food inventory management.

In food retailing, where competition is increasing, benefiting from artificial intelligence-based technologies in order to both increase operational efficiency and

maximize consumer satisfaction has become one of the important competitive advantages of companies within the sector. In the digital era where predicting consumer behaviour is more important than ever and data is readily available, it is recommended for food retailers to monitor and integrate these technologies into their systems to prevent overstocking and stockouts, avoid product waste, and increase customer revisit intentions.

One of the limitations of the study is that we examined only four food retailers in our study. However, these retailers stand out as strong examples of using AI technologies in food retailing. Therefore, these four retailers provide an important indication of how diverse and effective the use of AI in the industry can be. Future studies examining different food retailers will allow a broader perspective and comprehensive results to be obtained. Additionally, comparing retailers across different industries can reveal sectoral differences and similarities, contributing to solid inferences about broader trends. Another limitation is not focusing in detail on the entire range of artificial intelligence applications for each retailer, as the goal was to compare multiple retailers and highlight general trends. Therefore, rather than delving deeply into every application of each retailer, we aimed to provide a broader perspective by examining the differences and similarities between the selected retailers. Future studies could focus more in-depth on the full range of AI applications of a specific retailer, which would allow for more detailed and specific insights.

References

1. Buisman, M.E., Haijema, R., Bloemhof-Ruwaard, J.M.: Discounting and dynamic shelf life to reduce fresh food waste at retailers. Int. J. Prod. Econ. **209**, 274–284 (2019)
2. Ramanathan, R., Duan, Y., Ajmal, T., Pelc, K., Gillespie, J., Ahmadzadeh, S., Condell, J., Pelc, K., Gillespie, J., Hermens, I., Ramanathan, U.: Motivations and challenges for food companies in using IoT sensors for reducing food waste: some insights and a road map for the future. Sustainability **15**(2), 1665 (2023)
3. Gustavo, J.U., Jr., Trento, L.R., de Souza, M., Pereira, G.M., de Sousa Jabbour, A.B.L., Ndubisi, N.O., Jabbour, C.J.C., Borchardt, M., Zvirtes, L.: Green marketing in supermarkets: conventional and digitized marketing alternatives to reduce waste. J. Clean. Prod. **296**, 126531 (2021)
4. Adam, A.: Drivers of food waste and policy responses to the issue: the role of retailers in food supply chains (No. 59/2015). Working Paper (2015)
5. EU: Directive (EU) 2018/851 of the European Parliament and of the Council of 30 May 2018 Amending Directive 2008/98/EC on Waste. Off. J. Eur. Union, L 150/109 (2018)
6. Zhang, J., Wedel, M., Bloem, M.W.: Mitigating food waste in the retail supply chain: marketing solutions. J. Sustain. Market. **3**(2), 87–97 (2022)
7. Huang, I.Y., Manning, L., James, K.L., Grigoriadis, V., Millington, A., Wood, V., Ward, S.: Food waste management: a review of 'retailers' business practices and their implications for sustainable value. J. Clean. Prod. **285**, 125484 (2021)
8. European Court of Auditors (ECA): Combating Food Waste: an opportunity for the EU to improve the resource-efficiency of the food supply chain. European Court of Auditors, Luxembourg (2016)
9. Kor, Y.Y., Prabhu, J., Esposito, M.: How large food retailers can help solve the food waste crisis. Harvard Bus. Rev. **19** (2017)

10. Teller, C., Holweg, C., Reiner, G., Kotzab, H.: Retail store operations and food waste. J. Clean. Prod. **185**, 981–997 (2018)
11. Nascimento, A.M., Queiroz, A., de Melo, V.V., Meirelles, F.S.: Applying artificial intelligence to reduce food waste in small grocery stores. In: Information Systems in Latin America (2022)
12. Kuleshov, V., Seymour, H., Nemer, D., Meador, S., Fenner, N., Schwartz, M.: Towards a sustainable food supply chain powered by artificial intelligence. In: Proceedings of the 36 th International Conference on Machine Learning, Long Beach, California (2019)
13. Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., Fischl, M.: Artificial intelligence in supply chain management: a systematic literature review. J. Bus. Res. **122**, 502–517 (2021)
14. Hermansson, S., Lindgren, S.: Exploring Artificial Intelligence in Food Supply Chains: A Qualitative Study on the Effects of AI from the 'Retailers' Perspective. Umeå School of Business, Economics, and Statistics Umeå University (2024)
15. Amar, J., Rahimi, S., Surak, Z., von Bismarck N.: AI-driven operations forecasting in data-light environments (2022). https://www.mckinsey.com/capabilities/operations/our-insights/ai-driven-operations-forecasting-in-data-light-environments
16. Kayikci, Y., Demir, S., Mangla, S.K., Subramanian, N., Koc, B.: Data-driven optimal dynamic pricing strategy for reducing perishable food waste at retailers. J. Clean. Prod. **344**, 131068 (2022)
17. FAO: Food Wastage Footprint (2022). Retrieved from https://www.fao.org/nr/sustainability/food-lossand-waste/en/
18. United Nations Environment Programme (UNEP): Reducing Consumer Food Waste Using Green And Digital Technologies (2021). Rertieved from https://www.yunbaogao.cn/index/partFile/5/unep/2022-04/5_31028.pdf
19. Vadakkepatt, G.G., Winterich, K.P., Mittal, V., Zinn, W., Beitelspacher, L., Aloysius, J., Ginger, J., Reilman, J.: Sustainable retailing. J. Retail. **97**(1), 62–80 (2021)
20. Marchetti, E., Antonelli, C., Sgarbi, F.: Addressing Food Waste in the retail sector. Institute for European Environmental Policy (2024)
21. Beullens, P., Ghiami, Y.: Waste reduction in the supply chain of a deteriorating food item—impact of supply structure on retailer performance. Eur. J. Oper. Res. **300**(3), 1017–1034 (2022)
22. Riesenegger, L., Hübner, A.: Reducing food waste at retail stores—an explorative study. Sustainability **14**(5), 2494 (2022)
23. Felix, P.: Data-driven retail food waste reduction: a comparison of demand forecasting techniques and dynamic pricing strategies. Master's thesis, University of Twente (2018)
24. Haijema, R., Minner, S.: Improved ordering of perishables: the value of stock-age information. Int. J. Prod. Econ. **209**, 316–324 (2019)
25. Aichner, T., Santa, V.: Demand forecasting methods and the potential of machine learning in the FMCG retail industry. In: Serving the Customer: The Role of Selling and Sales, pp. 215–252. Springer Fachmedien Wiesbaden, Wiesbaden (2023)
26. Mitra, A., Jain, A., Kishore, A., Kumar, P.: A comparative study of demand forecasting models for a multi-channel retail company: a novel hybrid machine learning approach. In: Operations research forum, vol. 3, No. 4, p. 58. Springer, Cham (2022)
27. Borges, A.F.S., Laurindo, F.J.B., Spínola, M.M., Gonçalves, R.F., Mattos, C.A.: The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. Int. J. Inf. Manage. **57**, 1–16 (2021)
28. Riahi, Z., Hong, S.J., Rhim, J.W., Shin, G.H., Kim, J.T.: High-performance multifunctional gelatin-based films engineered with metal-organic frameworks for active food packaging applications. Food Hydrocolloids **144**, 108984 (2023)
29. Kondapaka, K.K.: Advanced artificial intelligence techniques for demand forecasting in retail supply chains: models, applications, and real-world case studies. Afr. J. Artif. Intell. Sustain. Dev. **1**(1), 180–218 (2021)
30. Anggraeni, M.C., Silaban, C.A., Anggreainy, M.S., Cahyadi, E.: Role of artificial intelligence in the management of food waste. In: 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS), pp. 1–6. IEEE, USA (2021)

31. Allal-Chérif, O., Climent, J.C., Berenguer, K.J.U.: Born to be sustainable: how to combine strategic disruption, open innovation, and process digitization to create a sustainable business. J. Bus. Res. **154**, 113379 (2023)
32. Adenso-Díaz, B., Lozano, S., Palacio, A.: Effects of dynamic pricing of perishable products on revenue and waste. Appl. Math. Model. **45**, 148–164 (2017)
33. Li, D., Wang, X.: Dynamic supply chain decisions based on networked sensor data: an application in the chilled food retail chain. Int. J. Prod. Res. **55**(17), 5127–5141 (2017)
34. Al-Turjman, F.: Cognitive Sensors and IoT: Architecture, Deployment, and Data Delivery. CRC Press (2017)
35. Dora, M., Biswas, S., Choudhary, S., Nayak, R., Irani, Z.: A system-wide interdisciplinary conceptual framework for food loss and waste mitigation strategies in the supply chain. Ind. Mark. Manage. **93**, 492–508 (2021)
36. Pasific Coast Food Waste Commitment: Using Artificial Intelligence to Reduce Food Waste in Grocery Retail Progress on the Path to Cut Food Waste in Half by 2030 (2022)
37. Perano, M., Cammarano, A., Varriale, V., Del Regno, C., Michelino, F., Caputo, M.: Embracing supply chain digitalization and unphysicalization to enhance supply chain performance: a conceptual framework. Int. J. Phys. Distrib. Logist. Manag. **53**(5/6), 628–659 (2023)
38. Hangl, J., Krause, S., Behrens, V.J.: Drivers, barriers and social considerations for AI adoption in SCM. Technol. Soc. **74**, 102299 (2023)
39. https://www.virasmart.co/en/the-largest-and-most-famous-food-chain-stores-in-the-world-in-2024/
40. Adams, H.S.: Top 10: Largest Food Retail Chains (2024). Retrieved from https://fooddigital.com/retail/top-10-largest-food-retail-chains
41. Eriksson, P., Kovalainen, A.: Qualitative methods in business research: a practical guide to social research. SAGE Publications (2015)
42. Kulikovskaja, V., Aschemann-Witzel, J.: Food waste avoidance actions in food retailing: the case of Denmark. J. Int. Food Agribus. Market. **29**(4), 328–345 (2017)
43. Musani, P.: Decking the aisles with data: how Walmart's AI-powered inventory system brightens the holidays (2023). https://tech.walmart.com/content/walmart-global-tech/en_us/blog/post/walmarts-ai-powered-inventory-system-brightens-the-holidays.html
44. Silverstein, S.: Walmart uses AI to predict demand (2020). Retrieved from https://www.supplychaindive.com/news/walmart-grocery-AI-demandoperations/585424/
45. TESCO (2024). Retrieved from https://www.tescoplc.com/sustainability/planet/food-waste/tackling-food-waste-in-our-own-operations
46. LS retail (2024). Retrieved from https://www.lsretail.com/resources/ways-uk-supermarket-brands-are-using-technology-to-drive-retail-transformation
47. Broekmeulen, R.A., van Donselaar, K.H.: Managing Retail Food Waste & Markdowns—Increasing Sales and Reducing Waste in the Fresh Supply Chain. ECR Retail Loss Group, Brussels, Belgium (2016)
48. Woods, A.: Carrefour: A customer-centric tech transformation. CPO Straregy Blog (2021)
49. CARREFOUR (2019). Retrieved from https://www.carrefour.com/en/news/artificial-intelligence-optimise-carrefours-supply-chain
50. Lee, G.: Realizing AI in Retailing (2024). Retrieved from https://erp.today/realizing-ai-in-retail/
51. CARRFOUR (2023). Retrieved from https://www.carrefour.com/en/news/2023/carrefour-integrates-openai-technologies-and-launches-generative-ai-powered-shopping
52. Stroh, K.: Aldi turns to AI to strengthen freight management. SupplyChainDive (2024)
53. Grace, R.: ALDI UK using AI to provide visibility to its flexible plastic packaging efforts. Interpack. Retrieved from https://www.interpack.com/en/Media_News/SPOTLIGHT_Magazine/Food_Industry_News/ALDI_UK_using_AI_to_provide_visibility_to_its_flexible_plastic_packaging_efforts
54. Guha, A., Grewal, D., Kopalle, P.K., Haenlein, M., Schneider, M.J., Jung, H., Moustafa, R., Hegde, D.R., Hawkins, G.: How artificial intelligence will affect the future of retailing. J. Retail. **97**(1), 28–41 (2021)

55. Annosi, M.C., Brunetta, F., Bimbo, F., Kostoula, M.: Digitalization within food supply chains to prevent food waste. drivers, barriers and collaboration practices. Ind. Mark. Manage. **93**, 208–220 (2021)
56. Stoyanov, T.: Forest Cooperatives and The Circular Bioeconomy. For. Sci. **60**. ISSN: 0861-007X (2024)

Pathway to Net-Zero Emissions in Supply Chain: A Comprehensive Analysis of Industry 4.0 Technologies and Emission Reduction Practices



Shobha Rathore, Rakesh Jain, and Gunjan Soni

Abstract This research endeavors to comprehensively examine the most recent scholarly contributions of implementing Industry 4.0 technologies within supply chain management and strategies aimed at emissions reduction with the objective of achieving net-zero emissions. We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method to systematically identify scholarly articles addressing Industry 4.0 technologies and emissions reduction methodologies for attaining net-zero objectives within supply chains. This investigation meticulously reviews existing literature, encompassing 65 articles thoroughly analysed, synthesized, and categorized across various dimensions (i.e., Industry 4.0 technologies, emission reduction strategies, type of industry, objectives, research methodologies, and key findings). The analysis identified 13 distinct technologies associated with Industry 4.0 and 13 corresponding strategies for emissions reduction, delineated by their fundamental characteristics. Furthermore, it investigates the industry 4.0 technologies and emission reduction practices (ERPs) deemed essential for realizing a net-zero supply chain. The study also elucidates these technologies' potential to mitigate resource consumption, energy utilization, and emissions of carbon or greenhouse gases. Subsequently, it examines the advantages and challenges associated with the concurrent adoption of I4.0 technologies and practices to achieve net-zero objectives at the organizational level. This investigation provides a simultaneous evaluation of various practices for emissions reduction in the supply chain alongside Industry 4.0, thus offering an in-depth analysis of two distinct yet interrelated approaches.

Keywords Industry 4.0 technology · Supply chain emissions · Net-zero supply chain · Net zero emissions · Emissions reduction practices

S. Rathore (✉) · R. Jain · G. Soni
Department of Mechanical Engineering, Malaviya National Institute of Technology, Jaipur, India
e-mail: 2022rme9072@mnit.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_37

1 Introduction

Global surface temperature is predicted to increase by 3–4 °C by 2100 compared to averages from 1986 to 2005 [9]. Researchers estimate that by 2052, the mean close-surface air average temperature will have increased by 1.5 °C if current trends continue [31]. Global greenhouse gas (GHG) emissions increased at the beginning of the twentieth century due to a dramatic rise in fuel used by various industrial operations and processes [19]. Supply chain (SC) emissions are also significantly higher than other direct and indirect emissions. With the growing urgency to address the global climate crisis, there has been an increasing emphasis on the critical role that supply chains must play in transitioning towards net-zero emissions. [18]. Recent studies have highlighted the significant environmental impact of supply chain activities, with the supply chain sector contributing to a substantial proportion of global greenhouse gas emissions [6]. This underscores the need for supply chain management practices prioritizing sustainability and environmental responsibility.

Developing a framework that satisfies the organization's sustainability and profitability objectives is the primary hurdle in establishing a net zero supply chain (NZSC). Net-zero business strategies can also aim for zero emission, zero waste, and so on, in addition to net-zero [10]. This broad definition makes it difficult for organizations to create successful green policies that simultaneously meet multiple requirements [20, 29, 30]. The adoption of NZSC procedures has become crucial due to the attainment of the objective of NZE in several sectors and the role of global governments [10].

In the rapidly evolving landscape of modern industry, Industry 4.0 (I4.0) has emerged as a beacon of efficiency and innovation [23]. The shift toward a NZSC is imperative for addressing the ecological challenges posed by traditional manufacturing practices and transportation [12]. Adopting I4.0 into supply chain management (SCM) offers a promising avenue for achieving NZE as we navigate the pressing global concern of climate change. I4.0 has led to a shift in how businesses and researchers focus on the cutting edge of the revolution in technology. The Internet of Things (IoT), blockchain (BCT), cyber-physical systems (CPS), big data analytics (BDA), cloud computing (CC), artificial intelligence (AI), and additive manufacturing (AM) are some of the key concepts that it incorporates both within and between businesses [26, 27]. According to multiple studies, I4.0 technologies are expected to generate interconnected manufacturing processes, allow real-time SC monitoring and optimization, and allow customization of production, goods, and services across the supply chain [2, 3]. The industry's GHG emissions cannot be considerably decreased by a single change [10]. It is necessary to research I4.0 technologies' potential from a net zero perspective.

The main objective of Industry 4.0, which is truly revolutionary, is to drive businesses to fulfill the essential requirements for achievement. Increasing efficiency, lowering resource and energy consumption, enhancing customer service, automating development, and integrating industrial processes are the most popular strategic objectives for using Industry 4.0 technology. This study addresses two objectives

by investigating Industry 4.0's contribution to attaining zero net emissions in the supply chain. The researcher can better grasp the primary goal of this study by consulting the research questions that came before it.

RQ1: What significant contributions do I4.0 technologies make to attaining net zero in a supply chain?
RQ2: What are the best emission reduction practices to create a net-zero supply chain?

In this section, we provide an overview of the I4.0 technologies that are facilitating the transition to net-zero supply chains, their urgent necessity, and the structure of this paper. This paper's subsequent sections are organized as follows: Sect. 2 presents the demographic analysis of the chosen articles and the review methodology. Section 3 presents the review's findings in answer to the research questions. Research gaps and their consequences for future research are thoroughly examined in Sect. 4. In Sect. 5, the conclusions are presented.

2 Methodology

This research uses the Prisma framework to perform a systematic literature review (SLR). PRISMA offers a replicable research approach demonstrating the entire review process and a great chance to discuss emerging themes [10]. This strategy was selected because it could answer the review questions and provide information on popular, current ideas and market trends. Tranfield [28] provided a four-stage framework for conducting an SLR, which we used in our analysis. There are four stages: Identification, screening, eligibility, and inclusion. A detailed description of the Prisma Framework is provided in Fig. 1.

3 Synthesis and Findings

3.1 Demographic Results

Year-wise Publications. An overview of the selected papers organized by year of publication is given in this section. Figure 2 shows that the number of articles has significantly increased in the last few years due to the rise in interest in I4.0 technologies and net-zero or sustainability goals in industry by academics and researchers.

Contribution by Country. Fig. 3 indicates that the nations with the most significant contributions in this field are the United States, China, Denmark, England, and India, based on the count of articles.

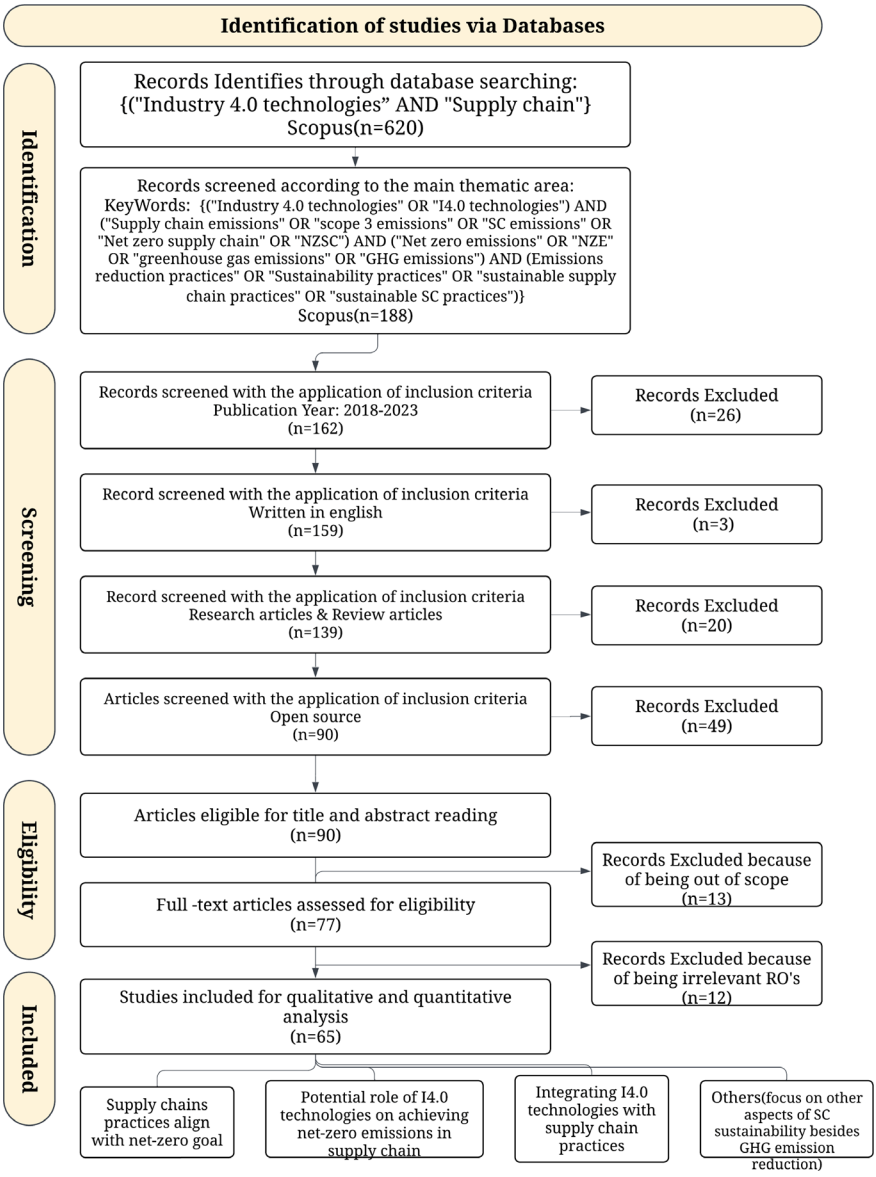


Fig. 1 Prisma framework for literature selection

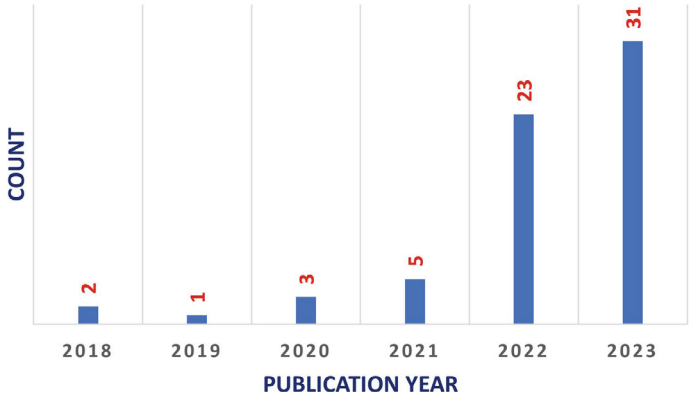


Fig. 2 Year-wise publication data

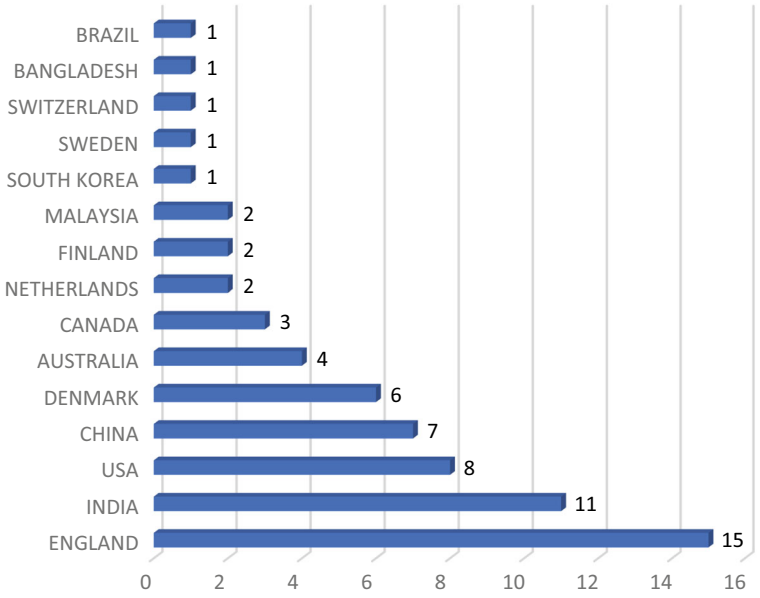


Fig. 3 County-wise publication details

Articles Focused on SDGs. Among all 17 Sustainable Development Goals, our selected studies mainly focused on SDG12: "Responsible Consumption and Production", SDG13: "Climate Action", SDG7: "Affordable and Clean Energy", SDG9: "Industry Innovation and Infrastructure", and SDG11: "Sustainable Cities and Communities". Figure 4 illustrates that the articles focus on responsible consumption sources, production processes, climate action, and clean energy.

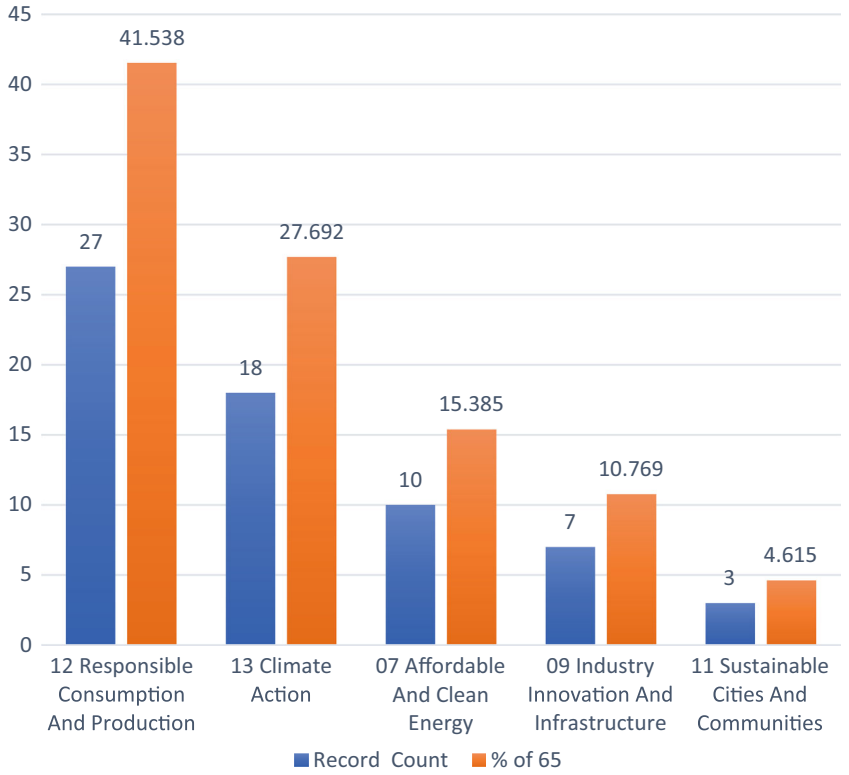


Fig. 4 SDGs focused articles

This study focuses on two terms, namely Industry 4.0 technologies and net-zero emissions in the supply chain. For Industry 4.0 technologies, we have identified Major supportive technologies of Industry 4.0 for Supply chain emission reduction: IoT, AI-ML, BDA, CPS, BCT, AR, AM, AR, CC, EC, DT, CS, and IoS. Moreover, a net-zero supply chain requires such practices: Net-zero economic practices (NZEPs), Sustainability practices (SPs), Carbon capture and sequestration practices (CCSPs), Carbon offsetting practices (COPs), Circular economy practices (CEPs), Decarbonization practices (DCPs), Carbon accounting practices (CAPs), GSCM practices (GSCMPs), Renewable energy integration practices (REIPs), Lean-green practices (LGPs), Stakeholder collaboration practices (SCPs), Resilience enhancement practices (REPs), Environmental, Social and Governance Practices (ESGPs) to reduce emissions.

3.2 Potential Contribution of Industry 4.0 Technologies Driving Net-Zero Goals

First, we present the significant contribution of technologies of Industry 4.0. IoT connects physical devices to the internet, enabling them to communicate and share data. In supply chains, IoT devices can monitor energy usage, track emissions, and optimize resource allocation [22]. For example, smart sensors in manufacturing plants can detect inefficiencies and provide data to reduce energy consumption and emissions [25].

In order to optimize supply chain operations, AI and ML systems examine enormous volumes of data. They can predict demand, streamline logistics, and enhance production efficiency, all of which contribute to reducing carbon footprints. AI-powered systems can also identify patterns and suggest improvements for energy savings and emissions reduction [6]. BCT offers an unchanging, visible ledger for monitoring and confirming supply chain transactions. It enhances traceability and accountability, ensuring sustainable practices and reducing fraud. Blockchain can track the carbon footprint of products from production to delivery, helping companies identify and address emission hotspots in their supply chain [26]. BDA provides a comprehensive view of supply chain operations. By analyzing large datasets, companies can uncover patterns and trends that lead to more informed, eco-friendly decisions [25]. For instance, data analytics can help identify areas where energy consumption can be minimized and waste can be reduced.

Real-time data collection and automation are made possible by CPS's integration of digital control systems with physical operations. This integration helps optimize supply chain operations and reduce inefficiencies. CPS can automate the adjustment of machinery settings based on real-time data to maintain optimal operating conditions, reducing unnecessary energy use and emissions [1]. Advanced robotics automate repetitive and labor-intensive tasks with high precision, reducing errors and waste. Robots can operate continuously without fatigue, increasing production speed and efficiency. Automation improves operational efficiency and reduces the energy required for production processes [2, 3].

Additive manufacturing, or 3D printing, produces objects layer by layer, minimizing waste compared to traditional manufacturing methods. It also allows for on-demand production, reducing the need for excess inventory and associated emissions. 3D printing parts locally reduces the need for transportation, cutting down on fuel use and emissions from shipping goods long distances [21]. Augmented Reality (AR) enhances real-world environments with digital information, aiding in training, maintenance, and repair activities. This can improve efficiency and reduce errors, lowering energy consumption and emissions. AR can guide technicians through complex repairs, reducing the time and resources needed to fix equipment and minimizing downtime and energy use [21].

Cloud Computing (CC) provides scalable and efficient data storage and processing solutions, reducing the need for energy-intensive on-site servers. It also enables better data sharing and collaboration across supply chain stakeholders [10]. By moving data

storage and processing to the cloud, companies can reduce the energy consumption of maintaining physical data centers [11]. Edge Computing (EC) processes data near the source point, reducing the need for data transmission to centralized servers. This lowers latency, enhances real-time decision-making, and reduces energy consumption associated with data transport. In a smart factory, edge computing can automatically analyze sensor data on-site to optimize machinery operations, reducing energy use and emissions.

Digital twins (DT) are digital replicas of real physical assets, processes, and systems. They allow for real-time monitoring and optimization, predicting potential issues before they become problems. Digital twins help optimize resource use and reduce emissions by simulating different scenarios [25]. Cyber Security (CS) measures protect IoT devices, AI algorithms, and digital networks from threats. Secure systems ensure data integrity, confidentiality, and availability, which is crucial for maintaining sustainable and resilient supply chains. Internet of Services (IoS) integrates services across the supply chain network, enhancing interoperability and efficiency. IoS supports sustainable practices through improved resource allocation and reduced operational redundancies by enabling seamless communication and coordination.

3.3 Emission Reduction Practices

This section presents different emissions reduction practices, explaining their different attributes. NZEPs involve economic strategies such as carbon pricing, carbon trading, and financial incentives to reduce greenhouse gas emissions across supply chains by imposing costs on carbon emissions or rewarding emission reductions [2, 3]. To reduce environmental damage and ensure long-term ecological balance, supply chains that use sustainable sourcing practices prioritize increasing resource efficiency, cutting waste, and promoting sustainable sourcing [26].

CCSPs involve technologies and practices that capture CO2 emissions from industrial processes or directly from the atmosphere, followed by storage underground or in other long-term storage solutions to prevent its release into the atmosphere [4]. By supporting initiatives like reforestation and renewable energy projects that cut or eliminate greenhouse gas emissions elsewhere, COPs make up for emissions that cannot be avoided [16]. CEPs seek to reduce waste and maximize resource efficiency while reducing environmental impact by encouraging actions like recycling, reusing, and remanufacturing within supply chains [5]. Implementing circular economy practices can help achieve self-sustainability regarding energy and materials and pave the way for net-zero manufacturing. [17].

DCPs involve strategies like electrification of processes, fuel switching to renewable sources, and optimizing energy-intensive operations to reduce the carbon intensity of supply chain activities [35]. CAPs encompass methods and tools for

measuring, reporting, and verifying greenhouse gas emissions for a particular business, enabling organizations to track emissions and identify GHG reduction opportunities. GSCMPs integrate environmental considerations into supply chain processes, including sustainable procurement, green logistics, and eco-design, to reduce environmental impact and enhance sustainability [7]. REIPs incorporate renewable energy sources such as solar, wind, and hydroelectric power into supply chain networks to reduce reliance on fossil fuels and lower greenhouse gas emissions.

LGPs combine lean manufacturing principles with green practices to optimize resource use, minimize waste, and reduce emissions across supply chain processes, achieving efficiency gains and environmental benefits (Yadav et al. 2023). SCPs involve partnerships, joint ventures, and multi-stakeholder initiatives to foster shared responsibility and collective action toward reducing supply chain emissions and achieving sustainability goals. REPs focus on building resilience within supply chains through risk assessment, adaptive planning, and investment in robust infrastructure to mitigate climate impacts and ensure continuity of operations. ESGPs encompass broader sustainability practices integrating environmental protection, social equity, and governance transparency into business strategies, impacting business operations toward net-zero goals.

4 Discussion

We found that companies are focusing on adopting essential emission-reduction practices. It shows the excellent reputation of the organization because of its net-zero concern. 26 (40%) out of 65 papers specifically discussed practices. On the other hand, 9 (~ 14%) out of 65 papers specifically discussed I4.0 T contribution and 30 (~ 46%) out of 65 discussed the I4.0 T and ERPs together regarding challenges, benefits, and issues related to NZSC. We can see from this data that concurrent study has a significant contribution, but more appropriate solutions are required in this field to attain the net zero target by 2050. This study identifies substantial insights and challenges of these existing studies.

The literature review concludes that CCSPs/CCUPs, CEPs, DCPs, and REIPs significantly contribute to reaching net-zero emissions in the supply chain. Nonetheless, many factors influence how Industry 4.0 technologies are implemented in businesses. Changing the company's current production process will cost enormous money. Second, we discovered that more significant information on the positive aspects of Industry 4.0 is needed, particularly for SMEs. It is, therefore, essential to train staff members on Industry 4.0 technologies, how to install them, and how to utilize them. Each Industry 4.0 technology, according to our investigation, has the potential to achieve net zero throughout the whole supply chain network. Moreover, I4.0 technologies significantly contribute to increased energy efficiency, real-time tracking of energy use and emission data, SC process optimization to reduce resource use, transparency of SC, and digitization to achieve net-zero targets.

4.1 Implications

This study highlights how Industry 4.0 technologies enable supply chain digitization and automation, which is essential for organizations aiming for net-zero emissions. Managers should train staff to adopt these digital practices, especially in SMEs where change is often resisted. The findings provide valuable guidance for companies, policymakers, and researchers, emphasizing collaboration with research institutions to stay updated on advancements toward NZSC. The study shows that key Industry 4.0 technologies like BCT, CPS, AI-ML, and IoT have reached maturity and are ready for enterprise adoption, enhancing value chains across sectors. The findings combine Industry 4.0 with ERP systems in a framework that clarifies each technology's role in sustainable supply chains, offering a roadmap for businesses.

4.2 Research Gaps, Limitations, and Future Research Opportunities

The study found two significant gaps in the supply chain's achievement of net zero. First, most research used quantitative data-gathering techniques such as modelling and survey approaches. This raises concerns about the practicality of the methods employed to reach the net zero target. Second, we only found a few studies that employed cross-sectional data; no longitudinal data articles were found for this study. Research gaps in this area are discussed in detail below in Table 1.

Research Limitations. This study only considered research articles and review papers, resulting in fewer papers being reviewed. The study's data were gathered from the Scopus database, although it is possible that some pertinent publications that should have been included were not indexed in Scopus. Consequently, future research endeavors might consider a more comprehensive range of papers. Second, due to the rapid speed of technological development in this area, research that is part of the review may need to encompass the most recent uses or implications of these technologies in SCs.

Scope in Future Research. The main emphasis of the upcoming research will be on implementing I4.0 technologies in various supply chain strategies: NZEPs, CEPS, SCPs, REIPs, DCPS, and others (discussed above), align with the net-zero emission goal. Future studies can also investigate other digital technologies (i.e., drones, RFID, intelligent factories) and ERPs and their potential to achieve net zero. Future research should continue to see how these identified practices are potentially helpful in different sectors like agriculture, food, and textile. We found that most SMEs must understand Industry 4.0 as a predictor of achieving net zero so future studies can discover opportunities for I4.0 technologies to fulfil the set targets of net zero in developing nations where SMEs comprise most firms.

Table 1 Identified research gaps

Gap related to	Identified gaps	Reference
I4.0 technology implementation	Lack of Comparative analysis of emerging technologies for successful net-zero implementation Complexity in the integration of emerging technologies	Agrawal et al. [2, 3]
Emission reduction Practice Adoption	Evaluating specific emission reduction practices at each stage of the supply chain is needed A comparative analysis of such practices is needed for a net-zero revolution There is a need for frameworks integrating different I4.0 technologies with emission reduction practices in the SC in depth	Chen et al. [5], Grassauer et al. [12], Hammer et al. [14], Khorasani et al. [16], Kumar et al. [17]
Benefits of I4.0 and ERPs integration	Lack of analysis of benefits from combining I4.0 technologies and ERPs and their dependence	El Jaouhari et al. [10], Wang et al. [29, 30], Yadav et al. [32]
Challenges of I4.0 and ERPs integration	Need to assess organizations' concurrent challenges in adopting I4.0 technologies and net-zero practices Analysis of more business initiatives and tactics needed to tackle/control GHG emissions	Chhabra and Singh [7], El Jaouhari et al. [10], Wang et al. [29, 30], Yadav et al. [32]

5 Conclusions

In order to investigate the role of I4.0 technologies in achieving net-zero emissions in the supply chain, this article examines 65 studies on the adoption of I4.0 technologies in SCs across several fields, including manufacturing, dairy, construction, and automotive. Given that previous models primarily focused on green or sustainable SCs, it emphasizes the necessity of a comprehensive integration of I4.0 technologies with ERPs for sustainability. The report highlights the advantages, difficulties, and the part that business and leadership play in the adoption of 13 important ERPs and 13 important I4.0 technologies. The results of the study, which focused on changing industrial processes to be emission-free, demonstrated a positive correlation between I4.0 technologies and emission reduction practices. Future research should concentrate on creating comprehensive models that integrate I4.0 technologies with ERPs

across various SC phases to maximize emission reduction. Studies might also investigate how I4.0 adoption affects supply chain emissions over the long run and create plans to get over technological and financial obstacles.

References

1. Abosuliman, S.: Deep learning techniques for securing cyber-physical systems in supply chain 4.0. Comput. Electr. Eng. **107** (2023)
2. Agrawal, R., Priyadarshinee, P., Kumar, A., Luthra, S., Garza-Reyes, J. A., Kadyan, S.: Are emerging technologies unlocking the potential of sustainable practices in the context of a net-zero economy? An analysis of driving forces. Environ. Sci. Pollut. Res. (2023). https://doi.org/10.1007/s11356-023-26434-2
3. Agrawal, S., Tyagi, M., Mangla, S., Garg, R.: Unlocking circular business model avenues to achieve net-zero emissions: a model-driven approach grounded on inter-valued intuitionistic fuzzy sets. Ann. Oper. Res. (2023)
4. Becattini, V., Gabrielli, P., Mazzotti, M.: Role of carbon capture, storage, and utilization to enable a net-zero-CO_2 emissions aviation sector. Ind. Eng. Chem. Res. **60**, 6848–6862 (2021). https://doi.org/10.1021/acs.iecr.0c05392
5. Chen, Q., Feng, H.B., de Soto, B.G.: Revamping construction supply chain processes with circular economy strategies: a systematic literature review. J. Cleaner Prod. **335**, Article 130240 (2022). https://doi.org/10.1016/j.jclepro.2021.1302405
6. Chen, Y., Jin, S.: Artificial intelligence and carbon emissions in manufacturing firms: the moderating role of green innovation. Processes **11** (2023)
7. Chhabra, D., Singh, R.K.: Analyzing barriers to green logistics in context of circular economy and Industry 4.0 in the Indian manufacturing industry. Int. J. Logistics Res. Appl. (2022). https://doi.org/10.1080/13675567.2022.2134847
8. Ching, N.T., Ghobakhloo, M., Iranmanesh, M., Maroufkhani, P., Asadi, S.: Industry 4.0 applications for sustainable manufacturing: a systematic literature review and a roadmap to sustainable development. J. Clean. Prod. **334**, Article 130133 (2022). https://doi.org/10.1016/j.jclepro.2021.130133
9. Duchêne, F., Hamdi, R., Van Schaeybroeck, B., Caluwaerts, S., De Troch, R., de Cruz, L., Termonia, P.: Downscaling ensemble climate projections to urban scale: Brussels's future climate at 1.5 °C, 2 °C, and 3 °C global warming. Urb. Clim. **46**, 101319 (2022). https://doi.org/10.1016/j.uclim.2022.101319
10. El Jaouhari, A., Arif, J., Samadhiya, A., Kumar, A.: Net zero supply chain performance and industry 4.0 technologies: Past review and present introspective analysis for future research directions. Heliyon **9**(11), e21525 (2023). https://doi.org/10.1016/j.heliyon.2023.e21525
11. Gammelgaard, B., Nowicka, K.: Next generation supply chain management: the impact of cloud computing. J. Enterp. Inf. Manag. (2023)
12. Grassauer, F., Arulnathan, V., Pelletier, N.: Towards a net-zero greenhouse gas emission egg industry: a review of relevant mitigation technologies and strategies, current emission reduction potential, and future research needs. Renew. Sustain. Energy Rev. **181**, 113322 (2023). https://doi.org/10.1016/j.rser.2023.113322
13. Grandsir, C., Falagán, N., Alamar, M.C.: Application of novel technologies to reach net-zero greenhouse gas emissions in the fresh pasteurised milk supply chain: a review. Int. J. Dairy Technol. **76**(1), 38–50 (2023). https://doi.org/10.1111/1471-0307.12926
14. Hammer, A.J., Millar, C., Hennige, S.J.: Reducing carbon emissions in aquaculture: using carbon disclosures to identify unbalanced mitigation strategies. Environ. Impact Assess. Rev. **96**, 106816 (2022). https://doi.org/10.1016/j.eiar.2022.106816

15. Hasan, M.F., Zantye, M.S., Kazi, M.K.: Challenges and opportunities in carbon capture, utilization and storage: a process systems engineering perspective. Comput. Chem. Eng. **166**, 107925 (2022)
16. Khorasani, M., Sarker, S., Kabir, G., Ali, S.M.: Evaluating strategies to decarbonize oil and gas supply chain: implications for energy policies in emerging economies. Energy **258**, 124805 (2022). https://doi.org/10.1016/j.energy.2022.124805
17. Kumar, R., Gupta, S., Rehman, U.: Circular economy a footstep toward net zero manufacturing: critical success factors analysis with case illustration. Sustainability **15** (2023)
18. Lazar, S., Klimecka-Tatar, D., Obrecht, M.: Sustainability orientation and focus in logistics and supply chains. Sustainability **13** (2021)
19. Lei, T., Guan, D., Shan, Y., Zheng, B., Liang, X., Meng, J., Zhang, Q., Tao, S.: Adaptive CO_2 emissions mitigation strategies of global oil refineries in all age groups. One Earth **4**, 1114–1126 (2021). https://doi.org/10.1016/j.oneear.2021.07.009
20. Liu, B., De Giovanni, P.: Green process innovation through Industry 4.0 technologies and supply chain coordination. Ann. Oper. Res. (2019). https://doi.org/10.1007/s10479-019-03498-3
21. Marinagi, C., Reklitis, P., Trivellas, P., Sakas, D.: The impact of Industry 4.0 technologies on key performance indicators for a resilient supply chain 4.0. Sustainability **15** (2023)
22. Mastos, T.D., Nizamis, A., Vafeiadis, T., Alexopoulos, N., Ntinas, C., Gkortzis, D., ... Tzovaras, D: Industry 4.0 sustainable supply chains: an application of an IoT enabled scrap metal management solution. J. Clean. Prod. **269**, 122377 (2020)
23. Naseem, M.H., Yang, J.Q.: Role of Industry 4.0 in supply chains sustainability: a systematic literature review. Sustainability **13**(17), 9544 (2021). https://doi.org/10.3390/su13179544
24. Peng, W., Xin, B.G., Xie, L.: Optimal strategies for production plan and carbon emission reduction in a hydrogen supply chain under cap-and-trade policy. Renew. Energy **215**, 118960 (2023). https://doi.org/10.1016/j.renene.2023.118960
25. Qi, Q.L., Tao, F.: Digital twin and big data towards smart manufacturing and Industry 4.0: 360 degree comparison. IEEE Access **6**, 3585–3593 (2018). https://doi.org/10.1109/access.2018.2793265
26. Saha, A.S., Raut, R.D., Yadav, V.S., Majumdar, A.: Blockchain changing the outlook of the sustainable food supply chain to achieve net zero? Sustainability **14**(24), 16916 (2022). https://doi.org/10.3390/su142416916
27. Sharma, M., Shah, J.K., Joshi, S.: Modeling enablers of supply chain decarbonisation to achieve zero carbon emissions: an environment, social and governance (ESG) perspective. Environ. Sci. Pollut. Res. **30**(31), 76718–76734 (2023). https://doi.org/10.1007/s11356-023-27480-6
28. Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. **14**(3), 207–222 (2003)
29. Wang, Y., Liu, J., Tang, X., Wang, Y., An, H., Yi, H.: Decarbonization pathways of China's iron and steel industry toward carbon neutrality. Resour. Conserv. Recycl. **194**, 106994 (2023). https://doi.org/10.1016/j.resconrec.2023.106994
30. Wang, Y., Yang, Y., Qin, Z., Yang, Y., Li, J.: A literature review on the application of digital technology in achieving green supply chain management. Sustainability **15**, 8564 (2023). https://doi.org/10.3390/su15118564
31. Wimbadi, R., Djalante, R.: From decarbonization to low carbon development and transition: a systematic literature review of the conceptualization of moving toward net-zero carbon dioxide emission (1995–2019). J. Clean. Prod. **256**, 120307 (2020). https://doi.org/10.1016/j.jclepro.2020.120307
32. Yadav, G., Luthra, S., Jakhar, S.K., Mangla, S.K., Rai, D.P.: A framework to overcome sustainable supply chain challenges through solution measures of industry 4.0 and circular economy: an automotive case. J. Clean. Prod. **254**, 120112 (2020). https://doi.org/10.1016/j.jclepro.2020.120112
33. Yadav, S., Samadhiya, A., Kumar, A., Majumdar, A., Garza-Reyes, J.A., Luthra, S.: Achieving the sustainable development goals through net zero emissions: innovation-driven strategies for transitioning from incremental to radical lean, green and digital technologies (2023). Resour. Conserv. Recycl. **197**, Article 107094. https://doi.org/10.1016/j.resconrec.2023.107094

34. Yang, L., Hu, Y., Wang, H., Li, C., Tang, B., Wang, B., Cui, H.: Uncertainty quantification of CO_2 emissions from China's civil aviation industry to 2050. J. Environ. Manag. **336**, 117624 (2023). https://doi.org/10.1016/j.jenvman.2023.117624
35. Zhao, C.P., Sun, J.J., Zhang, Y.: A study of the drivers of decarbonization in the plastics supply chain in the post-COVID-19 era. Sustainability **14**(23), 15858 (2022). https://doi.org/10.3390/su142315858

Integrating Automation and Orchestration in Security Incident Handling: A Review of SOAR Frameworks and Platforms



Pragya Rathore and Kishor Kolhe

Abstract Traditional manual incident response procedures are no longer sufficient in today's fast-paced digital environment due to the increasing complexity and frequency of cyber-attacks. Machine learning, automation, and orchestration are all combined in Security Orchestration, Automation, and Response (SOAR) platforms, which have become a potent tool for streamlining and improving incident response processes. In order to automate critical cybersecurity tasks, such as data collecting and analysis and real-time reaction execution, this review paper examines how SOAR systems help. Organizations can gather and compile data from several sources, analyze incidents with sophisticated algorithms, and carry out standardized responses with customizable playbooks by utilizing SOAR systems, which integrate a variety of security solutions. Organizations can improve overall security posture, eliminate human error, and cut reaction times by utilizing SOAR solutions, which automate complicated procedures, increase issue detection, and guarantee rapid resolution. This article examines the architecture, salient characteristics, and advantages of SOAR platforms and highlights their vital role in modernizing incident response, which in turn equips enterprises to remain ahead of the curve in terms of cyber risks.

Keywords SOAR · Playbook automation · Security operations · Incident response · Machine learning

1 Introduction

Organizations are facing an unparalleled increase in the complexity and frequency of cyber-attacks in the current digital era. Security teams now face a great deal of work due to the increasing complexity of cyberattacks and the volume of alerts produced by contemporary security solutions. Organizations are left susceptible to

P. Rathore (✉) · K. Kolhe
Department of Computer Engineering and Technology, School of Computer Science and Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India
e-mail: pragyarathore3@gmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_38

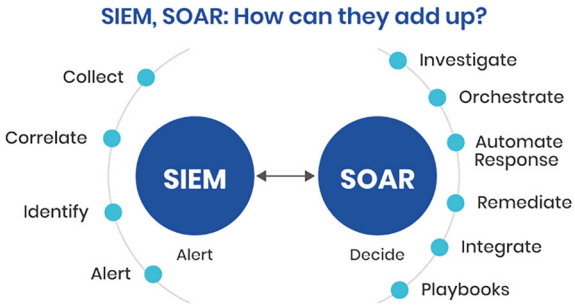


Fig. 1 SOAR versus SIEM

cybercriminals who are continuously refining their techniques, taking advantage of weaknesses, and focusing on a variety of attack surfaces. In these circumstances, the time-consuming and antiquated manual incident response techniques are no longer adequate due to their inability to keep up with the volume and velocity of contemporary cyberthreats. The increasing discrepancy between the threat picture and manual reaction capabilities emphasizes how urgently security operations need to be automated.

One innovative approach to addressing these issues is the use of Security Orchestration, Automation, and Response (SOAR) platforms. By automating tedious processes, integrating different security tools, and expediting incident response procedures, SOAR solutions are made to empower security professionals. The detection, analysis, and reaction stages of incident management can be automated by SOAR platforms by fusing machine learning algorithms, orchestration frameworks, and real-time threat intelligence. This guarantees a more uniform approach to managing security events and speeds up response times while also lowering the possibility of human error. The ineffectiveness of manual incident response in the context of constantly changing cyber threats is one of the most important cybersecurity challenges of our day, and SOAR systems' automation helps to solve it (Fig. 1).

1.1 Escalating Cyber Threats and the Need for Automation

A variety of attack vectors, including malware, ransomware, phishing, and distributed denial of service (DDoS) attacks, are among the many sophisticated cyberthreats that exist today. Threat actors are becoming more cunning than ever, deploying automated tools to take advantage of weaknesses in large quantities while adopting cutting-edge tactics to avoid detection. Manual incident response procedures are not only slow but also prone to error in such a volatile and hostile environment. Security teams frequently experience alert fatigue as a result of the exponential increase in security alert volume, which causes them to react slowly or overlook dangers. Furthermore,

manual processes usually need several steps and team engagement, which lengthens the time it takes to resolve issues and exposes enterprises to risk.

By streamlining and automating the whole crisis response process, SOAR platforms help to mitigate these issues. Tasks that are repetitive and time-consuming, such data collecting, log analysis, and basic threat mitigation activities, can be automated within SOAR systems, eliminating the need for human participation. The ability to concentrate on higher-priority duties, such threat hunting and strategic decision-making, is provided to security analysts by this time-saving method. The ability of an organization to quickly and effectively identify, address, and neutralize threats is greatly enhanced by SOAR systems' integration of automation into incident response.

1.2 Key Capabilities of SOAR Systems

The ability of a SOAR platform to gather and compile data from many security tools and systems is at its core. To safeguard their networks, modern enterprises use a variety of security solutions, including firewalls, vulnerability management platforms, intrusion detection systems (IDS), Security Information and Event Management (SIEM) systems, and endpoint detection and response (EDR) tools. Because of the enormous volumes of data that these systems provide, frequently in a variety of forms, security teams find it challenging to have a comprehensive grasp of the organization's overall security posture. In order to solve this problem, SOAR platforms centralize, normalize, and offer a single interface for analysis of data from all of these sources (Fig. 2).

Machine learning models and pre-established rules are used by SOAR platforms to evaluate and rank occurrences after data collection. SOAR systems can



Fig. 2 Capabilities of SOAR platform

discover possible dangers that might have escaped the notice of conventional detection methods thanks to the critical role machine learning plays in identifying patterns and anomalies within huge datasets. SOAR platforms offer actionable information by allowing speedier decision-making and lowering the possibility of false positives by connecting security events across multiple technologies. With automatic, standardized responses triggered by these insights, threats can be neutralized before they become more dangerous, potentially limiting the harm that an assault could inflict.

The capacity to integrate various security tools and facilitate the creation of unique playbooks is another crucial aspect of SOAR systems. Playbooks are preformatted procedures that specify how a company should react to particular kinds of problems. Using these workflows, security teams may respond to attacks in a standardized and consistent way by orchestrating operations across many security technologies. When a playbook identifies and detects malware infection, for instance, it may automatically start a sequence of steps that include removing the compromised system from the network, scanning it with an antivirus program, and raising the alarm with the security team. Even complex problems are handled well because of this degree of automation, which also speeds up response times.

1.3 Automation and Orchestration for Incident Response

By combining many security tools into a single system, SOAR systems' orchestration capabilities enable enterprises to automate whole incident response procedures. The ability of diverse security tools—each with unique skills and areas of expertise—to cooperate harmoniously is dependent on orchestration. These tools are coordinated via SOAR platforms, which allows them to interact and cooperate during an incident. For instance, without requiring human interaction, a SOAR system may autonomously start an intrusion detection system (IDS) to examine questionable network activity, use information from a security information and event management system (SIEM) to link the incident to earlier occurrences, and then activate a firewall to block the malicious IP address.

In addition, security teams can design unique playbooks based on their unique organizational requirements and risk profiles because of SOAR systems' versatility. From malware infections and phishing scams to insider threats and data breaches, playbooks may be created to handle a broad variety of security scenarios. The organization's distinct security rules and compliance needs are taken into account while designing these customized workflows, which guarantee incident response. SOAR solutions improve response times and the overall consistency and dependability of the security operation by automating the playbook execution process.

1.4 Real-Time Threat Analysis and Continuous Improvement

Another feature of SOAR systems that helps firms to identify and address issues as they happen is real-time threat analysis. Machine learning and orchestration frameworks are used by SOAR platforms to correlate and prioritize security incidents as they continuously monitor and analyze data from security tools. When a high-priority danger is recognized, SOAR systems can quickly carry out remediation activities because they automate the decision-making process. In order to minimize the impact of assaults like ransomware and advanced persistent threats (APTs), prompt action is crucial. This is where real-time response comes into play (Fig. 3).

Furthermore, by the integration of a feedback loop into the incident response procedure, SOAR systems promote ongoing enhancement. The results of security teams' playbook-executed responses to incidents are recorded and examined. The system's ability to respond to similar occurrences in the future is then enhanced by using this data to develop playbooks in the future. Furthermore, reporting and auditing capabilities are frequently included in SOAR platforms, which enable businesses to assess their security posture over time in order to decide what kind of cybersecurity investments to make going forward.



Fig. 3 Uses of SOAR platform

2 Related Work

In their work "Automatic Incident Response Solutions: A Review of Proposed Solutions' Input and Output," Karlzén and Sommestad use a methodical methodology to assess academic research on automatic incident response solutions. To group solutions according to their input and output, they apply the D3FEND Framework in conjunction with a Systematic Mapping Study approach that adheres to Petersen et al.'s recommendations. Once information from papers has been categorized using this framework, it is analyzed to find out how common various strategies are and how they compare to commercial products. This process is known as data extraction. Essential discoveries emphasize typical inputs such as network traffic analysis and asset inventories, with network isolation techniques frequently being the output. Academic and commercial solutions differ, though, and these disparities highlight opportunities for development and research shortages. Though more popular in research, practical insights advise playbook developers to concentrate on less typical strategies in practice. The limits of the review affect its validity and reliability. These include imprecise paper descriptions, possible biases, the omission of pertinent studies, and a restricted search scope. Despite its flaws, the study provides insightful information about automatic incident response methods that will direct future studies and improve cybersecurity incident management procedures [1].

The study article explores several approaches and strategies related to incident response, titled "A Comparative Study on Cyber Threat Intelligence: The Security Incident Response Perspective." The fundamental ideas of incident response are established, providing the framework for examining incident response formats. The structural components and advantages of six incident response forms are carefully examined. The strengths and differences of various formats are clarified through a comparative study, which facilitates their efficient use. The best format for a given use case can be found by categorizing objects according to fundamental ideas. Apart from that, the report assesses the wider product ecosystem that surrounds these formats and provides information on incident response systems and tools that are accessible. The results show that incident response formats consistently emphasize doable actions, and that playbooks are used to expedite response procedures. Responding skills are improved by the adaptability of combining several formats, led by fundamental ideas. But there are also drawbacks, such as restricted emphasis, difficulties in obtaining data, assessment-impairing proprietary features, and perpetual research difficulties [2].

A number of techniques to improve computer security incident response are presented in the paper "On Automation and Orchestration of an Initial Computer Security Incident Response by Introducing Centralized Incident Tracking System". The Incident Tracking System (ITS) is a crucial component that simplifies incident management by centralizing it for faster resolution. While Automation and Orchestration automate initial response and integrate user involvement as needed, the Host Isolating System provides techniques for isolating suspect hosts. The study

stresses that effective tracking requires monitoring incident ticket statuses and evaluating the dependability of SOC alerts. Results show that using ITS significantly improves workflow efficiency and reduces reaction times. Automation improves reaction efficiency and coordination, which results in better incident handling. Limitations, however, underscore the complexity of incident response despite automation efforts. These include problems with false positive detection and inconsistent alert reliability among SOCs. In order to effectively handle increasing security risks, the study emphasizes the necessity of continual process improvement, cooperation across security teams, and the continuous enhancement of response tactics [3].

In addition to introducing the Automated Incident Response (AIR) approach, the research article "Cyber threat intelligence enabled automated attack incident response" explores automated security incident response techniques, findings, and problems. By leveraging graph analytics and frameworks like D3FEND and ATT&CK, AIR integrates Cyber Threat Intelligence (CTI) with defensive measures to create the best possible security plans. By automating defensive measure selection, the methodology seeks to shorten the interval between breach discovery and reaction, hence increasing response efficiency. Test findings show that effective defenses may be predicted with high accuracy, reducing the advantages that attackers have. The Security Operation Centers (SOCs) can be fully automated by combining automated incident response and threat hunting, which strengthens the security posture overall. The Knowledge Graph (KG) has limited information, which presents challenges, especially with regard to defensive measures. Another issue is the reliance on reliable and high-quality data. The need for timely and thorough KG information highlights the necessity of exchanging threat intelligence and automatically generating attack graphs. Automated incident response has difficulties when defensive frameworks change, and the complexity of the best defense plans rises. The article presents AIR as a novel approach to automating incident responses, addressing the necessity of prompt and effective response in the face of constantly changing cyberthreats. It also emphasizes the importance of accurate, timely, and comprehensive threat information for the successful automation of security incident handling [4].

The research paper titled "PhiGARo: Automatic Phishing Detection and Incident Response Framework" explores automated security incident response techniques, discoveries, and difficulties, with a primary focus on phishing detection and management. With the use of honeypots, the framework records phishing attempts, uses the PhiGARo tool to automate incident handling, and keeps an eye on network flows to identify phishing situations. Furthermore, fictitious credentials are spread in order to obtain data for incident handling. Significant reductions in reaction times and manual labor were found, saving hundreds of hours of event handling time. After just one year of use, PhiGARo has shown to be cost-effective by enabling faster response times and preventing further harm caused by phishing attempts. The first spread of honeytokens, which necessitates human intervention, depending on user reports for event detection, and having restricted access to resources for installation and growth are some of the difficulties [5].

The study "On the Automation and Orchestration of Initial Computer Security Incident Response" explores the approaches, discoveries, and difficulties related

to the automation and coordination of security incident responses. It suggests an automated system that, upon obtaining the IP addresses of questionable hosts, can quickly identify and isolate them. This will allow for prompt reactions to notifications pertaining to potentially dangerous communications. To compute network topology and locate hosts automatically, the system makes use of a variety of automated components, including incident receiving, tracking, host locating, and isolation systems, in addition to protocols like LLDP and CDP. Results show a noteworthy decrease in the first response time, allowing enterprises to quickly isolate questionable hosts and possibly avert data breaches and virus outbreaks. There are still difficulties, though, such as the need to locate and isolate hosts manually, problems with unrecorded operations that impact accountability, and vendor lock-in restrictions that limit the flexibility of the equipment. Ultimately, the paper emphasizes how important automation and orchestration are to enhancing the efficacy and efficiency of secure computing initial incident responses. It also emphasizes how important it is to address related issues like vendor dependencies, accountability, and human dependence [6].

The goal of the research paper "SIEM-based framework for security controls automation" is to simplify information security management by introducing a comprehensive approach to security control automation. Examining suggested security controls from ISO/IEC 27001 and NIST SP 800-53 standards and assessing the automation potential of Security Information and Event Management (SIEM) technologies are the techniques used. Using SIEM systems for event correlation, risk analysis, and centralized control management are some of the techniques suggested. Other security controls that are automated include asset inventory, log management, vulnerability scanning, and incident management. Results show that ten essential security measures can be automated, accounting for almost thirty percent of the suggested controls. Platforms for SIEMs are acknowledged for their evolution into all-encompassing security centers that can automate threats, audits, and a variety of security measures. Among the difficulties mentioned are the intricacy of security management, problems with security tool compatibility, and restrictions placed on automation efficiency by social and human variables [7].

The study "Improving the Automation of Security Information Management: A Collaborative Approach" explores the field of automated security incident response and management, outlining difficulties, methodologies, and conclusions. The suggested architecture promotes the sharing of alerts and countermeasures among federated Security Information and Event Managers (SIEMs) across several domains, automating incident reporting procedures. The system uses Bash scripts with MySQL databases and Snort agents to efficiently detect and report possible attacks. Important results indicate that by sharing warnings and countermeasures among trustworthy domains, this cooperative strategy can improve decision-making among domain managers and increase overall security. But there are major obstacles to overcome, like information overload and integrating different technology. The study emphasizes how crucial it is to improve techniques for automatically creating and disseminating reaction plans in the event of reported assaults. Overall, the paper's

collaborative framework provides insightful information about how to use automation and teamwork to support security incident management across various domains. However, the successful implementation of this framework depends on resolving issues such as information overload and technology integration [8].

The study "Achieving automated intrusion response: a prototype implementation" explores approaches, strategies, conclusions, and difficulties related to automated security incident response and management. In order to adjust reaction choices according to incident circumstances, it suggests an architecture for an intelligent and adaptable automated response system. The architecture known as Flexible Automated Intelligent Responder, or FAIR, centralizes response management by initiating suitable actions in response to problems that are recognized. Results of the prototype implementation reveal how the system evaluates incident-related contextual parameters and illustrates how reaction choices change based on the circumstances of the occurrence. This demonstrates the versatility and flexibility of the system. Among the difficulties faced are the increasing amount and speed of network intrusions, which call for automated intrusion response. There are many obstacles to overcome, such as handling response in intricate networking infrastructures and growing incident reports. Beyond passive measures such as logging and alerting, there are additional challenges associated with automating response procedures, particularly active reactions [9].

The research article "A Comprehensive Semi-Automated Incident Handling Workflow" describes techniques, conclusions, and difficulties in an effort to automate security incident response and handling. To expedite incident management, it suggests a thorough approach that combines expert intervention with automated procedures. Triage procedures for sorting and ranking incidents, combining organizational, dynamic, and static knowledge for effective incident processing, and creating thorough incident handling protocols are important elements. The results highlight how efficiency can be increased gradually by reducing the need for human intervention while preserving adaptability to comply with corporate guidelines and business continuity plans. By covering several facets of the process, the workflow attempts to provide thorough coverage of incident handling principles. Automation attempts are, however, hampered by issues including the difficulty of managing incidents, the need for human expertise, and the high maintenance costs of expert teams. Notwithstanding these difficulties, the study emphasizes how crucial automation is to improving incident response capabilities—especially in large organizations—while also recognizing the necessity of striking a balance between automated and judgment from humans and experience [10].

3 Background

3.1 Overview of SOAR

In today's rapidly evolving cybersecurity environment, threats are becoming more common and complex at a rate never seen before. The development of more automated and effective solutions is required since manual incident response procedures are no longer sufficient to tackle these sophisticated intrusions. Platforms for Security Orchestration, Automation, and Response (SOAR) have become crucial parts of modern cybersecurity strategy because they automate and optimize the whole incident response lifecycle.

SOAR platforms are designed to integrate a wide range of security tools, including Intrusion Detection Systems (IDS), Endpoint Detection and Response (EDR), Security Information and Event Management (SIEM), and threat intelligence sources. By combining many technologies into one system, SOAR systems enable organizations to automate threat detection, analysis, and response processes with minimal human involvement. Security teams benefit from improved accuracy, faster reaction times, and lower operational overhead as a result. The fundamental components of a SOAR platform can be grouped into the following categories:

- Orchestration: By coordinating the interactions of many security tools, SOAR systems make it easier for them to communicate and work together. A SOAR platform may automatically extract indicators of compromise (IoCs) from an EDR system, improve the data using threat intelligence feeds, and forward relevant information to a SIEM for correlation analysis. This eliminates the need for manual processes and ensures smooth security operations across multiple systems.
- Automation: One of the key benefits of SOAR platforms is their capacity to automate mundane and repetitive processes that have traditionally burdened security professionals. The platform can reduce the amount of effort required to find and fix problems by automatically managing operations like threat hunting, data enrichment, and log analysis. Furthermore, automation ensures consistency in responses, which lowers the chance of human error, especially when managing large volumes of data.
- Reaction: Pre-written, editable playbooks that outline the exact steps needed to handle different incident types, including malware outbreaks, data breaches, and insider threats, are provided by SOAR solutions. The severity of the problem will determine whether these playbooks are executed automatically or require human approval. By allowing security teams to maintain control while automating less crucial processes, this flexibility guarantees the optimal possible balance between automation and human monitoring (Fig. 4).

In addition to these core components, contemporary SOAR platforms often include ***machine learning algorithms and threat intelligence*** to enhance decision-making. By using real-time data analysis and prediction models, SOAR systems may be able to identify potential threats more rapidly and accurately than manual

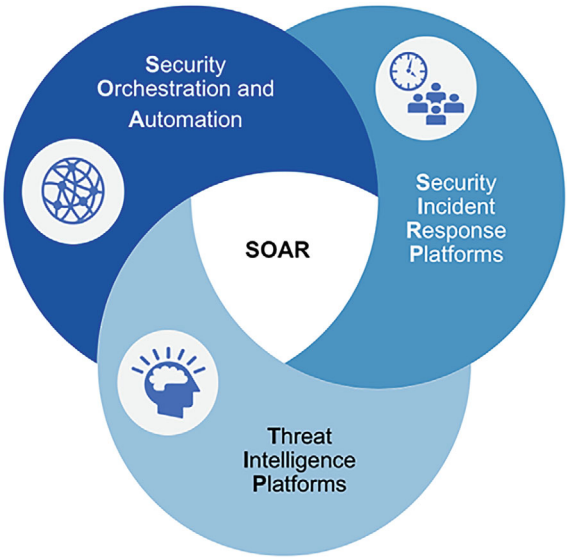


Fig. 4 Broad components of SOAR concept

methods. This reduces the possibility of overlooking crucial warnings and helps rank problems based on their seriousness.

The significance of SOAR platforms to the industry cannot be overstated. For businesses coping with an increasing number of cyberattacks, incident response systems that are automated, scalable, and agile are increasingly essential. By using SOAR systems, which provide a comprehensive approach to security operations, organizations may enhance their cybersecurity posture and stay ahead of emerging threats.

3.2 Previous Approaches to SOAR Architectures and Frameworks

Prior to the development of fully integrated SOAR platforms, cybersecurity teams relied heavily on specialist technology like SIEM systems, which collected and analyzed data from several sources. SIEM technology excelled at data collection and centralization, but they usually lacked automatic incident response capabilities. Analysts were required to manually correlate data, examine the impact of errors, and adopt corrective actions—a time-consuming and error-prone process.

Recognizing these limitations, researchers and industry professionals began to look into more holistic methods to security automation, which resulted in the development of frameworks that integrated orchestration, automation, and response seamlessly. Early research in this topic focused on how automation might be used to reduce incident response times while increasing accuracy. These efforts also highlighted the

importance of employing machine learning for anomaly detection, behavioral analysis, and threat prioritization, allowing SOAR platforms to move beyond rule-based automation.

Numerous academic studies have investigated how orchestration and automation can assist control complicated, multi-stage attacks. Some studies have pushed for designs that leverage automated decision-making to manage incidents from detection to remediation, while others have focused on open-source SOAR frameworks that allow businesses to customize their automation strategies. Despite these advances, many early frameworks had limited integration capabilities and necessitated substantial manual configuration.

Modern SOAR platforms have addressed these shortcomings by offering out-of-the-box integrations with a wide range of security solutions, as well as more advanced features like real-time collaboration and playbook customization. These solutions enable security teams to create dynamic and adaptive procedures while also automating responses to a wide range of attack scenarios.

3.3 Market Overview of Existing SOAR Platforms

The market for SOAR platforms has expanded considerably in recent years, with several critical solutions gaining mainstream acceptance in a range of industries. These platforms range in terms of features, capabilities, and integration options, but they all share the goal of boosting incident response operations. Here are some of the most popular SOAR platforms available today:

- Splunk Phantom, a well-known SOAR solution, focuses on automation and integration with a diverse variety of third-party security systems. It has a large library of pre-built playbooks and workflows, allowing businesses to automate complex security tasks without having to build custom scripts from scratch. Its flexibility and integration features make it a good choice for businesses seeking a completely tailored SOAR solution. However, Splunk Phantom has a steep learning curve, and its high cost may be prohibitive for smaller businesses.
- IBM Resilient, which is well-known for its intuitive incident response and playbook management features, provides security teams with a powerful platform for orchestrating and automating responses to a wide range of incidents. The platform focuses on real-time collaboration, allowing multiple analysts to work on the same issue. Furthermore, AI-powered automation enhances decision-making by making recommendations based on historical data and threat intelligence. Customization can be time-consuming, and maintaining integrations can be challenging, especially in several tool configurations.
- Palo Alto Networks developed Cortex XSOAR, a platform that unifies incident management, automation, and threat intelligence. One of its most remarkable features is its ability to readily integrate with a wide range of security technologies, making it an excellent solution for companies that rely on a diverse security

ecosystem. Its chat-based collaboration solution also stands out, allowing security personnel to connect and coordinate more effectively during incident response operations. Cortex XSOAR may experience performance lag with large processes, and there is a risk of vendor lock-in to the Palo Alto environment.

4 High-Level Architecture of SOAR Platforms

Usually, SOAR platforms include multiple separate levels in their architecture, each of which is essential to the operation of the system. The User Interface (UI), Application Logic, Automation, Data, Integration, Security, and Monitoring and Logging are some of these levels. Organizations are able to effectively automate security operations, coordinate incident response, and connect with a variety of security solutions when they have a well-designed SOAR architecture that guarantees smooth communication across various layers (Fig. 5).

4.1 Common Architecture Layers

User Interface (UI) Layer is the SOAR platform's front-end, where security teams interact with the system. This layer is in charge of visualizing security operations, creating dashboards for incident monitoring, providing drag-and-drop interfaces for playbook generation, and supporting team communication. The UI layer prioritizes user experience and accessibility, ensuring that security analysts can respond to threats with minimal effort.

Logic Application Layer is the brain of SOAR platform. It controls how playbooks, workflows, and other automated operations are carried out. The workflow engine, which controls the logic for threat analysis, incident triaging, and remediation procedures, is a component of this layer and is responsible for orchestrating various

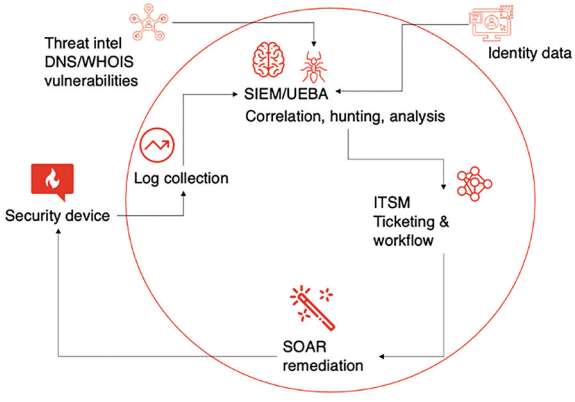


Fig. 5 Adoption of SOAR in incident response

security technologies. In order to improve decision-making based on real-time threat intelligence, it also interfaces with Machine Learning (ML) modules.

Automation Layer is the heart of SOAR platforms' immense capability. The predefined functions, like data enrichment, threat hunting, and log analysis, are carried out by this layer automatically without the need for human participation. Automated processes minimize human labor and provide uniformity in event management, enabling prompt reactions to risks. This layer's intricate orchestration of security operations across various systems is made possible by tools such as Ansible, Puppet, and Chef.

Data Layer manages playbooks, incident logs, user data and many more essential data. In order to facilitate the tracking of prior reactions and enhance processes in the future, SOAR platforms retain incident history and threat intelligence data. Scalability and performance are guaranteed while working with massive numbers of security data by this layer, which frequently makes use of databases like MySQL, PostgreSQL, or NoSQL alternatives like MongoDB.

Integration Layer within SOAR platform helps it to connect with Threat intelligence platforms, firewalls, SIEM systems, and endpoint detection and response (EDR) systems. It is ensured by this layer that the SOAR system can receive response actions from various sources and ingest data from them. In order to facilitate the easy integration of an organization's current security infrastructure with the SOAR platform, integration is typically accomplished through RESTful APIs, Webhooks, and precompiled connectors.

Security Layer enforces important security elements like encryption, authorization, and authentication. This layer uses role-based access control (RBAC) to make sure that only authorized users can access certain areas of the SOAR platform. Technologies like OAuth2 and JWT, along with encryption protocols like SSL/TLS, facilitate safe communication both within the SOAR system's various components and between the platform and external tools.

Monitoring and Logging Layer is crucial for monitoring the SOAR platform's overall performance and for reviewing security incidents and reactions. Centralized logging and real-time monitoring are frequently accomplished with tools like Prometheus, Grafana, and the ELK Stack (Elasticsearch, Logstash, Kibana). Logs are kept for later review, serving as an audit trail for issue responses and assisting in locating areas where the SOAR procedures could use some enhancement.

The overall objective of automating and streamlining security operations is furthered by the distinct roles that each tier in a SOAR platform plays. Security teams can readily interface with the system thanks to the UI Layer's focus on user experience. To make automatic judgments, the Application Logic Layer combines intelligence and carries out intricate procedures. By processing common operations quickly, the Automation Layer reduces the need for manual involvement. Critical information is structured and stored in the Data Layer, and the Integration Layer serves as a link among SOAR and other security systems. Finally, the Monitoring and Logging Layer offers transparency into platform performance and incident response actions. The Security Layer guarantees the integrity of the system and the data it processes.

4.2 Existing Frameworks and Architectures

Current SOAR architectures are influenced by a number of frameworks. Although they differ in terms of performance, scalability, and flexibility, platforms like Cortex XSOAR, IBM Resilient, and Splunk Phantom implement topologies comparable to the one described above. For example, Splunk Phantom excels in its ability to integrate with many third-party products, but it can be challenging to administer. IBM Resilient places a strong emphasis on real-time collaboration and incorporates AI-powered recommendations, although it could be difficult to adapt procedures to different situations. Cortex XSOAR may have performance issues during large-scale operations, but its focus is on smooth tool interaction and strong automation features.

In contrast, our suggested SOAR design aims to enhance scalability, usability, and performance while adhering to the same fundamental principles. One way to reduce reaction times is to concentrate on granular real-time monitoring, while another is to leverage lightweight containerization technologies to optimize workflow orchestration. Furthermore, our methodology places a strong emphasis on the integration of machine learning for proactive threat identification and the creation of adaptive playbooks that can change to respond to new threats. By resolving the shortcomings seen in previous solutions, this improves the dynamic and future-ready nature of our design.

5 Key Technologies Used in SOAR Platforms

SOAR platforms leverage a variety of technologies to integrate, automate, and coordinate security operations. These technologies include programming languages, databases, APIs, and automation tools, all of which are critical in developing a flexible and scalable SOAR solution. This section delves into the key technologies used in SOAR development, as well as the growing role of machine learning in security automation.

5.1 Programming Languages

SOAR platforms rely on programming languages that are adaptable, scalable, and capable of handling complex integrations. These languages allow developers to build robust systems capable of automating security processes while maintaining interoperable with a wide range of third-party solutions.

Python is widely used in SOAR systems due to its simplicity, readability, and extensive library support. Python's extensive ecosystem of modules, including requests for handling HTTP requests and pandas for data analysis, makes it ideal for constructing playbooks and automating repetitive security tasks. Its adaptability

allows security teams to swiftly develop custom scripts to automate incident response activities. Python is also well suited for interfacing with other security technologies via APIs.

In SOAR platforms, JavaScript is often used for front-end development, enabling for the creation of user-friendly dashboards and interfaces. Frameworks such as React and Vue.js are used to construct responsive, real-time interfaces for monitoring security events and executing playbooks. Because of its asynchronous nature, JavaScript is ideal for real-time data streaming and changes, which are necessary for monitoring and logging applications.

Known for its speed and scalability, go (or Golang) is commonly used at the backend of SOAR platforms where concurrency and performance are critical. Go's ability to handle multiple jobs at once makes it a good choice for constructing scalable SOAR systems capable of analyzing enormous amounts of security data.

5.2 Databases

SOAR platforms require efficient data storage and retrieval since they handle huge amounts of security-related data like as logs, alerts, and incident reports. Different database types are utilized depending on the nature of the data and the platform's performance requirements.

SOAR platforms commonly use relational databases, like as MySQL and PostgreSQL, to store structured data such user profiles, incident information, and playbook metadata. These databases support accurate and consistent querying, which is essential for report creation, user role maintenance, and incident history recording.

MongoDB is a popular choice for managing unstructured or semi-structured data, including logs, threat information, and real-time alerts. These databases offer flexibility in storing data types that do not naturally fit into relational structures. They are useful for storing large amounts of event data that need to be accessed fast for analysis and reporting.

5.3 APIs and Integration Technologies

SOAR platforms rely heavily on APIs (Application Programming Interfaces) for orchestration and integration. SOAR systems must integrate with a wide range of third-party security tools, including SIEMs, firewalls, and endpoint protection solutions. This is made possible by the use of APIs, which enable easy communication between different systems.

REST (Representational State Transfer) APIs are widely used in SOAR platforms due to its simplicity and compatibility with a variety of systems. RESTful APIs enable SOAR platforms to interface with security solutions using HTTP queries to retrieve

or send data. For example, a SOAR system can use REST APIs to query a SIEM for security alerts or to communicate remedial actions to a firewall.

Another common integration option, webhooks allow SOAR systems to receive real-time data from security solutions whenever an incident occurs. This is especially useful for triggering rapid responses to high-priority events. Webhooks minimize latency and enable real-time automation of incident response procedures.

SOAR platforms manage service communication and asynchronous data processing using technologies such as RabbitMQ and Apache Kafka. These message queuing systems ensure that events and alerts from several sources are processed consistently and in an ordered manner, allowing the SOAR platform to scale efficiently.

5.4 Automation Tools

Automation is key to SOAR platforms, allowing security professionals to automate repetitive tasks while optimizing workflows. Several automation tools are used to power these procedures.

Ansible is a popular open-source automation tool for SOAR platforms that automates tasks such as system configuration, application deployment, and incident response. Ansible playbooks define a set of automation operations that can be executed across several systems, enabling security teams to simply construct repeatable procedures.

SOAR platforms use Puppet and Chef for configuration management and automation. They allow security teams to automate the deployment and management of infrastructure, ensuring that systems follow security policies and guidelines.

Containerization technologies like Docker and orchestration platforms like Kubernetes are being used more frequently in SOAR systems to boost scalability and resilience. These tools enable SOAR platforms to incorporate microservices-based architectures, giving them greater flexibility in developing different system components independently.

5.5 Machine Learning in Security Automation

Machine learning (ML) is becoming an increasingly significant component of SOAR systems, enhancing its ability to detect, evaluate, and respond to cyberattacks. More complicated tasks, such as pattern recognition and predictive analysis, can be automated by SOAR systems using machine learning techniques.

Machine learning models identify patterns in past security data, allowing the SOAR platform to detect anomalies or potential threats that traditional rule-based systems may overlook. For example, machine learning models can detect unusual network behavior that may indicate a breach or insider threat.

Machine learning algorithms can assist in prioritizing occurrences based on historical data, threat intelligence, and the severity of the reported anomaly. This allows security personnel to focus on the most pressing concerns, reducing response times and preventing more serious incidents.

SOAR platforms can use machine learning to create adaptive playbooks that adjust their responses to shifting threat landscapes. By examining and learning from previous occurrences, machine learning models can improve workflows, making incident response more dynamic and effective.

Some SOAR platforms employ NLP technology to scan and analyze human-readable data, such as emails or chat logs, in order to detect phishing or other forms of social engineering. NLP enables security teams to automate actions that require the analysis of large amounts of textual data.

SOAR platforms are designed and operated with a mix of established and emerging technology. Python and JavaScript are programming languages used to create customizable automation workflows, whereas databases hold huge amounts of security data. APIs and message queuing systems ensure seamless integration with third-party products, and automation frameworks like Ansible and Docker offer incident response workflow scalability. Machine learning enhances these capabilities, making SOAR platforms more adaptive, intelligent, and responsive to the always changing threat environment.

6 Current Challenges in SOAR Platforms

The automation and orchestration of security operations have greatly improved thanks to Security Orchestration, Automation, and Response (SOAR) platforms; however, there are still a number of obstacles that businesses must overcome while implementing and expanding these solutions. These difficulties span a wide range of domains, from functional concerns like user cooperation and machine learning module efficacy to technological constraints like scalability and integration.

Scalability is among the most critical issues facing SOAR platforms. The number of data and incidents rises quickly as firms expand and use more security systems. Big data processing in real time, response orchestration across intricate infrastructures, and workflow growth without performance degradation are all tasks assigned to SOAR systems. With thousands of problems handled daily across internationally distributed locations, many SOAR solutions currently in use have difficulty maintaining performance under extreme loads. In order to effectively manage this size, certain SOAR platforms currently lack sophisticated architecture and optimization strategies.

Although SOAR solutions are made to automate incident response in real time, processing in real time can be particularly difficult. To effectively reduce threats, massive volumes of data from several sources must be collected, processed, and acted upon in a matter of seconds. When the system has to correlate and prioritize occurrences based on threat information, historical data, and real-time inputs, the

analysis step is usually when a bottleneck appears. Delays in responding may result from this, lengthening the period of vulnerability. Multiple events occurring simultaneously can also cause performance concerns because it slows down detection and reaction times.

SOAR platforms mostly depend on interfacing with other security tools including firewalls, endpoint detection and response (EDR) programs, SIEM (Security Information and Event Management) systems, and more. Integration is a major difficulty because these products are often dissimilar in terms of architecture, APIs, or data formats. While vendor technology and standards vary, inefficiencies, missing data, or inconsistent replies are frequently the result of this lack of seamless communication between instruments, which is necessary to guarantee efficient orchestration. Moreover, SOAR systems must continuously modify their integrations due to upgrades from outside vendors, which increases the complexity and expense of running a SOAR system.

The majority of incident response tasks can still be automated with SOAR platforms, but human supervision and cooperation are still crucial, especially in complicated situations when expert judgment is needed. Most SOAR platforms, however, don't provide enough assistance for user collaboration. Many analysts may find it challenging to work together on an event at the same time or to offer input on automated choices due to the interfaces' possibly nonintuitive design. Moreover, poorly designed communication capabilities inside SOAR platforms can make it difficult for teams to handle incidents in a cohesive manner, which would result in fragmented workflows and longer periods for resolving issues.

Many contemporary SOAR platforms depend heavily on machine learning (ML), which facilitates the analysis of huge datasets, the discovery of patterns, and the automation of decision-making procedures. Nevertheless, the accuracy and adaptability of ML modules in SOAR systems are frequently constrained. The imbalance in datasets (more data on benign occurrences than actual incidents) makes it difficult to train machine learning models on security data. Furthermore, when threats change, machine learning models need to be often retrained in order to remain successful. Because of this, SOAR platforms run the risk of producing false positives, which wears out alerts, or missing new threats, which lowers their overall efficacy in thwarting complex attacks. Moreover, a lot of machine learning algorithms utilized in SOAR platforms still rely on supervised learning, which defeats the purpose of complete automation since it requires human involvement for data labeling and refinement.

7 Future Directions and Emerging Trends

The Security Orchestration, Automation, and Response (SOAR) platform market is changing quickly in reaction to the increasing complexity and frequency of cyberattacks. A number of significant developments and new technologies are influencing how SOAR platforms will develop in the future as businesses seek to strengthen

their security posture. Security Operations Centers (SOCs) will benefit from these advancements in automation, machine learning (ML) and artificial intelligence (AI), real-time collaboration, and user experience optimization.

7.1 Trends in Automation and Orchestration Tools

The increasing complexity of automation and orchestration capabilities is one of the most important themes in SOAR platforms. The need for more sophisticated and adaptable automation technologies has grown as cybersecurity teams deal with an increasing number of occurrences. It is probable that SOAR platforms in the future will have improved automation frameworks that are capable of adapting to ever-changing settings and dynamic threats without the need for continual human intervention. This will be made possible by sophisticated workflow engines that provide more adaptability for managing intricate, multi-phase events.

Furthermore, "low-code" or "no-code" orchestration tools—which let security experts create unique workflows and playbooks without having advanced coding knowledge—should be integrated. This will facilitate quicker deployment and more effective incident response by making it simpler for enterprises to customize SOAR systems to meet their own operating requirements.

7.2 Role of AI/ML in Security Operations

Due to the requirement for improved decision-making, quicker reaction times, and more accurate threat identification, artificial intelligence (AI) and machine learning (ML) in SOAR systems are expected to play an increasingly larger role. Future AI/ML systems will be more self-sufficient, able to learn from past mistakes and adjust to new kinds of cyberthreats without human intervention. This will assist in shortening the time needed to identify and counter sophisticated assaults, particularly in settings where human capacity to handle the volume and complexity of data is exceeded.

The application of "unsupervised learning" algorithms is one approach that shows promise. By examining trends in network traffic, user activity, and other security data, these algorithms can assist uncover dangers that were either undiscovered or recently discovered without the need for labeled datasets. These developments in AI and ML will greatly improve SOAR platforms' predictive capacities, enabling them to proactively detect and reduce risks before they materialize into big crises.

In addition to the advantages of AI/ML, there's an increasing need for addressing the biases and limits present in these models. For automated security operations to continue to be trusted and accountable, future SOAR platforms must guarantee that judgments made by AI are clear and understandable to human operators.

7.3 The Impact of Real-Time Collaboration in Incident Response

When responding to incidents, cooperation across security teams is essential, especially when there are a lot of moving parts and stakes. Real-time collaboration tools that let analysts collaborate easily across time zones and locations will be emphasized more and more in upcoming SOAR systems. The creation of more sophisticated coordination and communication tools, such shared dashboards, incident discussion spaces, and integrated chat systems, will be necessary to enable different stakeholders to participate in the investigation and remediation process.

A newer approach is "collaborative automation," in which automated systems and human analysts collaborate to resolve issues. With this method, human specialists' critical thinking and context awareness are combined with the speed and efficiency of automation. With the ability for humans to modify or override automated activities as needed, such a model will help lower the risk of mistakes or lost opportunities during incident response.

7.4 Opportunities for Innovation in UI/UX for Security Operations Centers (SOCs)

User interfaces (UI) and user experiences (UX) in Security Operations Centers (SOCs) need to be enhanced as SOAR platforms become more essential to security operations. Conventional SOAR platforms can have confusing and sophisticated user interfaces that might overload analysts with information or make it difficult to locate important data. This may cause security personnel to have to work more and respond more slowly.

Future SOAR design efforts will probably concentrate on developing more intuitive, user-friendly interfaces that make it simple for analysts to handle workflows, retrieve important data, and move through occurrences. The application of visual playbooks, drag-and-drop workflow automation, and AI-driven dashboards that rank the most important occurrences and suggestions are a few examples of UI/UX innovations.

SOCs may also implement immersive, 3D interfaces that offer security analysts a more dynamic and interesting approach to handling events, given the growing popularity of augmented reality (AR) and virtual reality (VR) technologies. Long-term, it will be simpler to retain qualified cybersecurity personnel thanks to these UI/UX innovations, which will also increase productivity and lessen analyst burnout.

8 Conclusion

Organizations' approach to managing the increasing complexity and frequency of cyber-attacks has evolved significantly, as evidenced by the current state of SOAR (Security Orchestration, Automation, and Response) systems. SOAR platforms—which combine several security products, automate tedious activities, and optimize incident response workflows—have become essential solutions as manual issue response grows more and more inefficient. Although popular platforms such as Splunk Phantom, IBM Resilient, and Cortex XSOAR provide strong automation, orchestration, and threat intelligence integration, they also have drawbacks such as scalability problems, integration difficulties, and infrequent lack of flexibility in real-time collaboration.

There are still a number of shortcomings and difficulties with SOAR technologies. These include the requirement for enhanced real-time processing capabilities, more seamless integration with various security technologies, better scalability to handle increasing data quantities, and improvements in user feedback and collaboration processes. Furthermore, the adaptability and transparency of the AI/ML modules currently employed in SOAR systems are limited, which may affect the decision-making process during incident response.

Sustaining research and innovation in SOAR architectures is crucial, considering the vital role that SOAR plays in contemporary cybersecurity operations. To tackle these obstacles, more intelligent, flexible, and scalable platforms must be created in order to enhance user cooperation and real-time response. These platforms should also be better able to interface with a variety of security technologies. To further propel the next generation of SOAR systems, improving AI/ML capabilities and emphasizing user-friendly interfaces will be crucial. Future research and development in SOAR architectures will be essential for equipping enterprises to defend against increasingly complex assaults in this quickly changing threat scenario.

References

1. Karlzen, H., Sommestad, T.: Automatic incident response solutions: a review of proposed solutions' input and output. In: 18th International Conference on Availability, Reliability and Security (ARES'23)—ACM Digital Library (2023). https://doi.org/10.1145/3600160.3605066
2. Schlette, D., Caselli, M., Pernul, G.: A comparative study on cyber threat intelligence: the security incident response perspective. In: IEEE Communications Surveys & Tutorials, vol. 23, no. 4, pp. 2525–2556 (2021). https://doi.org/10.1109/COMST.2021.3117338
3. Ohmori, M.: On automation and orchestration of an initial computer security incident response by introducing centralized incident tracking system. J. Inf. Process. (2019). https://doi.org/10.2197/ipsjjip.27.564
4. Kaiser,F.K., Andris, L.J., Tennig, T.F., Iser, J.M., Wiens, M., Schultmann,F.: Cyber threat intelligence enabled automated attack incident response. In: 3rd International Conference on Next Generation Computing Applications (2022). https://doi.org/10.1109/NextComp55567.2022.9932254

5. Husák, M., Cegan, J.: PhiGARo: automatic phishing detection and incident response framework. In: Ninth International Conference on Availability, Reliability and Security, Fribourg, Switzerland, pp. 295–302 (2014). https://doi.org/10.1109/ARES.2014.46
6. Ohmori, M., Higashino, M., Kawato, T., Fujio, S., Kawamura, T.: On automation and orchestration of an initial computer security incident response. Information Processing Society of Japan SIC Technical Report, vol. 2018-IOT-40, pp. 155–160 (2018)
7. Montesino, R., Fenz, S., Baluja, W.: SIEM-based framework for security controls automation. Inf. Manag. Comput. Secur. **20**(4), 248–263 (2012). https://doi.org/10.1108/09685221211267639
8. Aguirre, I., Alonso, S.: Improving the automation of security information management: a collaborative approach. IEEE Secur. Priv. **10**(1), 55–59 (2012). https://doi.org/10.1109/MSP.2011.153
9. Papadaki, M., Furnell, S.M.: Achieving automated intrusion response: a prototype implementation. Inf. Manag. Comput. Secur. **14**(3), 235–251 (2006). https://doi.org/10.1108/09685220610670396
10. Hashemi, S.H., Babaeizadeh, M., Nowruzi, M., Jazi, H.H., Shahmoradi, M., Beigi Samani, E.B.: A comprehensive semi-automated incident handling workflow. In: 6th International Symposium on Telecommunications (IST), Tehran, Iran, pp. 1065–1070 (2012). https://doi.org/10.1109/ISTEL.2012.6483144

Trust in Social Media Through Deepfake Detection: A Literature Review



Debasmita Guha and Sumedha Sirsikar

Abstract This paper provides a comprehensive review of the growing cybersecurity challenges posed by deepfake technology and advanced language models such as ChatGPT. With the advancement of machine learning techniques like Generative Adversarial Networks (GANs), deepfake technology can produce increasingly realistic synthetic media, which malicious actors exploit for purposes ranging from misinformation to identity theft. The paper examines these technologies' dual-use potential, noting their roles in both enhancing and undermining cybersecurity measures. Current detection strategies, categorised into image, video, and text-based methods, are analysed alongside potential integrations with social media platforms for more robust defence mechanisms. Emerging technologies, such as blockchain, are also discussed for their potential to strengthen digital content authentication and protect against cyber deception. By exploring application scenarios, including social media engineering and the Internet of Things (IoT), the paper highlights key insights, challenges, and open questions. It concludes by recommending proactive measures and future research directions, emphasising the urgency of developing adaptive, multimodal detection systems to safeguard against deepfake-related threats in the evolving cybersecurity landscape.

Keywords Deepfake detection · Generative adversarial networks (GANs) · Deep learning · Digital content integrity · Explainable AI (XAI)

D. Guha (✉) · S. Sirsikar
Department of Computer Engineering and Technology, School of Computer Science and Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India

S. Sirsikar
e-mail: sumedha.sirsikar@mitwpu.edu.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_39

1 Introduction

1.1 Background

Deepfake technology, a product of rapid advancements in artificial intelligence (AI) and deep learning, has revolutionized the creation of synthetic media. By leveraging powerful generative models such as Generative Adversarial Networks (GANs), deepfakes produce hyper-realistic images, videos, and audio that can seamlessly mimic real human behavior and appearances. While this innovation opens avenues in entertainment, accessibility, and creativity, it also poses significant threats. Deepfakes have been weaponized for malicious purposes, including misinformation campaigns, identity theft, and fraud, undermining trust in digital media and public discourse. Consequently, the detection and mitigation of deepfakes have become a critical focus within cybersecurity, law enforcement, media, and AI research communities.

1.2 Core Mechanism: Generative Adversarial Networks (GANs)

Developed by Ian Goodfellow in 2014, GANs have revolutionized the field of synthetic media generation by leveraging an adversarial process that trains two networks against each other. In this setup:

- The generator creates synthetic images, video frames, or audio samples, starting with random data and gradually improving with training.
- The discriminator evaluates these synthetic samples by comparing them against real data to discern authenticity (Fig. 1).
- The adversarial relationship between these two networks allows the generator to improve with each iteration, producing more realistic results. This iterative

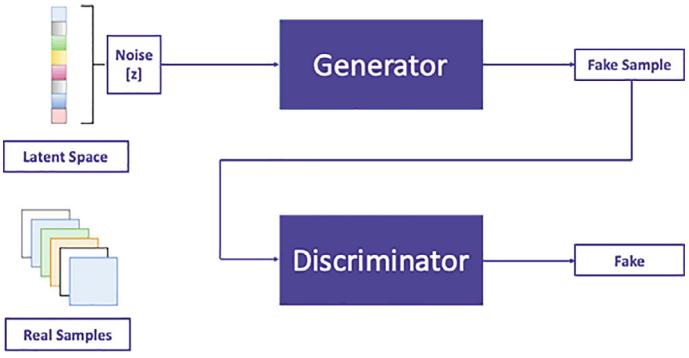


Fig. 1 GAN architecture

training process continues until the discriminator can no longer reliably distinguish between real and synthetic data. GANs have made it possible to create lifelike media by capturing subtle details, from facial expressions and voice intonations to background settings.

2 Related Work

The literature on deepfake detection has evolved significantly in response to the rapid advancement of deep learning technologies and the increasing prevalence of manipulated digital content. Deepfakes, which utilize Generative Adversarial Networks (GANs) to create realistic fake images and videos, pose serious threats to privacy, security, and the integrity of information. Researchers have explored various methodologies to combat this issue, focusing on the application of deep learning techniques for detection. Early detection methods primarily relied on traditional machine learning approaches, which often struggled with the complexity and variability of deepfake content. However, the introduction of deep learning frameworks has revolutionized the field, enabling more sophisticated analysis of visual features. Techniques such as Convolutional Neural Networks (CNNs) and deep belief networks have shown promise in identifying subtle artifacts and inconsistencies in manipulated media. Recent studies, including the one presented in this paper, emphasize the importance of feature extraction methods, such as fuzzy clustering, to enhance the detection process. By preprocessing images to remove noise and improve quality, researchers have achieved higher accuracy rates. The proposed deep belief network with loss handling through pairwise learning has demonstrated significant improvements, achieving detection accuracies of up to 98% across various datasets, including Face2Face and FaceSwap. Moreover, collaborative efforts, such as the Deepfake Detection Challenge initiated by organizations like Facebook and Microsoft, highlight the urgency of developing robust detection systems. These initiatives aim to foster innovation and improve the accuracy of detection methods, addressing the challenges posed by the evolving nature of deepfake technology [1].

The literature on deepfake detection has evolved significantly, focusing on various methodologies and technologies to combat the challenges posed by synthetic media. Recent studies emphasize the use of deep learning (DL) algorithms, particularly Convolutional Neural Networks (CNNs), which have shown promising results in classifying real and deepfake images. For instance, one study achieved an accuracy of 91.56% using transfer learning with CNN architectures, while another reported an even higher accuracy of 97.83% with recurrent neural networks. A critical aspect of this research is the integration of Explainable AI (XAI) techniques, such as Local Interpretable Model-Agnostic Explanations (LIME). LIME helps in understanding the decision-making process of models by identifying which features of an image contribute to specific classifications. This transparency is vital for validating the reliability of deepfake detection models and ensuring that researchers can interpret the results effectively. Moreover, the dataset used for training and testing these models

often comes from open-source platforms like Kaggle, which provides a diverse range of images for robust model training. The combination of advanced DL techniques and XAI not only enhances detection accuracy but also paves the way for future research by providing insights into model performance and feature importance [2].

The literature on Deepfakes encompasses a wide range of studies focusing on both the generation and detection of synthetic media. A significant body of work categorizes Deepfake generation techniques into various methods, including face swapping, face reenactment, and attribute manipulation. Yu et al. [3] provide a thorough review of video detection methods, classifying algorithms based on facial image manipulation goals and highlighting the limitations of current detection methods in real-world applications. Almars [4] further distinguishes between image and video content for detection, proposing categories based on biological signals and spatio-temporal features, while emphasizing the scalability issues faced by existing Deep Learning models. Juefei-Xu et al. [5] offer a comprehensive analysis of both generation and detection methods, discussing the ongoing battle between these two domains and introducing concepts such as adversarial attacks that aim to evade detection. Mirsky and Lee [6] focus on identifying the shortcomings of current defense solutions and suggest areas for further research, particularly in prevention and mitigation strategies. The literature also highlights the potential of new technologies, such as blockchain, to enhance cybersecurity and combat digital deception. The need for accurate, diverse, and large datasets is emphasized across studies, as well as the importance of exploring detection methods across various social media platforms beyond Twitter [7].

The literature on cybersecurity threats posed by advanced technologies, particularly deepfake algorithms and Massive Language Models (MLMs) like ChatGPT, has gained significant attention in recent years. Deepfake technology, primarily based on Generative Adversarial Networks (GANs), has been widely studied for its ability to create hyper-realistic fake images and videos, raising concerns about misinformation and identity fraud. Research indicates that deepfakes can be used in various malicious contexts, including social engineering attacks and the spread of disinformation, which can severely undermine trust in digital communications. Conversely, MLMs such as ChatGPT have been recognized for their potential to automate tasks and enhance user interactions. However, their capabilities also present risks, as they can generate misleading or harmful content that may be exploited by cybercriminals. Studies have highlighted the dual nature of these technologies, where their benefits in improving efficiency and accessibility are counterbalanced by the potential for misuse in cybercrime. The literature emphasizes the need for comprehensive cybersecurity strategies to address these emerging threats. Recommendations include user education to recognize deepfake content, the implementation of advanced security measures, and continuous monitoring of AI technologies to mitigate risks [8].

Deepfake technology, which utilizes advanced machine learning techniques such as Generative Adversarial Networks (GANs) and autoencoders, has gained significant attention due to its ability to create realistic synthetic media. Recent studies highlight the effectiveness of various GAN architectures in generating deepfakes across multiple modalities, including images, videos, and audio. For instance, Patel

et al. provide a comprehensive overview of multi-modal deepfake generation models, emphasizing the foundational principles of these technologies. In parallel, the detection of deepfakes has emerged as a critical area of research. Various detection methods have been proposed, focusing on different modalities. While some studies, like those by Tolosana et al., primarily address visual content, others, such as Nguyen et al., offer a broader perspective by including audio detection techniques. The literature indicates that detection strategies often rely on machine learning algorithms and forensic analysis, but challenges remain due to the rapid evolution of deepfake generation techniques, which can outpace detection capabilities. Moreover, ethical considerations surrounding deepfake technology are increasingly prominent, as the potential for misuse in misinformation and privacy violations raises significant societal concerns. Future research directions emphasize the need for robust detection algorithms that can adapt to new-generation techniques and the importance of integrating multi-modal approaches to enhance detection effectiveness [9].

Deepfake technology, which utilizes advanced artificial intelligence (AI) and machine learning (ML) to create realistic fake media, poses significant challenges across various sectors, including journalism, security, and social media. The detection of deepfakes has become an active area of research, with various methods and technologies evolving to address the growing threat. Early detection techniques relied on traditional machine learning algorithms that extracted handcrafted features from videos, such as pixel differences and facial landmarks, laying the foundation for more sophisticated approaches. The introduction of deep learning marked a major advancement in detection capabilities, with models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) significantly improving accuracy. For example, Hu et al. (2021) developed a two-stream CNN that effectively analyzes temporal information in videos. More recently, attention-based models have been incorporated into deepfake detection, enabling the identification of subtle inconsistencies in videos. Wang et al. (2021) proposed an attention-guided method that enhances detection by focusing on spatio-temporal artifacts. Additionally, innovative approaches such as blockchain-based federated learning (BFLDL) have emerged to enhance detection while preserving data privacy. This method standardizes data from diverse sources, helping address challenges related to data heterogeneity and security. Transfer learning techniques have also been employed to improve model performance, especially when dealing with limited labeled data, by allowing models to leverage pre-trained networks for better generalization. However, despite these advancements, deepfake detection continues to face several challenges, including the rapid evolution of generation techniques, the need for high-quality and diverse training data, and the demand for real-time detection capabilities [10].

The emergence of deepfake technology, primarily driven by advancements in Generative Adversarial Networks (GANs), has raised significant concerns regarding misinformation and digital media integrity. Early detection methods relied on traditional image analysis techniques, such as error-level analysis, which struggled to keep pace with the sophistication of deepfake generation [11]. With the advent of machine learning, particularly Convolutional Neural Networks (CNNs), detection capabilities improved significantly. Pretrained models like VGG16 and ResNet50

have been effectively fine-tuned for deepfake detection tasks, although most studies have focused on binary classification (Wu et al. 2017). Recent research has shifted towards multiclass deepfake detection, recognizing the need to categorize various types of manipulated content. The current study introduces a patch-wise deep learning model, leveraging local feature extraction to enhance detection accuracy, marking a significant advancement in the field. Additionally, the development of specialized datasets for training and evaluation is crucial for improving model performance. As deepfake technology continues to evolve, future research must focus on expanding datasets, integrating multimodal data, and addressing ethical implications to ensure the integrity of digital media. The ongoing advancements in detection methodologies are essential for combating the challenges posed by deepfakes and maintaining public trust in digital content [12].

The rise of deepfake technology has spurred significant research into effective detection methods, driven by growing concerns over the authenticity of digital content. Early detection efforts primarily relied on traditional machine learning techniques, such as Support Vector Machines (SVM) and Random Forests, which extracted handcrafted features from videos. However, these methods struggled to accurately detect deepfakes that closely resembled real media, exposing their limitations as deepfake generation techniques continued to evolve. The introduction of deep learning, particularly Convolutional Neural Networks (CNNs), marked a major advancement in detection capabilities. CNNs have proven highly effective in analyzing video frames, identifying subtle discrepancies in facial features and motion. The integration of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks further enhanced detection by capturing temporal information in video sequences, improving accuracy in identifying deepfakes. Generative Adversarial Networks (GANs) have also played a dual role in this field—both in the creation of deepfake content and in improving detection systems. Some studies have even proposed using GANs to generate large datasets of deepfake videos, which help train detection algorithms and strengthen their robustness. Recent innovations, such as the work by Hussain et al., have combined CNNs with optimization algorithms like the Jaya algorithm, leading to significant performance improvements. Their method achieved impressive accuracy rates of 99.3% on the DFDC dataset and 97.6% on the Celeb-DF dataset, demonstrating its effectiveness in identifying temporal anomalies indicative of deepfakes [13].

The literature on deep fake detection has expanded significantly in recent years due to the increasing prevalence of deep fake videos and images on social media platforms. This review highlights various deep learning techniques employed to identify and mitigate the impact of deep fakes, focusing on methods such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks. Deep fakes are generated using advanced techniques like Generative Adversarial Networks (GANs), which create realistic fake content that can mislead viewers. The societal implications of deep fakes are profound, as they can be used for misinformation, defamation, and manipulation, particularly targeting public figures and influencing public opinion. To evaluate the effectiveness

of deep fake detection methods, several metrics are commonly used, including accuracy, precision, and the area under the receiver operating characteristic curve (ROC AUC). These metrics help assess the performance of various algorithms in distinguishing between real and fake content. Recent studies have proposed a range of detection strategies, from traditional machine learning approaches to sophisticated deep learning models. The review emphasizes the importance of continuous research and development in this field to keep pace with the evolving techniques used in deep fake generation [14].

The emergence of deepfake technology has raised significant concerns about the authenticity of digital media. Deepfakes, which use advanced machine learning techniques to generate realistic but fabricated audio and visual content, pose threats to privacy, security, and public trust in media. This literature review synthesizes existing research on deepfake detection methods, particularly focusing on deep learning (DL) approaches, which have gained prominence for their ability to handle complex data patterns effectively. Deepfake technology primarily relies on Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) to produce synthetic media. GANs consist of two neural networks—the generator and the discriminator—that work in opposition to create increasingly realistic outputs. This technology has found applications across various domains, including entertainment, education, and misinformation campaigns, highlighting the dual-use nature of deepfake capabilities. However, the rapid evolution of deepfake generation techniques has outpaced detection methods, presenting significant challenges in identifying manipulated content. Key issues include the diversity of generation methods, the high quality of the output, and the need for real-time detection systems. Additionally, the inaccessibility of datasets and the absence of standardized evaluation metrics complicate the development of robust detection algorithms. Recent studies have explored various deep learning architectures for deepfake detection, with CNNs widely adopted due to their ability to extract spatial features from images. Research by Zhao et al. (2021) and Chen and Tan (2021) demonstrated CNNs' effectiveness in distinguishing between real and fake images by analyzing subtle artifacts often present in deepfakes. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have also been employed to capture temporal features in video data, further improving detection accuracy. Several comparative studies have been conducted to evaluate the performance of different deep learning models in detecting deepfakes, such as the work by Rana et al. (2022), which compared various algorithms and highlighted their strengths and weaknesses. However, the lack of standardized evaluation metrics remains a major challenge, making it difficult to compare results across studies. The literature calls for a unified framework to assess detection algorithms and facilitate progress in the field. Looking ahead, the need for more sophisticated models that can adapt to evolving deepfake generation techniques is urgent. Additionally, improving dataset accessibility and establishing standardized benchmarks for evaluation will be critical for advancing detection methodologies. Interdisciplinary approaches that incorporate insights from psychology, media studies, and ethics are also suggested to provide a more holistic understanding of deepfake technology and its broader implications [15].

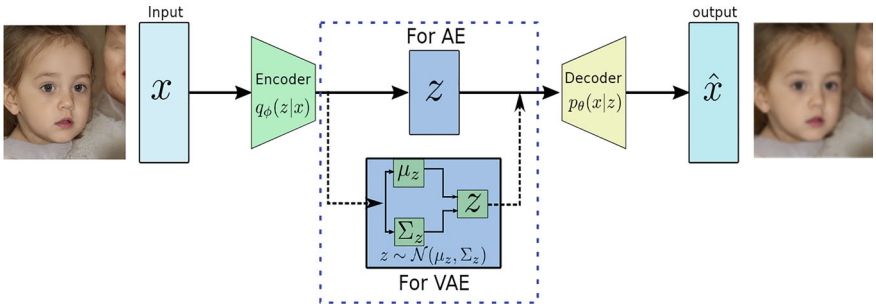


Fig. 2 VAE in deepfake technology

3 Variants of Deepfake Generation Techniques

While GANs are central to most deepfake applications, other architectures and techniques have been developed to refine specific aspects of synthetic media:

- **Autoencoders and Variational Autoencoders (VAEs)**: These compress and reconstruct images and are particularly effective in face-swapping applications. In a face-swapping scenario, an autoencoder trained on one person's face can encode facial features and expressions, which can then be reconstructed on a target face (Fig. 2).
- **Recurrent Neural Networks (RNNs) and Transformers**: These architectures are powerful for sequence-based data and are used in generating synthetic audio and text-based deepfakes. For example, RNNs and Transformer models can learn speech patterns and generate realistic dialogue, including speech synthesis in a specific individual's voice.
- **3D Modeling Techniques**: To enhance realism in video-based deepfakes, especially for applications requiring movement and 3D head tilts, 3D modeling techniques are sometimes integrated, generating more dynamic and versatile synthetic media.

4 Applications of Deepfake Technology

Deepfake technology has rapidly grown across various domains:

- **Entertainment and Gaming**: Synthetic media has been used to bring historical figures to life, enhance movie special effects, and create virtual avatars in gaming.
- **Social Media and Marketing**: Deepfakes have been used in marketing to personalize user experiences, with synthetic personalities or brand ambassadors interacting with users.

- **Education and Training**: Simulation training often uses deepfakes to create realistic scenarios, allowing for immersive learning experiences, particularly in medical or military training.
- Despite these legitimate uses, deepfakes have also emerged as a tool for malicious activities.

5 Ethical and Security Concerns

The proliferation of deepfake technology raises serious ethical and security concerns:

- **Misinformation and Disinformation**: Deepfakes have the potential to spread false information by manipulating influential figures' voices and appearances. In the context of political disinformation, for instance, fake speeches or statements can sway public opinion or incite discord.
- **Privacy Violations and Identity Theft**: By replicating someone's facial appearance or voice, deepfake technology can be used to create non-consensual media, such as unauthorized videos or voice recordings, potentially damaging reputations or compromising privacy.
- **Cybersecurity Risks**: Deepfakes introduce new risks in social engineering and cyberattacks, with malicious actors using synthetic media to impersonate executives, launch phishing scams, or deceive security systems reliant on biometrics.

6 Detection and Countermeasure Development

To counteract these risks, research on deepfake detection has grown alongside advances in generation techniques. Detection approaches often fall into one of several categories:

- **Deep Learning and Machine Learning Approaches**: Classifiers trained to detect specific artifacts—such as inconsistencies in blinking patterns, lighting, or pixel-level anomalies—are commonly used in deepfake detection.
- **Forensic Techniques**: Digital forensic methods analyze media for authenticity by examining metadata, analyzing frequency-domain artifacts, or leveraging digital watermarking to detect tampering.
- **Blockchain for Authenticity**: Emerging approaches suggest using blockchain to track and verify media authenticity through immutable records, providing a decentralized framework for validating content sources.

7 Evolution of Deepfake Technology

- **Role of GANs**: The development of deepfake technology is closely tied to the advent of Generative Adversarial Networks (GANs), introduced by Ian Goodfellow and his colleagues in 2014. GANs comprise two neural networks—a generator and a discriminator—that compete against each other in a process that refines the generation of highly realistic synthetic data. This adversarial training mechanism allows GANs to produce hyper-realistic images and videos, making them the backbone of deepfake technology. Initially used for artistic and creative purposes, such as image synthesis and style transfer, GANs have since become a driving force in the creation of deepfakes, enhancing their visual and audio realism to levels that can be indistinguishable from genuine content.
- **Applications and Risks**: Deepfakes present both opportunities and risks in various sectors. On the positive side, they can be utilized in creative fields such as film production, gaming, and virtual reality, where lifelike simulations enhance user experiences. Additionally, deepfake technology holds potential in accessibility applications, including voice synthesis for individuals with speech impairments and educational simulations. However, the rapid advancement of deepfakes also introduces serious threats. The ability to fabricate convincing fake content can lead to widespread misinformation and the erosion of public trust in digital media. Deepfakes have been weaponized for malicious purposes, such as political manipulation, identity theft, cyber harassment, and fraud. As deepfake generation techniques evolve, so too do their potential for misuse, underscoring the urgency for effective detection systems to safeguard privacy, security, and the integrity of digital content.

7.1 Deep Learning-Based Detection Techniques

7.1.1 Traditional Machine Learning Versus Deep Learning

- Early approaches to detecting deepfakes primarily relied on traditional machine learning methods that utilized handcrafted features, such as pixel-level inconsistencies, facial landmarks, and motion patterns. While these techniques laid a foundational basis for identifying manipulated media, they often struggled to keep pace with the increasing complexity and realism of deepfakes. As deepfake generation methods became more sophisticated, traditional algorithms demonstrated limited flexibility and accuracy in detecting subtle manipulations.
- In contrast, the rise of deep learning brought a transformative shift to the field of deepfake detection. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been widely applied to analyze visual and temporal patterns in media, enabling more robust identification of deepfakes. CNNs, in particular, excel at extracting intricate visual features, making them well-suited for detecting subtle artifacts and inconsistencies introduced during deepfake creation.

This evolution from traditional machine learning to deep learning has significantly improved detection accuracy and enabled scalable solutions.

7.1.2 Advanced Models

- Recent advancements in deepfake detection have focused on refining deep learning models to enhance their effectiveness and adaptability. For example, deep belief networks have demonstrated the ability to identify complex patterns and correlations within manipulated media, boosting detection accuracy. Moreover, attention mechanisms have emerged as powerful tools for improving detection performance by selectively focusing on critical regions of an image or video frame, such as facial expressions and movements. Techniques such as transfer learning have further optimized model performance by leveraging pre-trained neural networks, reducing the computational burden and enhancing generalization across different datasets.
- Collaborative efforts like the Deepfake Detection Challenge, led by organizations such as Facebook and Microsoft, have spurred innovation and the development of new models designed to address the evolving nature of deepfake technologies.

7.1.3 Explainable AI (XAI)

The integration of Explainable AI (XAI) tools, such as Local Interpretable Model-Agnostic Explanations (LIME), has become increasingly important for validating the reliability of deepfake detection models (Fig. 3).

By providing transparency into the decision-making process of models, XAI tools allow researchers and practitioners to understand which specific features of an image or video contribute to classification outcomes. This interpretability not only boosts

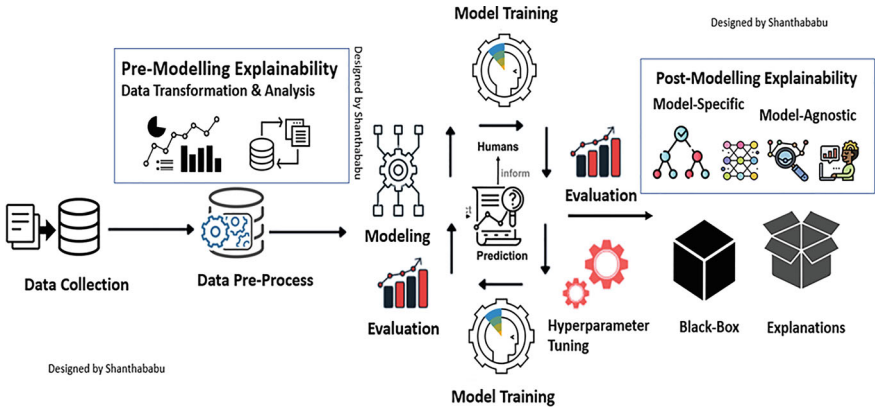


Fig. 3 Explainable AI for AI and ML Engineers

user trust but also aids in the refinement of detection algorithms by identifying areas for improvement. XAI's role in deepfake detection is crucial for ensuring accountability, enhancing model performance, and building confidence in the use of AI systems to combat digital media manipulation.

8 Challenges and Limitations of Current Approaches

- **Scalability**: One of the primary challenges in deepfake detection is scalability, particularly when dealing with large-scale datasets and real-world content. As the volume of synthetic media grows exponentially, detection models face increasing difficulties in processing and analyzing diverse content quickly and accurately. Real-world deepfakes often exhibit higher variability and complexity than those found in controlled datasets, resulting in performance degradation when detection models are deployed in practical scenarios. Ensuring scalability requires continuous improvements in algorithm efficiency, as well as innovative solutions capable of handling large, heterogeneous datasets without compromising accuracy.
- **Adversarial Attacks**: Deepfake detection systems are constantly challenged by the evolving nature of adversarial techniques, which aim to bypass detection models by creating subtle alterations that evade detection. Adversarial attacks exploit vulnerabilities in neural networks, generating modified content that remains highly convincing while eluding classification algorithms. This ongoing "arms race" between deepfake generation and detection necessitates the development of more robust models capable of adapting to new manipulation tactics. It also highlights the importance of building resilient systems that can detect even the most sophisticated and adaptive deepfake variations.
- **Dataset Limitations**: The effectiveness of deepfake detection models heavily depends on the quality and diversity of the datasets used for training and evaluation. Many current datasets suffer from limitations related to size, diversity, and authenticity. Models trained on restricted datasets may struggle to generalize to real-world scenarios, where deepfakes can vary significantly in terms of style, quality, and complexity. Ensuring diverse and representative datasets that capture the full spectrum of deepfake manipulation techniques is essential for improving model robustness. Furthermore, issues such as data scarcity, inconsistent labeling, and the rapid pace of deepfake evolution present additional challenges that need to be addressed through collaborative data collection efforts and standardization initiatives.

9 Emerging Trends and Innovations

- **Collaborative Efforts**: One of the most impactful trends in the fight against deepfakes is the rise of collaborative initiatives, such as the Deepfake Detection Challenge, led by organizations like Facebook and Microsoft. These efforts aim to bring together researchers, industry experts, and academic institutions to advance the development of robust detection technologies. By providing access to large datasets, benchmarking tools, and research incentives, these challenges encourage innovation and accelerate progress in deepfake detection. Such collaborations foster the exchange of ideas, drive innovation, and facilitate the creation of more effective and generalizable models capable of addressing the rapidly evolving deepfake landscape.
- **Federated Learning and Blockchain**: Innovative approaches like federated learning and blockchain-based methods are emerging as powerful tools for enhancing deepfake detection while preserving data privacy and security. Federated learning allows models to be trained on decentralized data sources, meaning that detection algorithms can learn from diverse datasets without compromising user privacy. By distributing the learning process across multiple devices and maintaining data locally, federated learning addresses issues related to data heterogeneity and security.
- Similarly, blockchain technology offers unique advantages in verifying the authenticity of digital content. By providing a tamper-proof and decentralized ledger for recording and tracking media provenance, blockchain can enhance the traceability and transparency of digital media, making it more difficult to alter or falsify content undetected. Together, these technologies present promising avenues for building robust and privacy-preserving deepfake detection systems.
- **Cross-Modal Approaches**: The integration of cross-modal approaches—combining image, audio, and video detection—represents a promising frontier in deepfake detection. As deepfake technology advances across multiple modalities, relying solely on visual analysis is no longer sufficient. Cross-modal approaches allow detection systems to analyze various forms of data, enhancing their ability to identify inconsistencies across different types of media. For instance, combining visual cues, such as facial expressions, with audio signals like voice tone and speech patterns can provide a more comprehensive assessment of a video's authenticity. By leveraging multimodal data, detection systems become more resilient and adaptable, increasing their accuracy and robustness against diverse and complex deepfake threats.

10 Future Directions

- **Enhanced Datasets**: To advance deepfake detection capabilities, there is a pressing need for improved datasets that are larger, more diverse, and reflective of real-world scenarios. Current datasets often suffer from limited diversity

and lack representation of various types of deepfake manipulations, making it challenging for detection models to generalize effectively. Proposals for creating and expanding public datasets include the development of open-source, high-quality datasets that encompass a wide range of content styles, resolutions, and manipulation techniques. Collaborative efforts to curate and standardize datasets can enhance the robustness and adaptability of detection models. Furthermore, building datasets that include new and emerging deepfake methods ensures that models stay relevant as manipulation techniques evolve.

- **Multimodal Data Integration**: Future advancements in deepfake detection will depend heavily on the integration of multimodal data. Given that deepfakes often combine visual, audio, and text elements to create highly convincing content, relying solely on a single modality is insufficient for comprehensive detection. Leveraging different data types allows for a more holistic analysis of manipulated media. For example, combining facial expression analysis with voice pattern recognition and linguistic cues can improve the accuracy and resilience of detection models. Integrating multimodal approaches also facilitates cross-verification, making it more challenging for deepfakes to bypass detection systems by manipulating only one aspect of the media. This direction holds the potential to significantly enhance the robustness of deepfake detection frameworks.
- **Ethical Implications**: As deepfake detection research progresses, it is crucial to address the ethical implications associated with the technology. Privacy and misinformation are among the most pressing concerns. The widespread availability of deepfake tools raises significant risks for individuals, including unauthorized use of their likeness and the potential for harassment or exploitation. Mitigating these risks requires the development of strong legal and ethical frameworks that govern the creation, dissemination, and detection of deepfakes. Researchers and policymakers must collaborate to ensure that detection systems do not inadvertently violate user privacy while effectively countering malicious deepfake content. Ethical considerations must also encompass transparency and fairness in AI-driven detection systems, ensuring that they serve the public good and uphold trust in digital media integrity.

11 Conclusion

Deepfake technology, driven by advancements in artificial intelligence and Generative Adversarial Networks (GANs), has transformed the digital media landscape, offering both innovative possibilities and serious risks. As deepfake content becomes increasingly sophisticated and accessible, the need for effective detection mechanisms grows ever more critical. This review has highlighted the evolution of deepfake detection methods, transitioning from traditional machine learning to deep learning techniques such as Convolutional Neural Networks (CNNs) and attention-based models, which have significantly improved detection accuracy. However, challenges remain, including issues related to scalability, adversarial attacks, and dataset

limitations. Emerging innovations, such as collaborative detection efforts, federated learning, and cross-modal approaches, show promise in addressing these challenges while preserving data privacy and enhancing model resilience. The integration of Explainable AI (XAI) further supports interpretability, fostering transparency and trust in detection models.

Looking ahead, the development of more diverse and comprehensive datasets, multimodal data integration, and ethical considerations will be crucial to staying ahead of evolving deepfake threats. As technology advances, continuous research and collaboration are essential to protect digital media integrity and combat the risks posed by deepfakes. By advancing detection systems, society can better navigate the challenges of this rapidly evolving field and ensure a safer, more trustworthy digital environment.

References

1. Ram, R., Kumar, M., Al-shami, T., Masud, M., Aljuaid, H., Abouhawwash, M.: Deep fake detection using computer vision-based deep neural network with pairwise learning. Intell. Autom. Soft Comput. **35** (2022). https://doi.org/10.32604/iasc.2023.030486
2. Khan, M., Alam, K.N.: Detecting deepfake images using deep learning techniques and explainable AI methods. Intell. Autom. Soft Comput. (2022). https://doi.org/10.32604/iasc.2023.029653
3. Yu, P., Xia, Z., Fei, J., Lu, Y.: A survey on deepfake video detection. IET Biometrics **10**(6), 607–624 (2021). https://doi.org/10.1049/bme2.12031
4. Almars, A.M.: Deepfakes detection techniques using deep learning: a survey. J. Comput. Commun. **09**(05), 20–35 (2021). https://doi.org/10.4236/jcc.2021.95003
5. Juefei-Xu, F., Wang, R., Huang, Y., Guo, Q., Ma, L., Liu, Y.: Countering Malicious DeepFakes: survey, battleground, and horizon. arXiv preprint arXiv:2103.00218 (2021)
6. Mirsky, Y., Lee, W.: The creation and detection of Deepfakes: a survey. ACM Comput. Surv. 54(1), Article 1–41. https://doi.org/10.1145/3425780 (2020)
7. Gambín, Á.F., Yazidi, A., Vasilakos, A., et al.: Deepfakes: current and future trends. Artif. Intell. Rev. **57**, 64 (2024). https://doi.org/10.1007/s10462-023-10679-x
8. Dash, B., Sharma, P.: Are ChatGPT and Deepfake Algorithms Endangering the Cybersecurity Industry? A Review (2023). https://doi.org/10.31873/IJEAS.10.1.01
9. Patel, Y., et al.: Deepfake generation and detection: case study and challenges. IEEE Access **11**, 143296–143323 (2023). https://doi.org/10.1109/ACCESS.2023.3342107
10. Heidari, A., Navimipour, N.J., Dag, H., et al.: A novel blockchain-based deepfake detection method using federated and deep learning models. Cogn. Comput. **16**, 1073–1091 (2024). https://doi.org/10.1007/s12559-024-10255-7
11. Rafique, R., Gantassi, R., Amin, R., et al.: Deep fake detection and classification using error-level analysis and deep learning. Sci. Rep. **13**, 7422 (2023). https://doi.org/10.1038/s41598-023-34629-3
12. Arshed, M.A., Mumtaz, S., Ibrahim, M., Dewi, C., Tanveer, M., Ahmed, S.: Multiclass AI-generated deepfake face detection using patch-wise deep learning model. Computers **13**(1), 31 (2024). https://doi.org/10.3390/computers13010031
13. Hussain, Z.F., Ibraheem, H.R.: Novel convolutional neural networks based Jaya algorithm approach for accurate deepfake video detection. mesopotamian. J. CyberSecurity 35–39 (2023).https://doi.org/10.58496/MJCS/2023/007

14. Mary, A., Edison, A.: Deep fake detection using deep learning techniques: a literature review. In: 2023 International Conference on Control, Communication and Computing (ICCC), Trivandrum, India (2023). https://doi.org/10.1109/ICCC57789.2023.10164881
15. Heidari, A., Jafari Navimipour, N., Dag, H., Unal, M.: Deepfake detection using deep learning methods: a systematic and comprehensive review. In: WIREs Data Mining and Knowledge Discovery, vol. 14, no. e1520 (2024). https://doi.org/10.1002/widm.1520
16. Ansari, M.F., Sharma, P.K., Dash, B.: Prevention of phishing attacks using AI-based cybersecurity awareness training. Int. J. Smart Sens. Adhoc Netw. **3**(3), 61–72. https://doi.org/10.47893/ijssan.2022.1221
17. Westerlund, M.: The emergence of deepfake technology: a review. Technol. Innov. Manag. Rev. **9**(11) (2019)
18. Sharma, P., Dash, B., Ansari, M.F.: Anti-phishing techniques–a review of cyber defense mechanisms. IJARCCE **11**(7), 153–160 (2022). https://doi.org/10.17148/IJARCCE.2022.11728.
19. Burroughs, S.J., Gokaraju, B., Roy, K., Khoa, L.: Deepfakes detection in videos using feature engineering techniques in deep learning convolution neural network frameworks. In: 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). https://doi.org/10.1109/aipr50011.2020.9425347
20. Dash, B., Ansari, M.F.: An Effective Cybersecurity Awareness Training Model: First Defense of an Organizational Security Strategy (2022)
21. Dash, B.: A hybrid solution for extracting information from unstructured data using optical character recognition (OCR) with natural language processing (NLP) (2021)
22. Dash, B., Sharma, P.: Role of artificial intelligence in smart cities for information gathering and dissemination (a review). Acad. J. Res. Sci. Publishing **4**(39) (2022). https://doi.org/10.52132/Ajrsp.e.2022.39.4
23. Güera, D., Delp, E.J.: Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE, USA (2018)
24. Gaur, L., Arora, G.K., Jhanjhi, N.Z.: Deep learning techniques for the creation of deepfakes. DeepFakes 23–34 (2022). https://doi.org/10.1201/9781003231493-3
25. Guarnera, L., Giudice, O., Nastasi, C., Battiato, S.: Preliminary forensics analysis of deepfake images. In: 2020 AEIT International Annual Conference (AEIT) (2020). https://doi.org/10.23919/aeit50178.2020.9241108
26. Kumar, B., Alraisi, S.R.: Deepfakes audio detection techniques using deep convolutional neural network. In: 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) (2022). https://doi.org/10.1109/com-it-con54601.2022.9850771
27. Nguyen, T.T., Nguyen, Q.V., Nguyen, D.T., Nguyen, D.T., Huynh-The, T., Nahavandi, S., Nguyen, T.T., Pham, Q.-V., Nguyen, C.M.: Deep learning for deepfakes creation and detection: a survey. SSRN Electron. J. (2022). https://doi.org/10.2139/ssrn.4030341
28. Sharma, M., Kaur, M.: A review of deepfake technology: an emerging AI threat. Adv. Intell. Syst. Comput. 605–619 (2021). https://doi.org/10.1007/978-981-16-5301-8_44
29. Almars: Media Forensics and Deepfake: systematic survey. In: Media Forensics and Deepfake – Systematic Survey · Issue 8 (pubpub.org) (2021)

Framework for a Non-intrusive Monitoring Solution for Independent Elderly



Halldor Broddi Thorsteinsson, Anne-Lena Kampen, Kai Goebel, and Maneesh Singh

Abstract As the global population ages, the need for reliable and non-intrusive systems to safeguard the well-being of elderly individuals living independently is growing. Traditional monitoring methods, such as wearable devices, can be inconvenient and inconsistently used, making them less effective. A preferable solution is a passive system that functions without requiring direct user interaction, ensuring privacy while providing continuous support. This study explores the use of floor-based vibration sensors to monitor movement patterns and detect anomalies, with an emphasis on identifying falls. The objective is to evaluate whether unique vibration signatures can reliably differentiate between routine activities and potential hazards, forming the foundation for cost-effective monitoring systems. By examining the accuracy and feasibility of this technology, the research seeks to improve senior people's independence and safety, while offering caregivers valuable insights to enable more proactive and informed care. First part of the project focuses on developing a home laboratory to gather data through vibration sensors and establishing a baseline of normal activity. Once this baseline is set, the sensors will monitor for deviations that could indicate a fall. The goal is to create a system capable of detecting falls, thereby enhancing the safety of older adults living independently. This paper presents a detailed overview of the project's motivation, objectives, and scope, as well as the proposed framework, setup, and software architecture.

Keywords Monitoring · Detection · Aging in place · Vibration signals

H. B. Thorsteinsson (✉) · A.-L. Kampen · M. Singh
Western Norway University of Applied Sciences, Bergen, Norway
e-mail: halldorорsteinsson@gmail.com

A.-L. Kampen
e-mail: anne-lena.kampen@hvl.no

M. Singh
e-mail: maneesh.singh@hvl.no

K. Goebel
SRI International, Menlo Park, CA, USA
e-mail: kai.goebel@rocketmail.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_40

1 Introduction

The World Health Organization (WHO) states that "*By 2030, 1 in 6 people in the world will be aged 60 years or over. At this time, the share of the population aged 60 years and over will increase from 1 billion in 2020 to 1.4 billion. By 2050, the world's population of people aged 60 years and older will double (2.1 billion). The number of persons aged 80 years or older is expected to triple between 2020 and 2050 to reach 426 million*" [1].

As the global population continues to age, the demand for autonomous and reliable systems to assist older adults living independently is growing. Ensuring their safety and well-being is essential, with a particular focus on detecting and responding to falls—a leading cause of injury. A report from WHO shows that "*Fall-related injuries are more common among older persons and are a major cause of pain, disability, loss of independence, and premature death. Approximately 28–35% of people aged 65 and over fall each year, increasing to 32–42% for those over 70 years of age*" [2].

The significance of this research is highlighted by the growing global challenge of an aging population, the increasing desire of older adults to remain at home, and the shortage of caregivers. By introducing a passive, non-intrusive monitoring system, this technology can ensure timely assistance, reducing response times and potentially saving lives. This research also supports the concept of "aging in place," enabling older adults to remain in their homes longer rather than transitioning to assisted living facilities. Aging in place benefits both individuals and the healthcare system by allowing seniors to maintain their social connections and daily routines [3].

While various solutions have been developed to address this need, each presents distinct challenges. Wearable devices, such as smartwatches and alarm bracelets, are common fall detection tools. These devices can notify caregivers or emergency services when a fall is detected or when the user manually activates an alarm. However, these technologies have notable limitations. Smartwatches are ineffective while charging, and alarm bracelets depend on the wearer being conscious and able to press the button—an action that may not be feasible following a severe fall.

Alternatively, in-home sensor systems—such as video cameras, audio detection, and motion sensors—provide a more passive approach to monitoring. While these systems can be effective, they often raise significant concerns regarding privacy and cost. Video and audio-based solutions are frequently seen as intrusive, causing discomfort for residents. Research indicates that the primary concern among smart home users is the collection and use of data through video and audio monitoring [4].

Vibration sensors are emerging as a promising, privacy-preserving method. In a study that focused on fall detection in industrial settings, the G-Link 200 vibration sensor, paired with a convolutional neural network (CNN) algorithm, achieved an impressive 99.6% detection rate [5]. Another study employed three geophones arranged in a square formation to create a sensor network, achieving a fall detection accuracy of 95% [6].

Some factors that make vibration-based detection particularly well-suited for fall detection are:

1. **Privacy Preservation**: Unlike cameras, which can be intrusive and capture sensitive information, vibration-based systems monitor movement discreetly, preserving individual privacy by avoiding the collection of personal data [4].
2. **No Dependence on Wearable Devices**: Vibration-based systems remove the reliance on wearable devices such as smartwatches or alarm bracelets, which often present challenges in maintaining continuous data collection. Wearables require frequent charging and are easily forgotten, especially by individuals in the early stages of Alzheimer's or dementia. A study found that in 80% of falls where individuals were alone, the wearable alarm was not activated [7]. Vibration-based monitoring provides uninterrupted coverage without the need for user interaction, ensuring more consistent and reliable detection.
3. **Room-Wide Coverage Without Line-of-Sight Issues**: Unlike camera-based systems, which struggle in low-light conditions, vibration sensors reliably monitor activity regardless of visibility, providing comprehensive coverage even in dark environments. Thus, vibration sensors can detect movement across an entire room, overcoming the line-of-sight restrictions faced by cameras and motion sensors [8].
4. **Independence from Room Layout and Furniture Configuration**: The key advantage of this approach lies in its ability to function independently of room layout or changes in furniture configuration, ensuring cost efficiency and non-intrusiveness. This adaptability allows for reliable fall detection regardless of variations in room configuration.

However, despite its effectiveness, the G-Link 200's cost—over 6500 NOK (600 USD) on eBay—limits its feasibility for residential use.

Researchers have also explored the combination of vibration sensors and microphones to detect sound waves generated by falls, resulting in a 97.5% detection rate [9]. Another study utilized gyroscopes, accelerometers, and mobile phones to achieve 97.37% accuracy in fall detection [10] demonstrating the potential of combining multiple sensors to improve accuracy. However, increasing the number of sensors can add complexity and raise costs, presenting challenges for widespread adoption.

2 Motivation and Aim of Research

2.1 Motivation

Detecting falls in elderly individuals is essential for safeguarding their well-being. As more seniors choose to live independently, there is a rising need for passive monitoring systems that enhance care without requiring constant human supervision.

Vibration-based detection offers a practical, non-intrusive solution by effectively monitoring movement and identifying falls.

A well-designed vibration detection system must reliably differentiate between various types of vibrations, accurately distinguishing routine daily activities, genuine falls, and false positives, such as dropped objects. This ensures the system provides precise fall detection while minimizing unnecessary alerts.

2.2 Aim of the Research

The ongoing project aims to assess the feasibility of using floor-based vibration sensors for accurate and reliable fall detection in elderly individuals living independently.

Through the analysis of vibration patterns generated by different movements, employing signal processing and machine learning techniques, the research seeks to identify fall events while minimizing false positives and distinguishing falls from routine daily activities.

The ultimate goal is to develop a non-intrusive, cost-effective monitoring system that enhances the safety of elderly individuals aging in place, while preserving their privacy and ensuring their comfort.

2.3 Scientific Novelty and Importance of Research

The scientific innovation of this research lies in adapting ISO standards—typically applied in industrial settings—to enable fall detection in home environments. Specifically, the study implements ISO 17359 as a methodological framework and ISO 13374-2 for software development. In alignment with ISO 17359, the project employs Failure Modes, Effects, and Criticality Analysis (FMECA) to classify types of falls, assess their potential consequences, and outline appropriate response strategies.

By following this structured framework, the project seeks to enhance current technology through the development of software tools capable of accurately interpreting data from vibration sensors. When necessary, the system will provide timely recommendations for intervention, contributing to improved safety for elderly individuals living independently.

3 Materials and Methods

3.1 Proposed Solution

This project proposes a fall detection system that leverages floor-based vibration sensors to monitor the movements of elderly individuals.

Building on previous work [11], this project will utilize geophones as the primary vibration sensors, paired with an Analog-to-Digital Converter (ADC) to digitize the signals. An amplifier and filter will be used to refine the data by enhancing relevant signals and reducing background noise. The sensor data will be transmitted to a microcontroller, which will process the signals and relay them to a Raspberry Pi over Bluetooth for further analysis.

Figure 1 illustrates the framework for key steps involved in setting up a condition monitoring system in accordance with ISO 17359.

The first three steps of the approach contribute to developing a maintenance plan using the Traditional Method:

- **Equipment Audit**: Aims to understand the function of each component and define the expected operational range for optimal performance.
- **Reliability and Criticality Audit**: Evaluates the machine's reliability, availability, maintainability, and safety (RAMS). The standard recommends applying Failure Modes and Effects Analysis (FMEA) or Failure Mode, Effects, and Criticality Analysis (FMECA) to gain deeper insights.
- **Selection of an Appropriate Maintenance Strategy**: Focuses on identifying a maintenance approach that ensures the machine operates efficiently and effectively.

The next three steps refine the maintenance plan by incorporating system condition insights:

- **Selection of Monitoring Method**: Guides the identification of parameters to measure and the appropriate monitoring techniques. Key factors considered at this stage include feasibility, accuracy, monitoring intervals, data acquisition rates, and the placement of monitoring equipment.
- **Data Acquisition and Analysis**: Focuses on preprocessing and analyzing collected data to facilitate diagnostic and prognostic assessments.
- **Determination of Maintenance Action**: Recommends actions necessary to ensure smooth machine operation or implement corrective measures related to performance and maintenance.

ISO 13374 Parts 1–4 form a series of international standards that provide guidelines for designing open condition monitoring and diagnostics (CM&D) software systems. ISO 13374-2 focuses specifically on the *Data Acquisition and Analysis* phase, offering a detailed framework for system architecture. Figure 2 illustrates the ISO 13374-2 recommendation to structure the system into six functional data

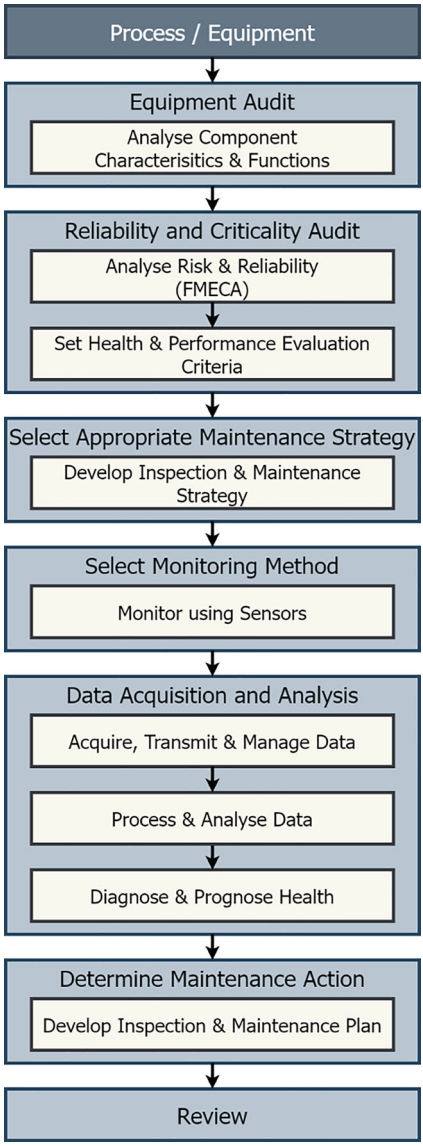


Fig. 1 Main steps of a monitoring system (Based on ISO17359)

processing blocks. Each block performs distinct tasks, progressively transforming and storing data as it moves through the sequence.

The six processing blocks in the architecture include:

- **Data Acquisition**: Collects raw, real-time data on structural and operational conditions using sensors.

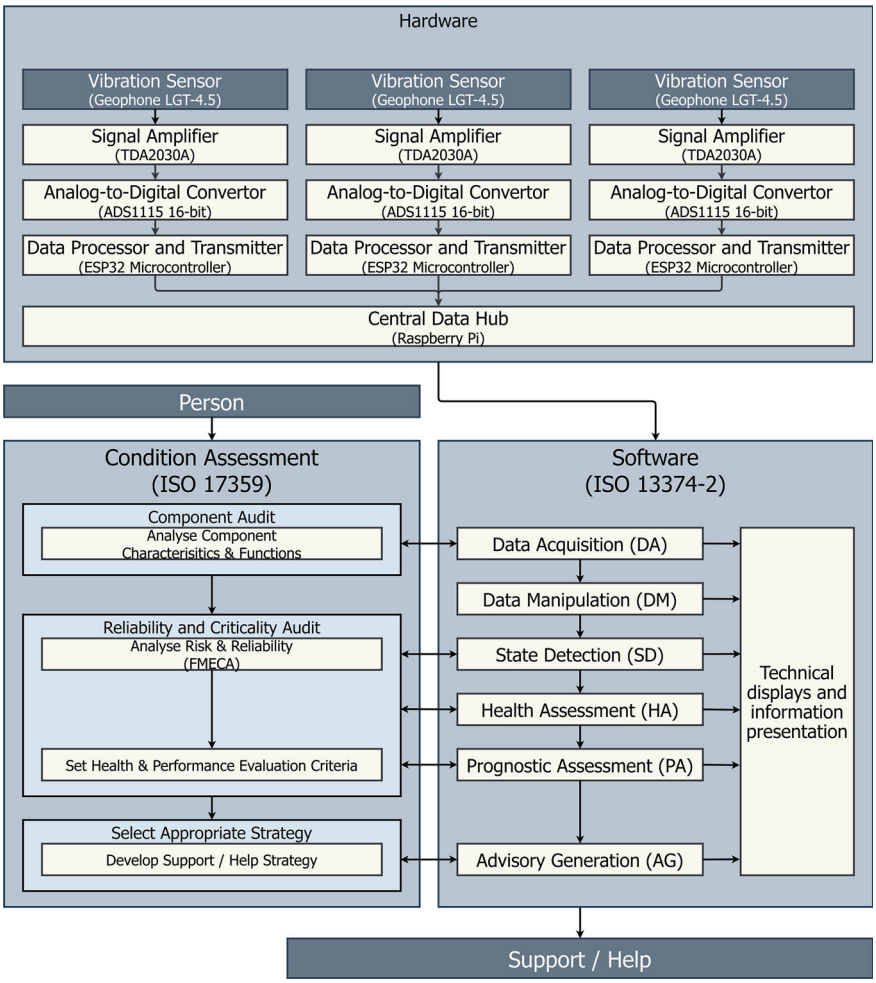


Fig. 2 Flowchart illustrating the concept for detecting fall

- **Data Manipulation**: Cleans and processes the acquired data by correcting for drift, noise, outliers, and missing values.
- **State Detection**: Analyzes the processed data to identify anomalies or deviations from normal operational behavior.
- **Health Assessment**: Evaluates the system's overall condition based on detected anomalies and operational data.
- **Prognostic Assessment**: Predicts fault progression and estimates the remaining useful life (RUL) of deteriorating components using status reports.
- **Advisory Generation**: Delivers actionable recommendations, guiding inspection and maintenance efforts.

Drawing inspiration from these standards, Fig. 2 also introduces a conceptual framework for monitoring elderly individuals to detect adverse events, such as falls. This framework consists of three key components:

- **Hardware**: Focuses on selecting and installing appropriate sensors and electronic components to monitor the individual's condition and transmit data to a logging system.
- **Condition Assessment**: Outlines the causes of falls, detection methods, potential consequences, and necessary response actions.
- **Software**: Establishes the framework for analyzing sensor data to detect falls and interpret the results, supporting informed decision-making to assist the individual.

3.2 Laboratory Setup

The laboratory setup is designed to simulate a home environment, enabling data collection, testing of fall detection accuracy, and assessment of system reliability to support ongoing refinement and improvement. It has the following hardware components:

- **Vibration Sensor (Geophone LGT-4.5)**: The LGT-4.5 geophone monitors floor vibrations and converts them into electrical signals, serving as the primary source of data for fall detection. Originally designed for seismology and structural monitoring, the LGT-4.5's high sensitivity to low-frequency vibrations allows it to detect subtle movements with precision. This makes it an ideal choice for fall detection in residential settings, where accurate monitoring of slight disturbances is crucial.
- **Signal Amplification (TDA2030A Amplifier)**: The TDA2030A amplifier boosts the analog signals generated by the LGT-4.5 geophone before they are processed by the Analog-to-Digital Converter (ADC). By amplifying these signals, the amplifier enhances clarity and ensures that even faint but critical vibrations, such as those caused by falls, are accurately detected and distinguished from background noise. In the laboratory setup, the TDA2030A amplifier is positioned between the LGT-4.5 geophone and the ADS1115 ADC, amplifying raw vibration data prior to digital conversion, ensuring the integrity and precision of the collected signals.
- **Analog-to-Digital Conversion (ADS1115 16-bit ADC)**: The ADS1115 Analog-to-Digital Converter (ADC) converts amplified analog signals from the geophone into digital data for further analysis. By sampling incoming signals at precise intervals, the ADC translates voltage levels into digital values that reflect the amplitude and frequency of detected vibrations. Positioned immediately after the TDA2030A amplifier, the ADS1115 ensures that amplified signals are accurately digitized, preserving data integrity for subsequent processing.
- **Data Processing and Transmission (ESP32 Microcontroller)**: The ESP32 Devkit microcontroller manages initial data processing and facilitates the transmission of sensor data to the central hub (Raspberry Pi). Selected for its robust

processing power and integrated wireless communication features, the ESP32 is ideal for efficient, wireless data transfer. Its built-in Bluetooth Low Energy (BLE) functionality ensures reliable, low-power communication with the Raspberry Pi.

- **Central Data Hub and Analysis (Raspberry Pi)**: The Raspberry Pi serves as the central hub for data collection and storage from multiple sensors. Serving as the primary aggregator, it receives data packets from each ESP32, consolidating inputs from various points to create a comprehensive overview of activity patterns within the environment. The Raspberry Pi enables real-time analysis and long-term data logging, supporting accurate fall detection and broader system performance monitoring.

3.3 Condition Assessment

The condition assessment draws inspiration from ISO 17359 to monitor elderly individuals and provide timely assistance in the event of a fall. This approach integrates condition assessment into the broader Failure Modes, Effects, and Criticality Analysis (FMECA) framework, outlining the causes, types, consequences, detection methods, and appropriate responses to fall events.

Component Audit focuses on identifying and analyzing various factors that contribute to falls, including physical health, environmental conditions, and technological support. This audit lays the groundwork for a comprehensive FMECA analysis by examining how these elements interact and influence fall risk.

- **Human Factors**: Evaluates the individual's physical health, cognitive function, and mobility. Key considerations include muscle weakness, balance issues, visual impairments, and cognitive conditions such as dementia, all of which can increase the likelihood of falling [12].
- **Environmental factors**: Assesses the living environment to identify potential hazards, such as poor lighting, uneven surfaces, or obstacles that could contribute to falls [12].
- **Technological support**: the use and effectiveness of assistive devices, such as canes, walkers, and other mobility aids [12].

By addressing these areas, the condition assessment aims to provide a holistic understanding of fall risks, supporting the development of more effective monitoring and intervention strategies.

Reliability and Criticality Audit (FMECA) expands the component analysis by applying a Failure Modes, Effects, and Criticality Analysis (FMECA) framework. This phase identifies potential fall scenarios, assesses their severity, and informs the design of targeted interventions to mitigate risks.

- **Causes of Fall (Failure Mechanism)**: A thorough analysis of failure mechanisms, such as physical weakness, medication side effects, and environmental hazards,

helps uncover the underlying causes of falls. Falls frequently result from a combination of factors, including reduced mobility, medications, environmental risks, and unsafe conditions. Understanding how these elements interact is crucial for accurately interpreting data and identifying fall events [13].

  - **Physical Conditions**: Balance issues, gait disturbances, vision impairments, sensory deficits, muscle weakness, cognitive decline, and fluctuations in blood pressure or heart rate all contribute to fall risk [12].
  - **Medications**: Certain medications can impair balance, alertness, or blood pressure, increasing the likelihood of falls [12].
  - **Environmental Hazards**: Uneven surfaces, poor lighting, slippery floors, obstacles, and unfamiliar surroundings pose significant fall risks [12].

- **Types of Fall (Failure Mode)**: Type of fall will depend upon the cause of fall. Falls are categorized based on different types of falls, such as slips, loss of balance, or sudden collapse. Identifying the type of fall aids in creating tailored support strategies.
- **Consequence of Fall**: Consequence of fall varies depending on its type and severity. Assessing potential outcomes, such as fractures, head injuries, or minor bruises, helps prioritize response measures. This evaluation is vital for determining the urgency of responsive actions [13].
- **Monitoring and Detection of Fall**: The selection of monitoring technologies and data analysis methods is influenced by the anticipated type of fall. In addition to vibration sensors, other technologies, such as wearable devices, may be integrated to enhance fall detection accuracy and ensure timely support when an incident occurs [6, 12, 13].
- **Required Actions**: An effective and timely response from caregivers or emergency services relies on accurately assessing the nature and impact of the incident. This ensures that the appropriate level of support and intervention is delivered, aligned with the severity and specific circumstances of the fall.

Health and Performance Evaluation: Focuses on assessing health and performance metrics to gauge the effectiveness of implemented strategies and the overall stability of the individual. Key criteria include fall frequency, severity of consequences, and recovery capability:

- **Fall Frequency**: Tracking the number of falls over time helps evaluate the reliability of interventions. An increase in fall frequency may signal deteriorating health or indicate insufficiency of the that current preventive measures.
- **Severity of Consequences**: Each fall is assessed based on the severity of physical injuries and psychological impacts. More severe outcomes may necessitate a reevaluation of strategies.
- **Recovery Capability**: The individual's ability to recover independently from falls is a critical indicator of resilience. Observations following a fall can reveal whether additional support or interventions are necessary to improve overall well-being.

Strategy Development: Builds on findings from condition assessments to design targeted interventions that support the individual's health and minimize the risk of future falls.

- **Preventive Strategies**: Strategies for making environmental adjustments to remove tripping hazards, plays a crucial role in reducing fall risk [14].
- **Emergency Response Protocols**: Establishing clear emergency response protocols ensures swift action in the event of a fall. These protocols incorporate fall detection technologies that immediately notify caregivers or medical services, facilitating prompt intervention.

3.4 Software Structure

Data acquisition is done by collecting data from vibration sensors connected to ESP32 microcontrollers and stored on a Raspberry Pi in.csv files for further analysis and record-keeping. Upon receiving data packets, a Python script running on the Raspberry Pi timestamps each entry with the current UTC time, ensuring precise chronological tracking of all events. This timestamp is essential for monitoring and distinguishing fall events from routine vibrations.

The data collection script then organizes each data packet into a structured format that includes several fields, like, Timestamp, Sensor ID, Vibration Amplitude, and Data Quality Indicator. This structured data is appended as a new row in the .csv file, forming the primary data archive.

Data manipulation (DM) block enhances fall detection accuracy by reducing noise in raw sensor data. Vibration signals often contain background interference that can hinder the identification of falls. To address this, the system applies filtering techniques to isolate significant vibration events, improving sensitivity to falls while minimizing false positives. A 20-Hz Butterworth high-pass filter is used to eliminate low-frequency noise, preserving vibrations associated with human movement. The Butterworth filter's flat frequency response in the passband minimizes distortion, ensuring a clean, stable signal ideal for fall detection [11].

Additionally, the DM block utilizes an energy-based dual-threshold mechanism to further differentiate between falls and non-fall events. This mechanism sets two calibrated amplitude thresholds: one to detect high-intensity signals indicative of falls, and another to filter out low-level background noise [11].

These processes are integrated into the Raspberry Pi's data pipeline, ensuring that only clean, high-quality data is fed into the fall detection algorithm, enhancing overall system reliability.

State Detection (SD) block evaluates processed data from the DM Block to classify events into specific states: "Nothing Going On", "Normal Activity", "Potential False Positive" and "Fall Detected". These states are determined based on vibration characteristics such as amplitude, frequency, event duration, and the context surrounding each event.

The "Fall Detected" state is further refined into two subcategories to capture variations in fall dynamics, recognizing that elderly individuals may fall in different ways. One category is the "Hard Fall," characterized by high-impact signals followed by inactivity, distinctly different from routine activities, which generally produce lower, continuous vibrations. The second is the "Controlled Fall," where an individual senses a potential fall and transitions more slowly to the ground. In this case, the data shows lower-intensity signals, followed by inactivity, which distinguishes it from regular movement.

A machine learning model, trained on labeled data from previous falls and normal activities, enhances classification accuracy. When a fall is detected, the system triggers an alert, while other states are logged for reference and monitoring.

Health Assessment (HA) block analyzes data from the SD Block to assess health trends and identify fall risk factors. Rather than focusing solely on individual falls, the HA Block tracks patterns over time to evaluate overall health stability. For instance, an increasing frequency of "Hard Falls" compared to "Controlled Falls" may indicate declining physical resilience, while a higher number of prolonged post-fall inactivity periods could suggest a growing risk of serious injury.

The HA Block monitors the duration of post-fall inactivity as one indicator of injury severity, comparing it with fall characteristics, such as impact intensity, to assess the seriousness of each event. It can help determine whether the individual might have lost consciousness or is struggling with recovery after falls. Additionally, the HA Block analyzes the time of day at which falls occur to identify possible behavioral or environmental contributors, such as fatigue, effects of medication, or other factors that may vary over a daily routine. The location of falls is also considered—if an individual frequently falls when rising from the sofa, for instance, this could indicate a need for assistive support in that area.

Increased Fall Frequency: The HA Block tracks and analyzes fall frequency to identify upward trends, signaling potential declines in balance or physical strength. A rise in fall incidents may trigger recommendations for proactive measures, such as physical therapy, balance exercises, or home modifications to reduce future fall risks.

Escalation in Fall Severity: By examining fall characteristics over time, the HA Block tracks and analyzes fall severity, from controlled, low-impact falls to more severe, high-impact events, may indicate a diminishing ability to brace or manage falls, heightening the risk of serious injury. This insight helps caregivers and medical professionals implement preventive measures, such as additional safety equipment or medical evaluations.

Indicators of Declining Mobility and Stability: Patterns of controlled falls or increased unsteadiness can indicate a gradual decline in mobility. The HA Block flags such trends, prompting considerations for assistive devices or enhanced monitoring. This predictive capability helps maintain the individual's independence by addressing emerging stability issues early.

Environmental and Behavioral Risks: The HA Block also assesses environmental and behavioral factors contributing to falls. By analyzing the time, location, and circumstances surrounding fall events, the system identifies high-risk scenarios. For example, if falls frequently occur at night or in specific areas (e.g., near the bed or bathroom), the system may recommend environmental adjustments, such as improved lighting or grab bars, to mitigate risks.

Long-Term Independence and Health Trajectory: Synthesizing data on fall frequency, severity, and recovery times, the HA Block offers a comprehensive view of the individual's long-term ability to live independently. A pattern of increasing falls, escalating severity, and prolonged recovery may signal the need for greater intervention, such as home care, medical assessments, or changes in living arrangements. This information supports caregivers and families in making informed decisions to safeguard the individual's health, safety, and autonomy.

By continuously analyzing these patterns, the HA Block helps caregivers and healthcare professionals proactively address fall risks and improve overall safety for elderly individuals.

Prognostic Assessment (PA) block analyzes historical data, fall patterns, and recovery trends to predict future outcomes for the monitored individual. Beyond immediate fall detection, this assessment anticipates potential scenarios, providing valuable insights into the individual's stability, fall risk, and overall health trajectory. A key focus is determining whether the person is likely to recover independently from future falls ("Get Up Independently"), while also identifying emerging risks that could affect their ability to live independently.

The PA Block evaluates past fall data to predict the likelihood of the individual recovering without assistance after future falls. If previous falls consistently show independent recovery with minimal support, the system predicts a high chance of similar outcomes in the future. Conversely, if trends indicate increasing post-fall inactivity or frequent reliance on assistance, the system may forecast a lower probability of independent recovery. This insight enables caregivers to plan for additional support, addressing potential mobility and resilience challenges.

Advisory Generation (AG) block is responsible for making recommendations and initiating actions based on insights from the Health Assessment (HA) and Prognostic Assessment (PA) Blocks. In a fall detection system, the AG Block's primary function is to determine if an emergency response is required following a fall. This decision considers the severity of the event, its characteristics, and patterns in the individual's recent activity data.

Criteria for Emergency Advisory: The AG Block evaluates whether to initiate an emergency call using predefined criteria, including:

- **Severity of the Fall**: Falls classified as "Hard Falls" (high-impact with prolonged inactivity) are more likely to trigger an emergency response. High-impact falls carry a greater risk of serious injuries, such as fractures or head trauma, requiring immediate intervention.

- **Post-Fall Inactivity Duration**: Prolonged inactivity following a fall signals potential injury or unconsciousness. If no movement is detected within a specified period (e.g., two minutes), the AG Block may automatically issue an emergency advisory.
- **Patterns of Declining Mobility or Frequent Falls**: If the PA or HA Blocks identify increasing fall frequency or reduced recovery capacity, the AG Block adjusts its sensitivity. This allows for quicker emergency responses, recognizing heightened risk.

Decision-Making Logic: The AG Block follows a structured, criteria-based decision-making process:

- **Immediate Emergency Response**: For "Hard Falls" with prolonged inactivity or signs of serious injury, the AG Block directly contacts emergency services.
- **Notification to Relevant Caregiver/Emergency Response Team**: For "Controlled Falls" (low-impact, self-managed), the AG Block notifies caregivers instead of immediately calling emergency services. Caregivers assess the situation and determine if further action is required.
- **Monitoring and Follow-Up**: If a minor fall or activity does not meet emergency criteria but suggests potential risk, the AG Block logs the event and monitors for subsequent incidents. If additional falls or concerning patterns emerge, the system escalates to an emergency advisory.

Advisory Options: Based on the assessment, the AG Block can initiate various responses:

- **Initiate an Emergency Call**: Directly contacts emergency services for high-risk falls.
- **Send Alerts to Caregivers or Family Members**: Notifies caregivers, family members, or neighbors via mobile app or messaging service, allowing them to check on the individual. This is used for less severe incidents that do not require immediate medical attention.
- **Log the Event for Health Tracking**: Records non-emergency falls for future health assessments, contributing to trend analysis and long-term health monitoring.

By tailoring responses to the severity and context of each fall, the AG Block ensures timely and appropriate action, enhancing the safety and well-being of elderly individuals while reducing unnecessary emergency calls.

4 Conclusions

This research contributes to the expanding field of fall detection solutions for elderly individuals through the use of monitoring devices. It presents a detailed framework for a monitoring system designed to support independent living among seniors.

The novelty of this solution lies in the application of ISO 17359 and ISO 13374-2 standards—typically used in industrial environments—to develop concepts for elderly monitoring. By adapting these established methodologies, the project introduces a structured, data-driven approach to enhance safety and well-being in home settings.

Beyond elderly care, the principles explored in this research have potential applications across various fields. Since the system primarily detects changes in vibration patterns, it could be adapted for use in areas such as structural health monitoring, industrial machinery, bridge traffic analysis, railway tracks, ship hull integrity, mining equipment, power transformers, pipelines, and more. This versatility highlights the broad potential of vibration-based monitoring systems.

Future work will involve the installation of sensors in a home environment and the development of the necessary software tools to operationalize the system. Additionally, the research will explore integrating multiple sensors—such as door magnets, power consumption meters, water flow monitors, and TV usage sensors—to enhance decision-making and provide a comprehensive overview of daily activities. This holistic approach aims to create a more reliable, multifaceted monitoring system to further support aging in place.

References

1. World Health Organization: Ageing and health WHO (2022) (Online Article). Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed: 25 Oct 2024
2. World Health Organization: Global age-friendly cities: a guide, WHO Press, Geneva, Switzerland (2007) (Online). Available: https://www.who.int/publications/i/item/9789241563536. Accessed: 25 Oct 2024
3. Ratnayake, M., Lukas, S., Brathwaite, S., Neave, J., Henry, H.: Aging in place: are we prepared? Dela. J. Public Health **8**(3), 28–31 (2022). https://doi.org/10.32481/djph.2022.08.007. PMID: 36177171; PMCID: PMC9495472
4. Haney, J.M., Furman, S.M., Acar, Y.: User Perceptions of Smart Home Privacy and Security, NISTIR 8330, National Institute of Standards and Technology, Gaithersburg, MD, USA (2021) (Online). Available: https://doi.org/10.6028/NIST.IR.8330. Accessed: 29 Oct 2024
5. Huang, Y., Chen, W., Chen, H., Wang, L., Wu, K.: G-fall: device-free and training-free fall detection with geophones. In: 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Boston, MA, USA, pp. 1–9 (2019). https://doi.org/10.1109/SAHCN.2019.8824827
6. Maharani, M.P., Putri, A.R., Lee, J.M., Kim, D.-S.: Vibration sensor-based fall detection in smart factory shop floor. In: 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, Republic of, pp. 762–765 (2021). https://doi.org/10.1109/ICTC52510.2021.9621173
7. Fleming, J., Brayne, C., and the Cambridge City over-75s Cohort (CC75C) study collaboration: Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. BMJ **337**, 2227 (2008) (Online). Available: https://doi.org/10.1136/bmj.a2227. Accessed: 25 Oct 2024
8. Igual, R., Medrano, C., Plaza, I.: Challenges, issues and trends in fall detection systems. BioMed Eng OnLine **12**, 66 (2013). https://doi.org/10.1186/1475-925X-12-66

9. Litvak, D., Gannot, I., Zigel, Y.: Detection of falls at home using floor vibrations and sound. In: 2008 IEEE 25th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, pp. 1–5 (2008). https://doi.org/10.1109/EEEI.2008.4736581
10. Basu, I., Ray, S., Sarkar, S., Saha, S., Panda, S.: Advancements in fall detection technologies: a comprehensive review and innovative solutions. In: 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, pp. 1–3 (2023). https://doi.org/10.1109/IEMENTech60402.2023.10423535
11. Wu, K., Huang, Y., Qiu, M., Peng, Z., Wang, L.: Toward device-free and user-independent fall detection using floor vibration. ACM Trans. Sens. Netw. **19**(1), article 5, 1–20 (2023) (Online). Available: https://doi.org/10.1145/3519302. Accessed: 20 Oct 2024
12. NIA: Falls and Fractures in Older Adults: Causes and Prevention. National Institute on Aging (2023). Available: https://www.nia.nih.gov/health/falls-and-falls-prevention/falls-and-fractures-older-adults-causes-and-prevention. Accessed: 8 Nov 2024
13. Berg, R.L., Cassells, J.S. (eds.) Institute of Medicine (US) Division of Health Promotion and Disease Prevention. The Second Fifty Years: Promoting Health and Preventing Disability. National Academies Press, Washington, DC (1992). Available: https://www.ncbi.nlm.nih.gov/books/NBK235613/
14. Karlsson, M.K., Vonschewelov, T., Karlsson, C., Cöster, M., Rosengen, B.E.: Prevention of falls in the elderly: a review. Scandinavian J. Public Health **41**(5), 442–454 (2013). https://doi.org/10.1177/1403494813483215

Assessment of *ProgPy* as an Open-Source Condition Monitoring and Diagnostics Tool



Magnus S. Nordeide, Anne-Lena Kampen, Chetan S. Kulkarni, Mayank S. Jha, and Maneesh Singh

Abstract Traditional maintenance programs, such as corrective and preventive strategies, may lead to high costs and operational inefficiencies. Condition Monitoring and Diagnostics (CM&D) aims to improve these maintenance strategies by enabling timely insights into equipment health and performance. However, implementation of CM&D can be challenging without a robust framework that manages data efficiently, supports interoperability and simplifies integration. To address these challenges ProgPy, an open-source Python-based prognostics tool developed by NASA Ames Research Center, offers a structured solution for broader Prognostics and Health Management (PHM) applications. An ongoing research is assessing the feasibility of implementing ProgPy as a Condition Monitoring and Diagnostics (CM&D) solution by comparing its framework to the guidelines for open CM&D systems recommended in the ISO 13374-2 standard. This evaluation aims to highlight ProgPy's strengths and identify opportunities for improvement, thereby, contributing to its advancement as an effective tool for Prognostics and Health Management (PHM). This paper presents the results of an initial assessment of the ProgPy toolbox. To demonstrate the working of ProgPy, it also includes diagnostics and prognostics of a gearbox using open-source datasets.

M. S. Nordeide (✉) · A.-L. Kampen · M. Singh
Western Norway University of Applied Sciences, Bergen, Norway
e-mail: magnus.nordeide@gmail.com

A.-L. Kampen
e-mail: anne-lena.kampen@hvl.no

M. Singh
e-mail: maneesh.singh@hvl.no

C. S. Kulkarni
NASA Ames Research Center, KBR Inc., Mountain View, CA, USA
e-mail: chetan.s.kulkarni@nasa.gov

M. S. Jha
Polytech Nancy, Université de Lorraine, Vandœuvre-Lès-Nancy, France
e-mail: mayank-shekhar.jha@univ-lorraine.fr

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_41

Keywords Condition-monitoring · Diagnostics · Failure · Gearbox · ISO 13374 · Maintenance · Prognostics · ProgPy

1 Introduction

Traditional maintenance strategies for industrial equipment, for example, wind turbine gearboxes, have predominantly been corrective and preventive in nature. Corrective maintenance, the most common approach, is a reactive approach where repairs or replacements are made only after a failure occurs. While this method is commonly used, this approach often results in unplanned downtimes. In large systems such as wind turbines, these downtimes can be particularly prolonged and costly due to the complex designs and challenging accessibility. The unpredictability of failures under corrective maintenance not only disrupts energy production but also results in significant maintenance expenses.

To reduce chances of unplanned failures, preventive maintenance involving scheduled inspections and routine servicing based on predetermined time-intervals, usage metrics or risk assessments are commonly used. This proactive approach aims to reduce the likelihood of equipment failure by addressing potential issues before significant degradation takes place. However, preventive maintenance does not consider the actual condition of equipment, hence components may be replaced or serviced unnecessarily, resulting in increased maintenance costs without proportional improvements in reliability [1].

Preventive and corrective maintenance strategies each have distinct strengths and weaknesses. Traditional maintenance approaches are widely adopted across industries due to their familiarity, ease of implementation, predictability and minimal reliance on high-quality data. However, despite these advantages, they can sometimes be unreliable, inefficient, and costly [2].

To address the limitations of traditional maintenance approaches and improve equipment reliability, condition monitoring solutions have been proposed. These solutions integrate various techniques for data acquisition, processing, diagnostics, prognostics, and predictive maintenance planning. By continuously monitoring the equipment and analyzing the collected data, this approach may be able to predict the remaining useful life Remaining Useful Life (RUL), End-of-Discharge (EOD), and End-of-Life (EOL) of components and determine optimal times for inspection and maintenance.

Several standards guide structured procedures and frameworks for condition monitoring, supporting early fault detection and recommending predictive maintenance. Notable among them, ISO 17359 provides general guidelines for machine condition monitoring, while ISO 13374 focuses on data processing and management. ISO 13379 and ISO 13381 offer guidance on data interpretation and prognostics, helping estimate time-to-failure. Together, these standards enable consistent, reliable monitoring practices essential for proactive maintenance.

NASA's open-source Prognostics Python Package (ProgPy) provides a comprehensive framework for addressing key prognostic metrics, including RemainingUseful Life (RUL), End-of-Discharge (EOD), and End-of-Life (EOL) by facilitating use of physics-based, data-driven, and hybrid modeling approaches. The package includes tools for state estimation, enabling precise health state assessments essential for accurate predictions. Its uncertainty quantification framework supports evaluation of RUL, EOD, and EOL under diverse conditions, ensuring reliability in safety–critical applications. A notable application of ProgPy is in battery degradation analysis, where it models phenomena such as capacity fade and internal resistance growth. These analyses offer insights into battery performance limits, predicting EOD—the point at which a battery can no longer deliver sufficient power—and EOL, typically defined by a capacity threshold (e.g., 80%) below which a battery is considered non-functional. ProgPy integrates state-of-charge (SOC) estimation with degradation models to compute probabilistic distributions for EOD and EOL, accounting for environmental and operational uncertainties. Beyond battery systems, ProgPy has been applied across industrial and aerospace sectors, demonstrating its flexibility in adapting to various systems. Its ability to accommodate diverse modeling paradigms and deliver robust predictions makes it a valuable tool for advancing prognostics, supporting predictive maintenance, and enhancing reliability engineering [3].

2 Motivation and Aim of the Research

2.1 Aim of the Research

The aim of this research is to evaluate the feasibility and effectiveness of using ProgPy as an open-source tool for condition monitoring and diagnostics.

The study involves a comparative analysis of ProgPy's architecture against the framework recommended by the ISO 13374-2 standard. ProgPy's performance in prognostics and health assessment is being tested using a publicly available dataset from a faulty gearbox.

The findings are expected to highlight key similarities and differences between the two architectures. Any gaps or limitations in ProgPy's alignment with ISO 13374-2 may reveal opportunities for enhancing ProgPy to better meet industry standards.

2.2 Scientific Novelty and Importance of Research

This paper evaluates ProgPy's suitability for gearbox prognostics and health assessment. The novelty of this work lies in the comparative analysis of ProgPy's architecture against the framework recommended by ISO 13374-2 for open Condition Monitoring and Diagnostics (CM&D) software systems.

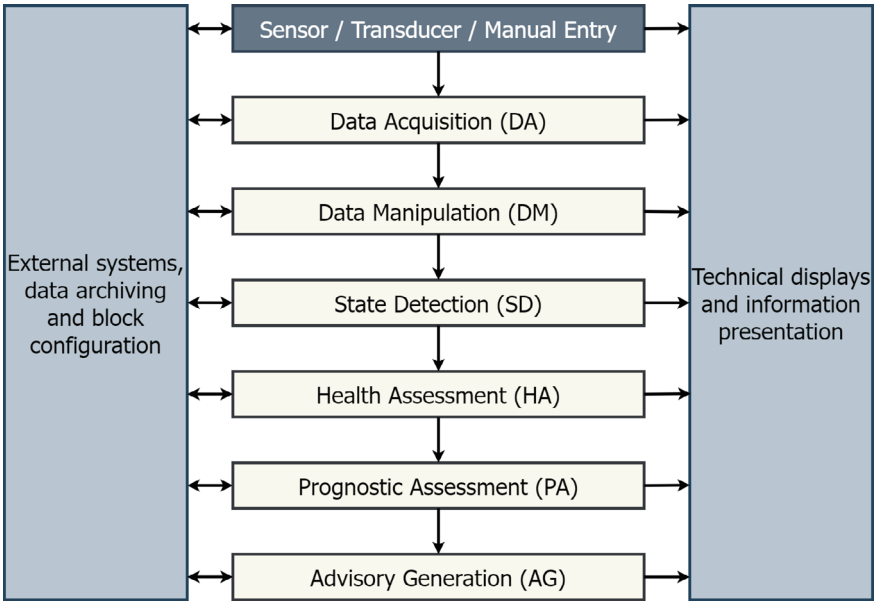


Fig. 1 Flowchart depicting the architecture of an open condition monitoring and diagnostics (CM&D) software system as recommended by ISO 13374-2

3 Architecture of a Condition Monitoring and Diagnostics (CM&D) Software System

3.1 ISO 13374 (Condition Monitoring and Diagnostics of Machines)

ISO 13374 Parts 1–4 are a set of international standards that provide guidelines for creating open Condition Monitoring and Diagnostics (CM&D) software systems. Figure 1 shows recommendation of ISO 13374-2 (Data Processing, Communication, and Presentation) for organizing the architecture of the system into eight functional data processing blocks. Each block performs specific tasks, transforming and archiving data as it moves to the next block in the sequence.

3.2 Data Acquisition (DA)

The Data Acquisition block is responsible for collecting raw, real-time data on structural and operational conditions using transducers or sensors (such as accelerometers, thermocouples, and acoustic sensors) mounted on equipment (example, gearbox).

The objective is to ensure data is collected in a structured, consistent and reliable manner for further processing in subsequent stages. This block receives inputs in the form of analog, digital, and/or manual signals, while its outputs consist of digitized data, often including both Coordinated Universal Time (UTC) and local time-stamped records.

3.3 Data Manipulation (DM)

The Data Manipulation block receives input from the Data Acquisition block and cleanses it by reducing imperfections such as drifts, noise, outliers, and missing values. It then processes the cleansed data using signal processing techniques, including filtering, normalization, Fourier transforms, wavelet analysis, and feature extraction, to transform raw data into clean, rational information ready for further analysis by subsequent blocks.

3.4 State Detection (SD)

The primary function of the State Detection block is to analyze the cleaned data from the Data Manipulation block to identify deviations or anomalies from expected normal behavior in structural and operational conditions. The goal is to classify the system's state and detect the early signs of potential issues. Outputs from this block include anomaly alerts, rate-of-change notifications, threshold boundary alarms, and enumerated state indicators [4].

3.5 Health Assessment (HA)

When deviations or anomalies have been detected by the previous block, the Health Assessment block performs a thorough evaluation to assess the overall condition of the system. Using techniques such as signature analysis, it diagnoses and quantifies faults to provide a comprehensive health status. In gearbox condition monitoring, this includes detailed analysis to identify and measure degradation mechanisms like bearing wear, gear tooth damage, or gear pitting. The output from this block includes a report detailing the current condition of the equipment [4].

3.6 Prognostic Assessment (PA)

The Prognostic Assessment block utilizes the equipment status report from the previous block to predict the progression of faults and estimate the remaining useful life (RUL) of degraded components. This is achieved using predictive models—physical, data-driven, or hybrid—that factor in anticipated future load conditions. Expected outputs from this block include RUL estimates, projections of future health conditions, and the likelihood of potential failures.

3.7 Advisory Generation

The Advisory Generation block is the final stage in the Condition Monitoring & Diagnostics (CM&D) process. It provides actionable recommendations based on the analyses conducted in all preceding block. This block thus aids in planning inspection and maintenance activities.

3.8 External Systems and Block Configuration

External Systems, Block Configuration support all previously described blocks by providing configuration, integrating external data, and preserving historical information for continuous improvement. External systems provide access to maintenance records and operational data, such as start/stop events and load histories. Block configuration involves the setup and tailored configuration to each block to operate them effectively. To support long-term reliability, historical data is archived to enable statistical analysis. Together, these components ensure the condition monitoring system is capable of supporting timely and informed decision-making.

3.9 Technical Displays and Information Presentation.

Technical displays and Information Presentation also support all previously described blocks by facilitating the visualization and interpretation of their outputs. ISO 13374-2 specifies that technical displays should provide data analysts with essential information to support corrective action decisions by converting complex data into a clear and presentable format, enabling operators to understand the system's state. When anomalies occur, users may also need the capability to delve into the State Detection, Data Manipulation, and Data Acquisition displays for deeper analysis [4].

4 Overview of ProgPy

ProgPy's architecture is built on three core components: Prognostic Models, Prognostic Engine and Prognostic Support Tools [5]. These components enable researchers and developers to build predictive models, estimate system health states, and visualize outcomes across applications.

- The **Prognostic Models** can be physics-based, data-driven or hybrid type.
- With the **Prognostics Engine**, users can apply these models to estimate health states and make prognostic predictions, creating a flexible, extendable framework.
- The **Prognostics Support Tools** assist with model tuning, benchmarking and visualization, making it easier to develop and refine predictive solutions.

In Fig. 2 Prognostic Models is represented by "Prognostic Models", Prognostics Engine is represented by "State Estimator", "Predictor" and "Analysis", while the Prognostic Support Tools is not explicitly outlined in the figure. However, the "Analysis" component can be understood to incorporate these tools, as it provides functionalities for evaluating, visualizing and benchmarking prognostic results.

ProgPy has been implemented in several NASA projects and has proven to be a valuable tool. For example, use of ProgPy for the prognostics and health management of UAV batteries offers operators with essential insights into battery status and provides support for safe operations [3].

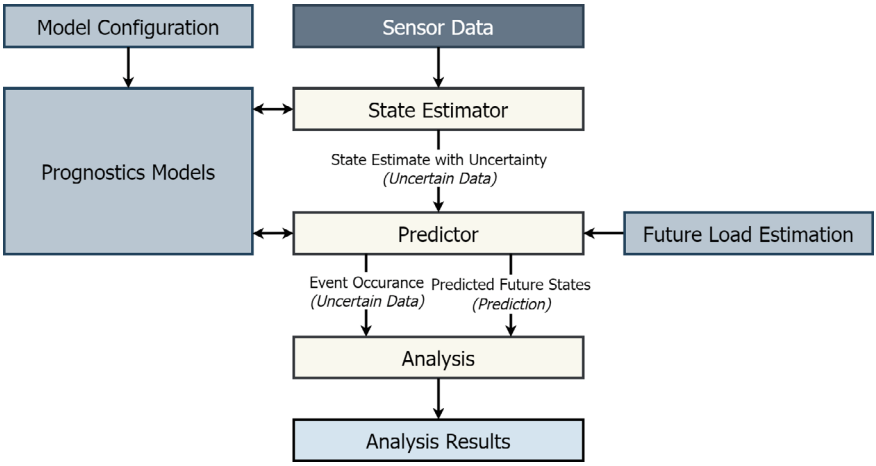


Fig. 2 Flowchart depicting the architecture of ProgPy

4.1 Prognostic Models

ProgPy takes all the input information and passes it through a complex prognostics models to predict the current (through State Estimator) and future (through Predictor) state of the system. For these tasks ProgPy already has a library of prognostic models, including, some physics-based models for batteries, pumps, DC motors, etc., and data-driven models based on LSTM [6]. More models can be added as required.

Physics-based models simulate system dynamics using first-principle equations, making them ideal for systems with well-understood processes. For instance, in battery systems, these models capture degradation phenomena like solid electrolyte interphase (SEI) growth and lithium-ion diffusion, enabling accurate predictions of Remaining Useful Life (RUL), End-of-Discharge (EOD), and End-of-Life (EOL).

On the other hand, Data-driven models, utilizing techniques such as regression, neural networks, and ensemble methods, infer patterns from historical data. They are particularly suited for complex systems where the underlying physics is unknown or too intricate to model explicitly. These models excel when abundant operational data is available, offering insights into trends and degradation mechanisms.

Hybrid models merge the strengths of both approaches, using physics-based constraints to guide data-driven algorithms. This reduces overfitting while enhancing interpretability and robustness. For example, hybrid models can effectively model degradation in mechanical systems or rotary machines by combining physical parameters with data-driven adaptability.

In the case of high-fidelity models that are computationally expensive, substitute models can be made which serve the purpose of approximating the behavior of the model to make inference time shorter. “Surrogate models” are models which serve the purpose of approximating the behavior of another model to shorten inference time or for testing purposes [7]. Can be useful in the cases of high-fidelity models.

In addition, ProgPy also supports “Composite” and “Ensemble” models [8]. A Composite Model combines multiple models to represent interconnected systems, such as linking a gearbox model with a motor model. Ensemble models is a mixture of models to leverage different strengths, useful for cases where different models predict aspects of system behavior. Ensemble models mix different models to benefit their unique strengths, for example when various models predict different aspects of system behavior.

Core elements for the configuration of these prognostic models include “Inputs”, “Outputs”, “States” and “Events” [7].

- Inputs represent how the system is loaded and external factors influencing the system, such as the voltage and external load applied to a motor or the temperature surrounding the motor.
- Outputs include measureable characteristics of the system, such as the torque or speed of a motor, vibration, current drawn etc.
- States define the internal conditions of the system, encapsulated by the StateContainer in ProgPy, which assess the system’s health and operational capacity.

- Events mark specific points of interest, like when a failure threshold is passed. In the context of prognostics, these are often Remaining Useful Life (RUL) and Time-of-Event (ToE). Systems may have multiple failure modes, each tracked as a separate event.

ProgPy models operate primarily as "state transition models", predicting the next state based on the current state, external and internal factors (inputs and outputs). When the state of the system passes an explicitly defined threshold, an event is triggered, and relevant data is stored and used for analysis. ProgPy can also operate through a "direct-prediction" approach that estimates RUL directly from current state of system without any state tracking. This open framework provides a comprehensive and intuitive pipeline for extensive PHM applications [7].

With integrated tools for state estimation and uncertainty quantification, ProgPy supports robust analysis across diverse scenarios. This versatility makes it invaluable for predicting and analyzing critical failure modes, such as battery degradation and mechanical wear, across a broad range of applications.

4.2 Sensor Data

Sensor data may include temperature readings, vibrations, torque, current and voltage, pressure and more. Multiple sensor input may improve the state estimation of a system [9].

4.3 State Estimator

State estimation is necessary if the state of the system isn't directly measurable [7]. It is a description of a system's state and should be available to assess health and to propagate for further prognostics.

State Estimator estimates the internal state of a system using calibration methods (Kalman filter [7, 8], Unscented Kalman filter [7, 8], Particle filter [7], etc.). ProgPy introduces flexibility to its framework by providing helper methods along with already implemented filters, including the option to add custom state estimators. ProgPy's state estimate comes with the respective uncertainty quantification, represented by the ProgPy Uncertain Data structure.

4.4 Predictor

After estimating the current state of the system, ProgPy's Predictor can be used to predict its future states or faults. ProgPy generalizes faults as events, which may also include other points of interest within the system [7].

For a prediction to take place, an initial state estimation, a prognostics model and an estimate of future load are required. The prognostics can be done without uncertainty quantification and is then referred to as "Simulation" by ProgPy, or it can be done with an uncertainty quantification using methods such as Monte Carlo Simulation or Unscented Transform [7].

4.5 Analysis

After results from preceding blocks (state estimation and prediction) are generated, Analysis in ProgPy is designed to process and interpret the outputs.

ProgPy provides various data structures to manage simulation and prediction data. The results of State Estimation are stored in an object of type UncertainData, which includes several methods for analyzing the distribution, such as calculating the mean, median and displaying a summary of key distribution metrics in a readable format [7]. The methods and output of the analysis is also stored under object type Prediction. Predictions can be made of multiple types of data, including future internal states, inputs, outputs and state of health (ProgPy event state). These analyses give information on statistics of the distributions and metrics.

4.6 Additional Features of ProgPy

ProgPy provides a Service-Oriented Architecture test implementation, designed to perform prognostics on engineering systems through a REST API [7]. The Prognostics-as-a-Service approach includes a server providing access to prognostic functions, and a client for interfacing with the server. Given the increasing shift towards cloud-based analysis and storage in prognostics [6], ProgPy is in step with this trend by implementing Prognostics-as-a-Service functionality.

5 Method

The research is conducted in two parts. The first part compares ProgPy's architecture with the architecture recommended by the ISO 13374-2 standard.

In the second part, ProgPy is evaluated using the publicly available MCC5-THU gearbox fault diagnosis dataset. The dataset has been generated for a two-stage parallel gearbox operating under variable speeds and loads by manually introducing a spectrum of fault types and severity levels, ranging from single faults (such as gear wear, pitting, wear, cracks, and missing teeth) to compound faults involving both gears and bearings through precise machining [10].

The dataset comprises of 240 time series recordings of vibration, speed, and torque signals, sampled at 12.8 kHz over 60-s intervals. By capturing the vibrations of the 36-tooth gear on the intermediate shaft and its support bearings, this dataset offers valuable insights into how faults manifest in vibration signals across different operational conditions [10]. Figure 3 shows the organization of the dataset.

In this example, the dataset has been transformed into visual representations as spectrograms by converting the time-domain signal into a frequency-domain signal using the Short-Time Fourier Transform (STFT). Figure 4 presents a spectrogram generated from a five-second segment of an experiment where a fault has been introduced by removing a tooth from a gear on the intermediate shaft. The spectrogram displays changes in frequency intensity at regular intervals, offering a comprehensive view of all frequencies and enabling visualization of fault signatures.

6 Study of ProgPy's Architecture

In this part each functional block in ProgPy is compared to the corresponding block in the ISO 13374-2 standard architecture (Fig. 5). The comparison examines the operation of individual blocks to identify ProgPy's strengths and limitations relative to the guidelines from the ISO standard.

6.1 Data Acquisition

ProgPy has capability to directly collect data streams from sensors, but data acquisition is not managed directly; instead, it relies on external systems or tools for any prior data cleanup.

In the gearbox prognostics example, the data was obtained in digital format and was directly utilized for further analysis.

6.2 Data Manipulation

As discussed earlier, ProgPy does not include built-in data manipulation functionalities such as feature extraction, normalization, or filtering. These preprocessing steps

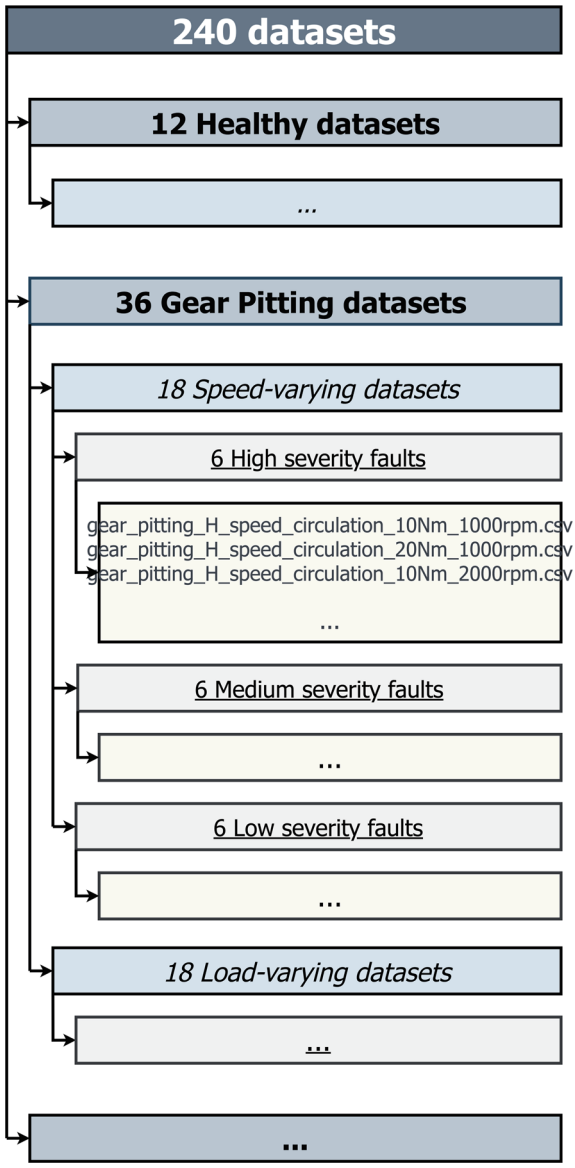


Fig. 3 Flowchart representation of structure of MCC5-THU gearbox fault diagnosis dataset

are expected to be performed externally using tools like NumPy, SciPy, or Pandas before the data is input into ProgPy.

Initial approaches for the data manipulation of the gearbox prognostics example may include envelope analysis and filtering for isolating frequencies of interest, time–frequency analysis, time-domain and frequency-domain analyses, including FFT spectrum analysis. Filtering the signal may be relevant because vibration signals

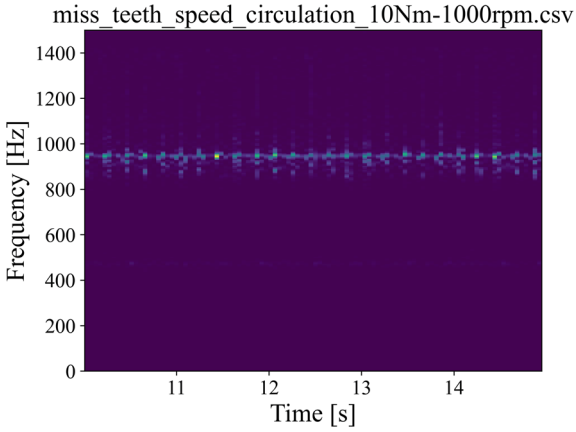


Fig. 4 Spectrogram generated from vibration data of a gear with a missing tooth

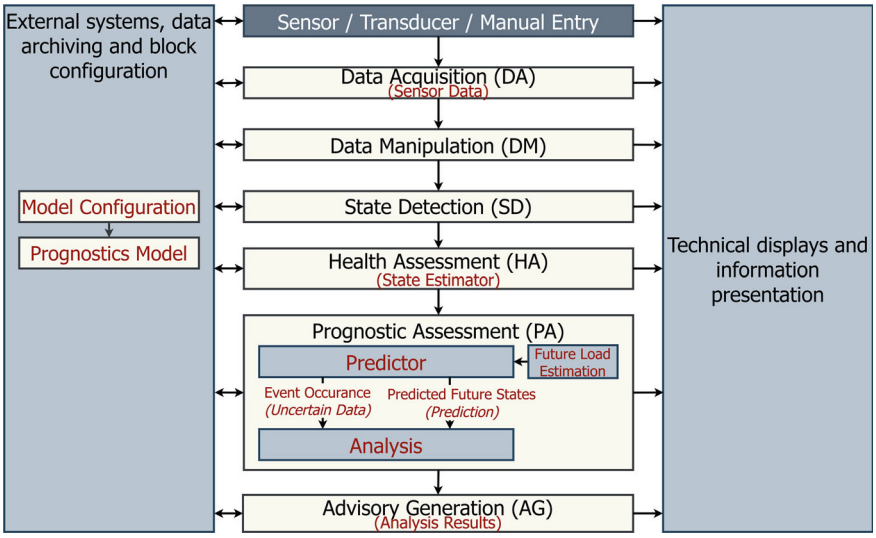


Fig. 5 Flowchart illustrating the architecture of ProgPy compared to the architecture recommended by the ISO 13374-2 standard project flowchart

indicative of faults can be overshadowed by other signals with higher amplitudes [11, 12].

Additionally, if the prior data acquisition process resulted in inconsistencies such as missing rows, sampling issues or data corruption, cleaning the data could also be part of the data manipulation stage. However, since the data in the gearbox prognostics example was acquired in a clean and organized manner, additional structural cleaning hasn't been investigated.

6.3 State Detection

ProgPy features a component known as State Estimation, which assesses the current health or operational state of a system based on input data. ProgPy leverages selected prognostic models to continuously evaluate the system's condition.

A key difference between the State Detection function in ISO 13374-2 and ProgPy's State Estimation is that the ISO approach typically recommends to identify discrete states, such as "normal," "warning," or "failure," using rules, thresholds, or basic pattern recognition techniques. In contrast, ProgPy's State Estimation provides a more detailed continuous assessment—often estimating a numerical degradation level by applying physics-based or machine learning models.

State Detection detects anomalies by identifying deviations from baseline operation. While ISO 13374-2 recommends a standalone feature for State Detection, ProgPy does not have this independent functionality; rather it integrates this block with the subsequent block of Health Assessment.

In the gearbox prognostics example, faults can be detected by comparing spectrograms of healthy and faulty gears using techniques like Convolutional Neural Networks (CNNs).

6.4 Health Assessment

The Health Assessment block in ISO 13374-2 focuses on evaluating the type and severity of detected faults. In ProgPy, the State Estimation functional block partially fulfils this role. It generates degradation metrics that quantify the system's health on a continuous scale, providing valuable insights into fault severity. However, since diagnostic tools are not included in ProgPy, any fault identification or classification is carried out externally.

In the gearbox prognostics example, each fault type produces a distinct spectrogram pattern, which can be used for identification. Figure 6 shows the speed profile of the gearbox run and the time-interval of interest for generation of spectrograms. Figure 7 displays spectrograms generated from five-second segments of experiments with various introduced fault types in the frequency range of 400–1250 Hz. The frequency dips in the last second of Fig. 7 spectrograms are caused by speed decrease, showing a discrepancy between the set speed profile, as shown in Fig. 6, and the actual speed profile. These spectrograms can be further analysed using techniques like Convolutional Neural Networks (CNNs). Additionally, it may be possible to quantify fault severity by calibrating frequency intensities against fault degrees.

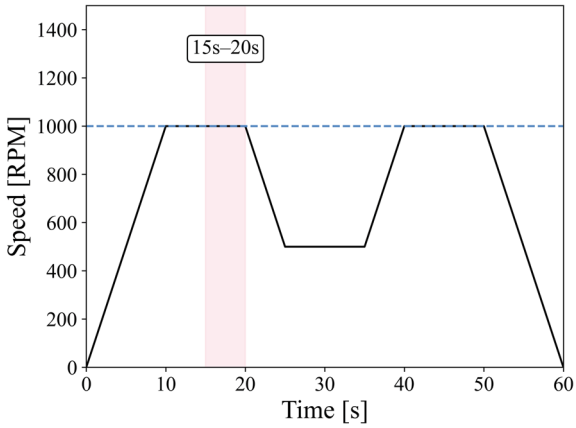


Fig. 6 Speed profile of gearbox run and time-interval of interest

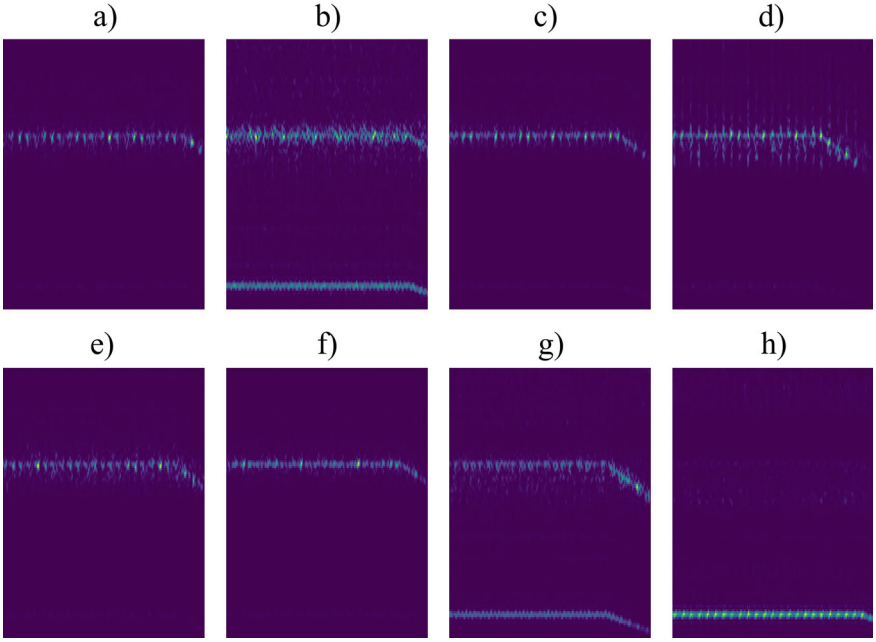


Fig. 7 Spectrograms from interval 15–20 s during stable RPM at 1000 rpm and constant torque at 10 Nm. **a** Healthy **b** Gear pitting **c** Gear wear **d** Missing teeth **e** Teeth break **f** Teeth crack **g** Teeth break and bearing inner ring **h** Teeth break and bearing outer ring

6.5 Prognostic Assessment

ProgPy includes a specialized Prognostic Assessment block (Predictor), making it well-suited for estimating remaining useful life (RUL). This block allows for

modeling system degradation patterns and predicting RUL under varying load conditions, along with associated uncertainties. For effective performance, this block requires reliable prognostic models, accurate inputs from preceding blocks and estimates of future loading. When the system's current condition and expected usage is well understood, ProgPy can deliver precise RUL estimates, along with quantified uncertainties.

In the context of ISO 13374-2, the Predictor aligns with the standard's Prognostic Assessment block, where future states and time-to-event estimations are calculated to enable timely maintenance actions. This predictive capability supports proactive decision-making by estimating events such as RUL, aligning with the structure of ISO 13374-2.

In the gearbox prognostics example, this step has not yet been investigated.

6.6 Advisory Generation

In ProgPy, the "Analysis" function closely aligns with the Advisory Generation block in ISO 13374-2. ProgPy's analysis can include visualizing predictions and generating actions based on model outputs, offering valuable recommendations to support informed decision-making.

In the gearbox prognostics example, this step has not yet been investigated.

6.7 External Systems and Block Configuration

In ProgPy, the Block Configuration house the details of prognostic models, which are utilized for making predictions and calculating the remaining useful life (RUL). In addition, it allows for saving results from simulations, enabling storing of outputs such as estimated current states, estimated future states, and event predictions such as RUL predictions.

ProgPy does not include built-in interfaces for accessing external systems such as maintenance records and operational data. However, this limitation can be addressed by using external tools to interface with it.

6.8 Technical Displays and Information Presentation

ProgPy includes features for visualizing calculations using various plot types. For more complex or specialized visualizations, users can generate external plots using the calculation outputs. Its analysis tools provides detailed outputs of the distributions

connected to state estimation and prediction, however, to fully conform to the standard's technical display requirements, additional features for interactive drill-down, visualization, and comprehensive anomaly analysis may be necessary.

7 Conclusions

This paper explores the suitability of ProgPy as an open Condition Monitoring and Diagnostics (CM&D) software tool. A comparison of its architecture against that recommended by the ISO 13374-2 standard reveals that ProgPy's design aligns closely with the recommended structure. ProgPy's flexible architecture enables users to integrate various NASA technologies into a cohesive framework. Its open structure allows models, state estimators, and predictors to be easily modified, eliminating the need for users to adapt their cases to fit the software.

The comparative study has also identified some areas for improvement. ProgPy may be improved for broader adoption in industry by including a library of tools for data manipulation, state detection, diagnostics, interfacing with external systems and additional features for technical displays.

References

1. Huang, C., Bu, S., Lee, H.H., Chan, C.H., Kong, S.W., Yung, W.K.: Prognostics and health management for predictive maintenance: a review. J. Manuf. Syst. **75**, 78–101 (2024)
2. Wagner, C., Hellingrath, B.: Implementing predictive maintenance in a company: industry insights with expert interviews. In: 2019 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 1–8 (2019)
3. Teubert, C., Jarvis, K., Corbetta, M., Kulkarni, C., Daigle, M.: ProgPy: Python packages for prognostics and health management of engineering systems. J. Open Sour. Soft. **8**(87), 5099 (2023)
4. International Organization for Standardization (2004) ISO 13374-2: Condition Monitoring and Diagnostics of Machines—Data Processing, Communication, and Presentation—Part 2: Data Processing (2004)
5. Keeter, B.: 2024 Software of the Year Award Co-Winner—Prognostics Python Packages (ProgPy) (2024). https://www.nasa.gov/organizations/otps/2024-software-of-the-year-award-co-winner-prognostics-python-packages-progpy/. Last accessed 11 Aug 2024
6. Guo, J., Li, Z., Li, M.: A review on prognostics methods for engineering systems. IEEE Trans. Reliab. **69**(3), 1110–1129 (2020)
7. Teubert, C., Jarvis Griffith, K., Corbetta, M., Kulkarni, C., Banerjee, P., Watkins, J., Daigle, M.: ProgPy Python Prognostics Packages (2023). https://github.com/nasa/progpy. Last accessed 11 Nov 2024
8. Chindriş, G., Ilieş, A.I., Pitică, D.: A comparison between state of charge estimation methods: extended Kalman filter and unscented Kalman filter. In: 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME), pp. 376–381 (2020)
9. Jung, R., Weiss, S.: Modular multi-sensor fusion: a collaborative state estimation perspective. IEEE Robot. Autom. Lett. **6**(4), 6891–6898 (2021)
10. Chen, S., Liu, Z., He, X., Zou, D., Zhou, D.: Multi-mode fault diagnosis datasets of gearbox under variable working conditions. Data in Brief, no. 54 (2024)

11. SKF: Vibration Diagnostic Guide (2019). https://skftechnicalsupport.zendesk.com/hc/en-us/article_attachments/360042513054. Last accessed 15 Nov 2024
12. Tiboni, M., Carlo, R., Bussola, R., Amici, C.: A review on vibration-based condition monitoring of rotating machinery. Appl. Sci. **12**(3), 972 (2022)

Development of Human Cognition Inspired Digital Twins for Optimized Performance of Process Equipment



Maneesh Singh, Venkataramana Runkana, Knut Øvsthus, Arvind Keprate, and Anne-Lena Kampen

Abstract Advances in the domains of electronics, communication technologies, artificial intelligence, and computing hardware have led to the development of *Cyber-physical Systems (CPS)*; where physical structures, such as, machines, process equipment, buildings, offshore installations, integrate seamlessly with cyber systems to enhance operational efficiency. Data, from numerous sensors deployed to monitor various attributes of the equipment, operation and environmental conditions, is collected and transmitted over the communication network to processors. This data is analyzed using various algorithms and the output is used to optimize the performance and health of processes and equipment. *Biomimicry* is an approach for solving modern day technological challenges by taking inspirations from the solutions that have been successfully developed and tested by bio-organisms over the course of millions of years. This paper presents a proposed human cognition inspired framework for developing digital twins of processes or equipment. The framework consists of three components (a) *Primary Component*—inspired by the "old brain" in mammals for *Procedural Decision Making* used for taking initial decisions that are needed for safe operations (corresponding to Basic Needs in an organisms); (b) *Secondary Component*—inspired by the "new brain" in mammals for *Deliberative Decision Making* used for optimizing operations (corresponding to Functional Needs in organisms); and (c) *Arbitrator Component*—used for resolving decisions based on conflicting needs. Additionally, the paper discusses a bio-inspired digital twin framework specifically designed for iron ore sintering.

M. Singh (✉) · K. Øvsthus · A.-L. Kampen
Western Norway University of Applied Sciences (HVL), Bergen, Norway
e-mail: maneesh.singh@hvl.no

V. Runkana
TCS Research, Tata Research Development and Design Centre (TRDDC), Tata Consultancy Services, Pune, India
e-mail: venkat.runkana@tcs.com

A. Keprate
OsloMet—Oslo Metropolitan University, Oslo, Norway
e-mail: arvindke@oslomet.no

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_42

Keywords Artificial intelligence · Biomimicry · Cyber-physical systems · Digital twin · Process optimization

1 Introduction

Advances in Information and Communication Technology (ICT) have allowed data to be acquired, transmitted, stored, organized, and converted into useful information at an unprecedented level. This has introduced new possibilities for integrating physical processes with digital systems (sensors, controllers, electronics, computing hardware, etc.) to form *Cyber-Physical Systems (CPS)*. This term, coined by Helen Gill at the National Science Foundation (NSF), describes complex systems that advances beyond the traditionally integrated equipment, sensors, and control systems [10]. The research on CPS has primarily been in the domains of fundamental studies around concept, architecture, and associated technologies. Recently, the growth of these domains has been assisted by the developments in the fields of Industry 4.0, Internet-of-Things, big data, multi-physics simulation, non-invasive sensors and sensor networks, data management, data processing, and a drive towards a data-driven and digital manufacturing future [21, 15].

As a result of the rapid growth in the development of embedded systems, modern installations come with a multitude of sensors that constantly acquire data for further processing. While collection, transmission, and storage of this data have become relatively easy, the challenge remains in the development of digital representations that can identify and extract relevant information from the available data [28], and then use the information to synchronize the physical system and its virtual counterpart. As a result, no generic virtual system can take a holistic approach to processing available data, taking decisions and optimally controlling the physical system.

Over time, the development and expansion of various digital components with varying capabilities have led to a situation similar to the story of the blind men and the elephant. These digital representations have been given different names based on personal understanding and commercial interests. One way of categorizing digital representations is based on the capability of its *Digital Thread*. Depending on the type of *Digital Thread* between the physical system and its virtual representation, a digital representation can be categorized into *Digital Model*, *Digital Generator*, *Digital Shadow* and *Digital Twin* [17, 19].

One of the very promising methods for synchronizing the physical system and its virtual counterpart is the *Digital Twin* [11]. Since then, the idea has been used extensively both in academia and industry. Since its inception, the definition of *Digital Twin* has evolved based on its functionalities, hierarchy, input technologies, modelling approaches capabilities, applications and purposes [1]. For the purpose of this paper, a *Digital Twin* should have an autonomous bi-directional data exchange between the physical system and the virtual system. A *Digital Twin* utilizes data from sensors to depict the present condition and forecast the future development of its physical equivalent. By analyzing the present condition and anticipating future outcomes,

it becomes possible to command actuators to regulate the behavior of the tangible property [24, 25, 6].

The application of digital twins in process industries can assist in real-time process monitoring, predictive maintenance, leakage detection, dynamic process optimization, hazard identification and operator training [18]. Thus, in the context of the applications of digital twin in process industry, it can be defined as "*as a virtual representation of a physical asset enabled through data and simulators for real-time prediction, optimization, monitoring, controlling, and improved decision making*" [25, 23].

Over millions of years, bio-organisms have evolved to operate and survive under different environmental conditions. Emulating bio-organisms' structure and behavior may help to develop more efficient processes and equipment.

The practice of taking cues from biological organisms and mimicking their strategies to enhance the quality of products or processes is called *Biomimicry*. According to the Biomimicry Institute, *the distinctive feature of biomimicry is the study and emulation of functional strategies to create sustainable solutions that also embody the (re)connect and ethos elements.*

A study of bio-organisms shows distinct similarities between the workings of bio-organisms and cyber-physical systems. Both have: physical body (human body versus equipment/structure), data collecting sensors (sensory organs versus sensors), data transfer mechanism (nervous system versus digital data transportation and storage), and components for memory and analysis (brain versus data analytics). Thus, it should be possible to learn from humans how their body reacts to external stimuli and apply those concepts for developing self-optimizing and self-preserving assets (equipment or process).

Application of biomimicry for solving technical challenges is relatively new. Some work has been done in mimicking design of physical body ("painless" needles, "sharkskin" swimming suit, etc.), sensors (tactile array of sensors inspired by whiskers of animals, ultra-sensitive vibration detector inspired by spiders, etc.) and data communication networks (AI enabled telecommunication networks, etc.) [Biomimicry Institute]. Relatively less attention has been paid to the use of biomimicry as an active component in decision-making. By mimicking the cognitive process by which bio-organisms regulate their processes in response to the changing environment, it may be possible to develop a cognitive system for a digital twin. Such a system may enable an equipment to provide recommendations for optimizing its performance or ensuring good health.

This paper presents the ideas behind an ongoing research work being carried out to develop digital twin of process equipment inspired by the human cognitive system. Appreciating the vastness and complexity of human cognitive system that covers various aspects of biological, neurological, sociological, and functional human behavior, we are aware that several parallel and conflicting theories have been proposed to explain different subjects. In our approach, we have often taken simplified versions of these theories and at times adopted a mix and match approach while using them. Thus, this work takes inspiration from the human cognition system

but in no way claims to replicate it. This work is still under development and many important features are still missing.

2 Cognition in Humans and Equipment

All living organisms adjust their behavior to suit their environment, similarly, all industrial equipment and processes must adapt to suit the changing environmental and processes. An understanding of human cognition system may help develop a decision making system for the digital twin of the equipment or the process.

2.1 Inspirations from "Old Brain" for Procedural Decision Making

A brain consists primarily of two parts—"old brain" and "new brain". All animals have the old brain that is made up of many structurally different components or organs that are hard-wired to perform specific tasks like premeditation and impulsive aggression (amygdala), basic movements (spinal cord), and digestion and breathing (brain stem). Together they carry out basic functions like eating and sleeping, and behavioral functions like reflexes, feelings, emotions, desires, etc. These basic skills are coded in the genes and most of them occur without conscience decisions [12].

Some of the hard-wired decisions and skills are activated by signals from sensory organs. A sensory organ sends the signal first to the spinal cord which, if required, takes reflexive decisions. For example, an accidental touching of a hot surface can result in a reflex action based on processing at the spinal cord. The decisions taken for reflex are mostly regulated by what is known as *Procedural Decision Making* mechanism of the brain [30].

Taking inspiration from the old brain, digital twins should have decision-making modules for safe equipment operation. Addressing safety issues that require immediate attention, such as sudden failures, is crucial. These situations are either too critical to wait for detailed data collection or difficult to predict accurately with time-dependent models. This component should take decisions based on approaches that require minimum data and rely on past experiences, like rule-based reasoning (expert systems, decision trees, fuzzy logic, etc.) or case-based reasoning, to generate robust and safe decisions.

2.2 *Inspirations from "New Brain" for Deliberative Decision Making*

Mammals have an additional component, often referred as the "new brain" (neocortex or "new outer layer"). Unlike the old brain, neocortex is composed of one large structure that looks similar throughout, yet different regions of the neocortex perform different functions, from interpretation of signals from the sensory organs to understanding the environment (forms, shapes, etc.) to abstract thinking (language, mathematics, philosophy, etc.). This suggests the possibility that all the regions work on the same general algorithm, giving every region the same ability to learn and perform. The differences in the results of different regions arise due to the differences in the applications and connections to the sensory organs [12].

The signal from a sensory organ goes first through the spinal cord and then via brain stem and thalamus to the neocortex for processing. In the example of accidental touching of a hot surface, according to the sequential model, after the reflexive decision has been taken by the old brain, the signal reaches the neocortex (somatosensory cortex) for detailed processing (sensory perception, information processing, decision making and learning). Similarly, in reading of a braille code, there is no reflexive action, and the neocortex interprets the writing based on the *Deliberative Decision Making* mechanism of the brain [30].

Neocortex with its knowledge and understanding of the world, while actively involved, can only facilitate the old brain in pursuit of its "survival goals and motivation". Thus, neocortex is an enabler for motivations that arise in the old brain. Additionally, since neocortex is not directly connected to the sensors and cannot directly control the movement, it has to rely on the old brain that controls the sensors and behavior [12].

Taking inspiration from the new brain, a digital twin should have a component that contains generic tools and has capabilities for data analytics for optimized operation and failure prognostics associated with time-based failure. This component should contain deterministic, statistical or machine learning/artificial intelligence models for performing specialized tasks like risk analysis, fitness-for-service analysis, etc. This component would require more data but would also produce more reliable and optimized decisions.

Separation of functions of the old and new brains creates a natural built-in security barrier. For instance, a person can consciously choose to hold their breath through the new brain's deliberative decision-making, but if survival is at risk, the old brain's procedural decision-making takes over to ensure safety. Applying this concept of functional separation introduces an effective security feature for digital twins.

2.3 Arbitration of Decisions Suggested by "Old Brain" and "New Brain"

The old brain and the new brain are not disjoint organs; they operate in a coordinated manner. Yet, at times, the independence of two separate decision making components results in disparate decisions when:

Case 1 old brain has taken a quick decision, while the new brain awaits for more data
Case 2 old brain and new brain have reached conflicting decisions
Case 3 the sense organs produce conflicting inputs
Case 4 new brain sets a goal and the old brain ensures proper operation

Case 1 arises due to delayed decision making by the neocortex, and delay may be caused by the time it takes for the brain to collect data and produce a clear situation awareness. The new brain takes additional time to interpret the situation thereby causing a time gap between the reflex and the conscious response. Kahneman [14] discusses the relations between the two systems, which he refers to as System One and System Two. Sometimes it is wise to delay the response/conclusion and await some reasoning before responding. On the other hand, waiting can cause "The Centipede's Dilemma" which occurs when a person starts to reflect on an automated process, like the centipede who could not dance when she started reflecting on how to move her feet. The main takeaway is that slower consciousness should not obstruct instantaneous responses (understanding fast reflexes). The interplay between the two systems, the fast and the slow, must be carefully tuned [2, 16].

Similarly, the time available to realign a process based on the external changes could be long or short. Within the available period of time, the adjustments can either be fixed (*Procedural Decision Making*) or built on the available options (*Deliberative Decision Making*). This decision making requires evaluating two or more alternatives having different attributes within a certain time duration. When making decisions, it is important to consider how delaying that decision could impact its outcome. Thus, the time available to take a decision is constrained by (a) the duration of availability of each option, accounting for each option's rate of change (example, time available before pressure drops below an acceptable level), (b) time required to implement the decision (example, time required to open a valve after issue of command), and (c) time required to achieve the consequence of the decision (example, time required to raise pressure back to the acceptable level). Each of these is decided by design issues and choices made after accounting for the feedback loop time delays, processing time, etc.

Case 2 arises when the decisions made by the two may come in conflict. To resolve the differences during life-threatening situations, the old brain has the power to override the neocortex's commands. This is because various functional areas of the neocortex connect to distinct regions of the old brain, which is responsible for regulating muscle movement. Since the neocortex doesn't have a direct connection

to the muscles, it can't control movement directly and manages movement through the old brain [12].

Similarly, in the proposed framework of the digital twin, the two components should take independent decisions based on different goals and methodologies. An arbitration component is required to resolve the differences caused by the two decision processes. It selects appropriate decisions based on the two components' recommendations. For example, just as the old brain over-rides the decisions taken by the new brain in case of safety issues, so too can the arbitration component of the digital twin favor safety over optimization.

For any structural, environmental, or operational change that takes place, an organism gets a number of independent inputs in terms of location (where?), form (what?) and time (when?). The multiple data streams and corresponding interpretation have separate associated neocortex columns, yet the organism perceives it as a single event, because of the "voting" among different neocortex columns involved in the process. Once consensus has been reached among the neocortex columns, the brain can unite different sensory inputs into a single representation of the event being sensed. In case a sensor system is unable to send a single opinion, it sends multiple possibilities that are accounted for during voting [12]. A digital twin can also apply the same principle of voting to resolve ambiguity of results among different sensors by concepts of "sensor fusion" and multi-criteria decision making.

Case 3 arises when the decisions based on the interpretations of the sensor inputs are not unanimous. In this case there is a need to resolve the differences and give higher weightage to those sensors that are more reliable or take conservative approach.

Case 4 arises when the neocortex sets goals like participating in a skyscraper stair race. Bipedal running is a complex movement involving coordination between muscle movements, vision, oxygen supply, etc. While neocortex, sets the goals and motivations, the old brain activates and controls these activities through its components (amygdala, spinal cord, brain stem, etc.).

Digital twin of an autonomous factory containing CNC machines can take inspiration from this division of responsibility between goal setting and operation. Such a factory can implement a "new brain" system for scheduling orders without human interventions. To ensure cost-effective and reliable production, the orders may be selected based on several criteria, such as deadline, the conditions of the machines and the associated tools, stock of raw materials, time until the next service/refill, etc. Once a decision has been reached, the execution can be automated mimicking the functioning of the old brain. Hence, the operation can be properly planned and executed in steps, involving locating the raw material, transporting it to the saw, loading the sawed unit into the CNC machine with precision and predictability, fetching the finished unit, and preparing it for delivery. Thus, as in the staircase runner, in the autonomous factory the new brain will set the goals and the old brain shall execute them in a safe and reliable manner.

3 Proposed Framework of a Bio-Inspired Digital Twin

Figure 1 shows the proposed framework for a digital twin that has been inspired by the human cognition system. The proposed digital twin has three distinct components:

- **Primary Component (corresponding to the "old brain")**. The *Primary Component* contains the regions that are required specifically for the purposes that are unique for the operation (corresponding to *Functional Needs*) and maintenance (corresponding to *Basic Needs*) of process equipment. Analogous to the working of the old brain in animals, the *Primary Component* carries out *Procedural Decision Making*. The results of *Procedural Decision Making* are used for taking safety critical or time constrained decisions. The *Primary Component* consists of five regions dedicated to: (a) *Goals and Motivations*, (b) *Safety and Security*, (c) *Memories*, (d) *Awareness and Attention*, (e) *Procedural Decision Making*.

- **Secondary Component (corresponding to the "new brain")**. The *Secondary Component* receives preprocessed data from the *Primary Component* and carries out analytics for the digital twin. Here advanced analytics are carried out to enable *Deliberative Decision Making*. The results of *Deliberative Decision Making* are used for optimizing the process or health of the equipment. The *Secondary Component* consists of four regions that are dedicated to: (a) *Memories*; (b) *Failure Profile Analysis*; (c) *Risk Assessment*; and (d) *Deliberative Decision Making*.

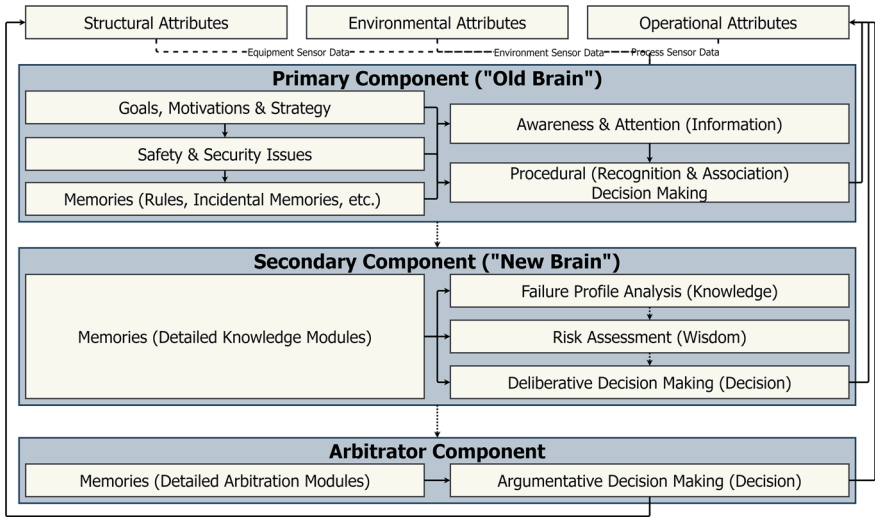


Fig. 1 Proposed architecture for a biologically inspired digital twin

- **Arbitrator Component.** The decisions taken by the *Primary Component* and the *Secondary Component*, have to account for a number of potentially conflicting requirements. If necessary, the *Arbitrator Component* resolves conflicting decisions. This component receives decisions from the *Primary Component* and *Secondary Component* and if the differences between the decisions exceed some pre-set thresholds, then arbitration among the decisions are carried out using the *Arbitration Methodology*.

4 Structural—Environmental—Operational Attributes

Every physical object, biological or process equipment, has a unique structure that is designed to help operate in its living/working environment. In case of an equipment, this structure can be characterized based on the physical dimensions, materials of construction, protections (example, coatings and linings), insulations, lubrications, etc. The external and/or internal environment(s) in which the object operates can be characterized by parameters like, layout, temperature, pressure, humidity, fluid type, fluid flow rate, etc. The interplay between the structure (biological or process), operation and environmental characteristics determines not only the capabilities and performance of an equipment, but also its degradation and subsequent failure. An equipment's components are subjected to a number of degrading mechanisms, like load, wear, corrosion, and erosion during its operation. These attacks may damage and reduce the structural integrity of the object. For example, for an offshore wind turbine, the environment attribute (wind speed) affects the operational attribute (rotational rate or RPM) of the gearbox, resulting in changes to the structural attributes (temperature and wear rate) of bearings.

The structure (biological or process) also houses the organs that help it sense the condition of the structure, operation and environment in which it exists. These senses may be simple, like feeling just the presence of light, to complex, like vision (IR, UV or visible light). Corresponding to the sense organs in animals, equipment use different types of sensors equipment for monitoring corresponding parameters. Some sensors may have equivalence in the animal world, for example, visual or IR cameras (vision), acoustic or ultrasonic sensors (hearing), thermocouple (temperature), accelerometer (touch), gas sensors or digital nose (smell) and chromatographs (taste), while others may monitor specific parameters that have no equivalence in animal world.

The sense organs convert changes in physical parameters or energy potential to electro-chemical signals and send them to the brain via sensory nerve fibers. Correspondingly, sensors in an equipment convert the changes in physical parameters or energy potential to electromagnetic signals and transmit them to the processor (computer) via cables. Thus the brain or the processing unit receives real-time data about the structure, operation and environment, which it can use to regulate and optimize its behavior.

5 "Old Brain" Inspired Procedural Decision Making

5.1 Needs, Goals and Motivations

In an organism, the old brain dictates the creation of desires and emotions in pursuit of the needs, goals and motivations. It can also enable the animal to lead a complex life, like navigation, searching for food, mating, caring for young; but is not able to help it understand more complex intellectual pursuits, like understanding physical forms and abstract thoughts [12].

Needs, goals and motivations drive an organism's activities. According to cognition theories, the contents of these exist as knowledge (cognition) representations in memories.

To meet the goals, the process requires optimization based on the cause-and-effect relationship between major process variables during its entire life cycle. For example, in a digital twin of a pump, the goal could be the desired level of output and the motivation is the result of the cost–benefit analysis. This optimization is achieved by the motivations dictated by the cost–benefit analysis [22, 27].

5.2 Awareness and Attention

Awareness and attention can be seen as cognitive filters that modulate the processing of sensory input. Awareness refers to our subjective experience of being conscious of something, attention refers to the selective allocation of cognitive resources to specific stimuli or information. Changes in stimulation can trigger orienting responses to direct attention towards the change in the environment. So, attention can be considered as control valve of data transformation in-between old brain and new brain. If information gets attention, then it moves into a new brain for further analysis (active thinking/learning) [7].

5.3 Procedural Decision Making

The expected process equipment performance depends on its structure, operating environment, and operational parameters, most of which are unique and intimately associated with the purpose of the equipment. Just as the uniqueness of a bio-organism is stored in its genes and reflected on the memories in the old brain; the *Primary Component* also stores features that are unique to the process equipment. Hence, the *Primary Component* contains the necessary information and models required to carry out its fundamental tasks. For example, it holds details about:

- Structure in terms of spatial designs (3D CAD, 3D point cloud, 2D CAD, etc.), materials information, safety features, location and characteristics of sensors, etc.
- Identification of: threats (failure causes) from operating conditions, misuse, mishandling, etc.; attacks (degradation mechanisms) by corrosion, erosion, fatigue, etc.; and resulting damages (degradation), like cracks, wall thinning, deformation, etc.
- Cache of "best practices" regarding process control, diagnostics and prognostics (aging, wear, sudden failure), protection (like, process operation adjustment) and preservation (maintenance strategy).

The signals that come from the sensors monitoring structural, environmental and operation parameters are first preprocessed (noise reduction, handling of the missing or spurious values, etc.) in the *Awareness and Attention Region* of the *Primary Component*. The preprocessed measured values are matched against the threshold values or predicted values to detect any anomaly. If an anomaly is detected then *Procedural Decision Making* region is activated.

When an animal encounters an incidence or accident, it takes decision and corresponding action within short reaction time, and these actions are often characterized as "reflexive actions". The short decision time is achieved by a combination of recognition processes and their associated decisions [22]. Similarly, *Procedural Decision Making* region uses schemas present in the *Memories* region to identify predefined failure modes and type of damage. This knowledge is then used for estimating safe working conditions of the damaged equipment.

6 "New Brain" Inspired Deliberative Decision Making

6.1 Memories

At birth a mammal's neocortex does not have any knowledge about its host's living environment. As time passes, sensory data is received through the old brain. However, instead of storing this information in the neocortex as a library of facts, the neocortex uses it to develop predictive models. Mammals are not born with pre-existing models in their neocortex, but possess the ability to create them. As an animal explores its ever-changing environment, it continuously develops new models and refines existing ones. In this way, the brain builds and recalls hundreds of thousands of models to "sense" and interpret the real world.

Just as the neocortex has several models to represent the environment, the Secondary Component of a digital twin has a number of models to represent the process equipment. These models analyze the sensor data, interpret the real-time process equipment's condition, predict its behavior and control it. This feed-forward approach enables creation of a proactive and not just a reactive digital twin.

Depending on the requirements of the process equipment, various types of models are used, such as computational fluid dynamics (CFD) models to predict fluid flow,

finite element analysis (FEA) for structural analysis, and degradation models for assessing integrity loss. These models should work together, enabling equipment conditions to be interpreted through multiple sources of data and models, rather than relying on a single data source. This integration of heterogeneous sensors and complementary models provides a more comprehensive understanding of equipment health.

6.2 Deliberative Decision Making

Deliberative decision-making is a slow and computationally intensive process. It requires sufficient data and computational power. Hence, *Deliberative Decision Making* Region takes only the critical decisions and provides detailed output regarding failure prognostics, optimized operation control and recommends inspection-maintenance schedule.

Humans carry out deliberation by means of "episodic future thinking" that takes place in two steps: (a) imagination to create a representation of a future scenario; and (b) evaluation of that scenario. The consequences of each possible future scenario are then compared and the best option is selected [22].

Similarly, for a particular goal, *Deliberative Decision Making* Region considers various possible scenarios and evaluates corresponding risks and outcomes in terms of safety, cost–benefit, etc. These results are then compared and the best option is used to optimize and control the process according to the current condition.

7 Argumentative Decision Making

The old and new brains are not separate organs; they function together as a coordinated team. At times, they may reach conflicting decisions that need resolution. For example, there may be a conflict between the new brain's goal to hold one's breath and the old brain's instinct to supply oxygen to the body. In such cases, the old brain's decision often takes precedence over the new brain's choice [12].

An *Arbitrator Component* resolves decisions based on conflicting goals in the proposed framework. This component receives decisions from the Primary Component and Secondary Component. At times, because of the optimization among multiple goals, the Secondary Component may generate multiple decisions. If the differences between the decisions exceed some pre-set thresholds, then arbitration among the decisions is carried out using Arbitration Methodology [9].

8 Bio-Inspired Digital Twin of a Sinter Plant

Using the framework outlined in the previous section, a bio-inspired digital twin system can be designed for a sinter plant within an integrated steel plant. This system would leverage biological cognition principles discussed previously to create a responsive, adaptive model that continuously learns and refines its understanding of plant operations. The digital twin can then help in enhancing real-time monitoring, predictive maintenance, and process optimization (Fig. 2).

The sinter plant processes iron ore fines into a porous sinter cake by utilizing fluxes like limestone and dolomite, along with fuel such as coke breeze. Key unit operations include raw material blending, granulation in a rotary drum with water and calcined lime as binders, sintering of granules on a moving strand, crushing of the sinter cake, and screening of the final sinter product [20]. In this setup, the digital twin's Primary Component (Old Brain) encompasses data processing algorithms, control laws, a rule engine, knowledge bases, data-driven models, and safety locks to maintain plant stability and prevent equipment from entering unsafe conditions. The Secondary Component (New Brain) conducts computationally intensive simulations, such as multiphase computational fluid dynamics, discrete element modeling for particle flows, finite element stress analysis, and an integrated plant model. The Arbitrator Component incorporates process optimization models, equipment maintenance and scheduling algorithms, risk analysis, and multi-criteria decision-making to evaluate inputs from the Primary and Secondary Components, ultimately determining optimal plant operations.

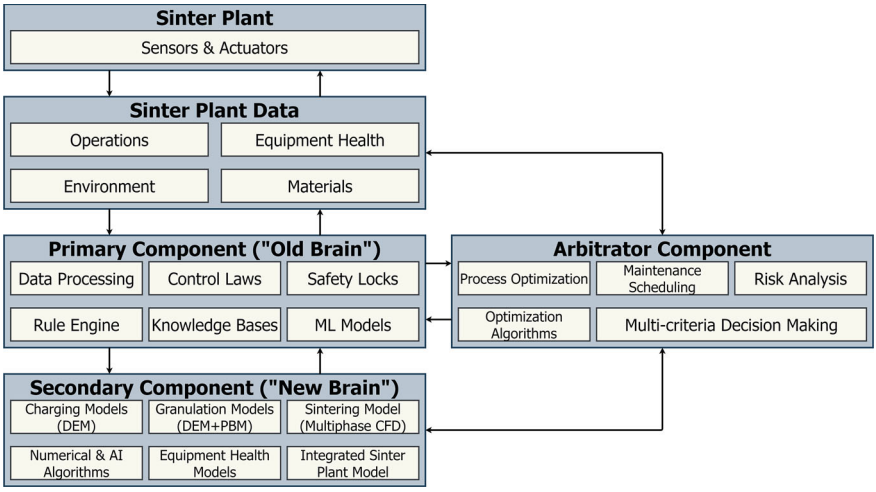


Fig. 2 Biologically inspired digital twin of a sinter plant

9 Conclusions

Applying biomimicry to solve technical challenges is an emerging field, with successful applications in developing innovative technologies for objects such as the Shinkansen Bullet Train and the Eastgate Building in Harare, Zimbabwe. However, to our knowledge, biomimicry has not yet been explored for creating digital twins of equipment or processes. Here, a human cognition-inspired framework for developing a digital twin, designed to leverage real-time data to recommend optimal decisions for the operation and maintenance of equipment or processes has been proposed.

It is expected that the framework can help to:

- Clearly define distinct goals for operation and maintenance tasks to support decision-making amidst multiple, often conflicting, requirements.
- Systematically manage diverse data types (numerical, pseudo-numerical, linguistic, etc.), along with relevant information and knowledge, to support effective decision-making.
- Ensure access to a range of mathematical and statistical tools as needed.
- Facilitate structured decision-making by integrating:
  - **Procedural Decision-Making**: recommends decisions based on limited data or information, prioritizing safety and stability of the process.
  - **Deliberative Decision-Making**: recommends decisions based on detailed analysis, focusing on process optimization.
  - **Argumentative Decision-Making**: considers multiple or conflicting goals, offering balanced recommendations accordingly.

It is acknowledged that the bio-inspired digital twin is still at a conceptual stage and this work is a step in that direction. Effort was made to draw the parallels between the human brain and the cognitive digital twins and how it is possible to learn from the biological systems to design the digital twin systems for process industries. Implementation of the framework proposed here, and its application will be described in a future work.

The proposed bio-inspired digital twin remains at a conceptual stage. This work has aimed to draw parallels between the human brain and cognitive digital twins, demonstrating insights from biological systems can guide the design of digital twin systems for process industries. The implementation of this framework and its practical applications will be discussed in future work.

References

1. Abbas, M., Liu, Y., Cai, B.: Digital twin-based prognostics and health management for subsea systems: concepts, classification, opportunities and challenge. In: Castanier, B. et al. (ed) Proceedings of the 31st European Safety and Reliability Conference (ESREL 2021), Research Publishing, Singapore (2021). ISBN: 978-981-18-2016-8; https://doi.org/10.3850/978-981-18-2016-8070-cd

2. Ariely, D., Zakay, D.: A timely account of the role of duration in decision making. Acta Psychol. **108**(2), 187–207 (2001)
3. Atkinson, R.C., Shiffrin, R.M.: The control of short-term memory. Scient. Amer. **225**(2), 82–91 (1971)
4. Bartnes, M.: "Safety vs. Security?". In: Proceedings of the 8th International Conference on Probabilistic Safety Assessment and Management, May 14–18, 2006, New Orleans, Louisiana, USA (2006)
5. Biomimicry Institute. https://biomimicry.org/, Accessed 10 Feb 2023
6. Botín-Sanabria, D.M., Mihaita, A.-S., Peimbert-García, R.E., Ramírez-Moreno, M.A., Ramírez-Mendoza, R.A., Lozoya-Santos, J.d.J.: Remote Sens. **14**, 1335 (2022)
7. Bridgman, T., Cummings, S., Ballard, J.: Who built Maslow's Pyramid? A history of the creation of management studies' most famous symbol and its implications for management education. Acad. Managem. Learn. Educ. **18**(1), 81–98 (2019)
8. Cowan, N.: In: Attention and Memory: An Integrated Framework. Oxford University Press 1998
9. Fishbach, A., Ferguson, M.J.: The goal construct in social psychology. In: Social Psychology: Handbook of Basic Principles, pp. 490–515. (2007)
10. Fridman, L., Ding, L., Jenik, B., Reimer, B.: Arguing machines: human supervision of black box ai systems that make life-critical decisions (2018). Available at: https://arxiv.org/, Accessed: 27 Feb 2023
11. Gill, H.: NSF perspective and status on cyber-physical systems. Austin (2006). Internet: http://varma.ece.cmu.edu/CPS/Presentations/gill.pdf [zuletzt aufgesucht am 1.04. 2015]
12. Grieves, M.: Digital twin: manufacturing excellence through virtual factory replication. White paper, 2014—Florida Institute of Technology (2014)
13. Hawkins, J.: A Thousand Brains—A New Theory of Intelligence. Basic Books, New York (2021)
14. Health and Safety Executive.: "Key Programme 3—Asset Integrity Programme". In: A report by the Offshore Division of HSE's Hazardous Installations Directorate HSE (2007). www.hse.gov.uk/offshore/kp3.pdf
15. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B.: Characterising the digital twin: a systematic literature review. CIRP J. Manuf. Sci. Technol. **29**, 36–52 (2020)
16. Kahneman, D.: In: Thinking, Fast and Slow, Farrar, Straus and Giroux (2011). ISBN 978-0374275631
17. Klapproth, F.: Time and decision making in humans. Cognit. Affect. Behav. Neurosci. **8**(4), 509–524 (2008)
18. Lee, J., Cameron, I., Hassall, M.: Improving process safety: what roles for digitalization and Industry 4.0? Process Saf. Environ. Prot. **132**, 325–339 (2019)
19. Newrzella, S.R., Franklin D.W., Haider, S.: 5-dimension cross-industry digital twin applications model and analysis of digital twin classification terms and models. IEEE Access **9**, 131306–131321 (2021). https://doi.org/10.1109/ACCESS.2021.3115055
20. Newrzella, S.R., Franklin D.W., Haider, S.: Three-dimension digital twin reference architecture model for functionality, dependability, and life cycle development across industries. IEEE Access **10**, 95390–95410 (2022). https://doi.org/10.1109/ACCESS.2022.3202941
21. Nistala, S., Singh, K., Vakkantham, P., Pareek, A., Masampally, V., Kumar, R., Runkana, V.: Virtual Sinter®: digital twin for integrated sinter plants. Trans. Indian National Acad. Eng. **7**, 943–955 (2022)
22. Onosato, M., Iwata, K.: Development of a virtual manufacturing system by integrating product models and factory models. CIRP Annals **42.1**, 475–478 (1993)
23. Rasheed, A., San, O., Kvamsdal, T.: Digital twin: values, challenges and enablers from a modeling perspective. IEEE Access **8** (2020)
24. Redish, A.D., Schultheiss, N.W., Carter, E.C.: The computational complexity of valuation and motivational forces in decision-making processes. Current Topics in Behav. Neurosci. **27**, 313–333 (2016)

25. Runkana, V.: Model-based optimization of industrial gas-solid reactors. KONA Powder Particle J. **32**, 115–130 (2015)
26. Runkana, V.: "Digital Twins". In: Reimagining Research, pp. 97–103. Tata Consultancy Services Ltd. (2018)
27. Schacter, D.L.: Implicit memory: history and current status. J. Experim. Psychol.: Learn. Memory Cognit. **13**(3), 501 (1987)
28. Simpson, E.H., Balsam, P.D.: The behavioral neuroscience of motivation: an overview of concepts, measures, and translational applications. Current Topics in Behav. Neurosci. **27**, 1–12 (2016)
29. Singh, M., Hetlevik, S.: Data–information–knowledge hierarchy based decision support system for risk based inspection analysis. Int. J. Syst. Assurance Eng. Managem. **8**(Suppl. 2), S1588–S1595 (2017)
30. Subramanian, S., Runkana, V., Parameswaran, S., Shah, N., Sandipan, M., Mehrotra, A.N., Padsalgi, M.S., Manna, R., Kumar, R., Nistala, S., Pandya, R., Pareek, A., Baikadi, A.K., Deodhar, A.: System and method for development and deployment of self-organizing cyber-physical systems for manufacturing industries, Patent Application No. IN202021021089 (2020)
31. Vancouver, J.B.: Goal constructs in psychology: structure, process, and content. Psychol. Bullet. **120**(3), 338–375 (1996)

Application of a Large Language Model (LLM) for Failure Modes and Effects Analysis (FMEA)



Kamal V. Charan, Surender Redhu, Gunjan Soni, and Maneesh Singh

Abstract During its operational lifespan, equipment used in chemical, mechanical, and metallurgical process plants is subjected to various degradation mechanisms, including corrosion, erosion, fatigue, and deformation. These factors can severely impact equipment integrity, increasing the likelihood of failure. While complete elimination of failures is impractical, an effective asset management program focuses on maximizing reliability and availability through well-structured inspection and maintenance strategies. However, developing and implementing such programs is complex and costly. Risk-based approaches, such as Risk-Based Inspection (RBI) and Reliability-Centered Maintenance (RCM), offer systematic methods to optimize resource allocation and minimize accident risks. Failure Modes and Effects Analysis (FMEA) is a widely used technique for risk assessment. Primarily qualitative, FMEA can incorporate quantitative elements by estimating failure likelihood and consequences. While benefiting from human expertise, traditional FMEA is often criticized for being time-intensive, prone to human error, and inherently static. Recent advancements in Generative AI have opened avenues for automating the FMEA process. A current project is developing a framework that combines open-source Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG) to automate the creation of FMEA/FMECA tables using only pre-vetted, credible documents. At the proof-of-concept (PoC) stage, the project has shown promising results in automating failure analysis. This paper introduces the proposed framework and presents initial findings.

Keywords Failure modes and effects analysis (FMEA) · Large language models (LLMs) · Retrieval augmented generation (RAG)

K. V. Charan · G. Soni
Malaviya National Institute of Technology, Jaipur, India
e-mail: gsoni.mech@mnit.ac.in

S. Redhu
Kongsberg Digital, Oslo, Norway

M. Singh (✉)
Western Norway University of Applied Sciences, Bergen, Norway
e-mail: maneesh.singh@hvl.no

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_43

1 Introduction

During its operational lifetime, equipment that forms part of a chemical, mechanical, or metallurgical process plant (e.g., an offshore oil and gas installation) or a structure (e.g., a bridge) is subjected to various degradation mechanisms, such as corrosion, erosion, fatigue, and deformation. These mechanisms can significantly compromise the integrity of the equipment, increasing the risk of failure. While it is impossible to eliminate all failures or their causes, a well-structured asset management program focuses on maximizing reliability and availability by selecting the most appropriate inspection and maintenance strategies.

An effective technical asset integrity management program seeks to optimize resource availability at minimal cost while ensuring compliance with safety and regulatory standards. This is achieved by balancing costs, opportunities, and risks to attain the desired asset performance [9]. Furthermore, this approach ensures that the asset management system meets its objectives, identifies potential improvements, minimizes risks, and fosters continuous enhancement [10].

Currently, inspection and maintenance planning typically follows one of two approaches: the *Traditional Approach* or the *Condition-Based Approach*. In the traditional approach, the failure profile of equipment—including failure causes, mechanisms, modes, and rates—is analyzed using various data sources, such as design specifications, process data, and historical inspection and maintenance records. This understanding informs the development of a maintenance strategy and, ultimately, a maintenance plan. Two widely used methodologies for this purpose are:

- *Reliability-Centered Maintenance (RCM)*: Originally developed in the aviation industry, RCM is used to analyze complex equipment, such as pumps, valves, and compressors.
- *Risk-Based Inspection and Maintenance (RBIM)*: Developed in the oil and gas industry, RBIM is primarily applied to static equipment, such as pipelines and storage tanks.

Both RCM and RBIM rely on risk (criticality) management to formulate inspection and maintenance plans. While the word "risk" and associated terms have been defined in many ways across standards and recommended practices, this paper adopts the definition from NORSOK Z-013 [15] where risk is defined as *a combination of the probability of occurrence of harm and the severity of that harm*. In this context, "probability" is interchangeable with "frequency", and "consequence" corresponds to "severity" [8, 11].

A clear understanding of risk is fundamental to "risk management", which is defined as *systematic application of management policies, procedures and practices to the tasks of identifying, analyzing, assessing, treating and monitoring risk* [1]. Essentially, risk management serves as the overarching framework—encompassing philosophy, principles, procedures, and structures—used to identify, assess, and control accident risks in processes, installations, or structures.

By applying risk management principles, operators can strategically allocate resources among assets based on their perceived risk of failure. This approach enables the clear classification of assets into "low-risk" and "high-risk" categories. As a result, resources can be concentrated on high-risk assets to enhance overall system safety, while unnecessary expenditures on low-risk assets can be minimized, leading to cost efficiency.

Industries such as oil and gas and chemical processing have successfully implemented these methodologies, resulting in more effective inspection and maintenance planning, reduced accident rates, and lower operational costs.

One widely used method for conducting risk analysis is *Failure Modes and Effects Analysis (FMEA)* and its extension, *Failure Modes, Effects and Criticality Analysis (FMECA)*. According to ISO 31010 [12], *Failure modes and effects analysis (FMEA) is a technique used to identify the ways in which components, systems or processes can fail to fulfil their design intent*. This analysis includes identifying potential failure modes, their underlying mechanisms, associated consequences, and possible mitigation strategies.

A typical FMECA worksheet consists of a table with the following key components:

- Component (Part)
- Description and Function
- Failure Mode
- Failure Mechanism/Cause
- Consequence of Failure (Local and Global)
- Likelihood of Failure
- Likelihood of Detection
- Risk Priority Number (RPN) = (*Consequence of Failure* × *Likelihood of Failure* × *Likelihood of Detection*).

Traditionally, FMEA/FMECA tables have been prepared manually, hence, they have been criticized for being:

- *Time-Consuming:* Requiring extensive manual effort and analysis.
- *Prone to Human Error:* Increasing the risk of oversight or misjudgment.
- *Rigid and Static:* Lacking adaptability to changing asset conditions or operational demands.

In recent years, efforts have been made to automate aspects of this process using Artificial Intelligence (AI) techniques, enhancing efficiency and accuracy.

Fuzzy logic-based techniques have been applied to evaluate the Risk Priority Number (RPN) [4, 13]. These models utilize a set of "if–then" rules and logical operators to establish relationships between:

- *Input Variables:* Example, "Consequence of Failure", "Likelihood of Failure", "Likelihood of Detection".
- *Output Variable:* Example, "Risk Priority Number".

Both input and output variables are expressed as *linguistic variables*, which consist of multiple *linguistic terms* or *fuzzy variables* (example "High", "Medium", "Low", etc.). These linguistic variables are defined using *membership functions*, which quantify the degree of compatibility between a fuzzy variable and the input values.

The model relies on a *linguistic rule base* that encapsulates expert knowledge in the form of linguistic IF–THEN rules. Each rule typically consists of linguistic statements connected by the AND operator, followed by a conclusion. A standard linguistic statement follows this structure:

Linguistic variable – Symbol of Comparison – Linguistic term

This framework enables the model to process uncertain or imprecise data and generate an RPN as the output, improving risk assessment accuracy.

The Fuzzy Logic-based approach for calculating the Risk Priority Number (RPN) has been further enhanced through the integration of advanced methodologies, including:

- *Bayesian Network* [24],
- Convolutional Neural Network (CNN) [5]
- Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [14, 22].

These enhancements aim to improve the precision, robustness, and applicability of RPN evaluation across different industries.

Beyond fuzzy logic, *Machine Learning (ML)* techniques have also been applied to automate the FMECA process. ML models have proven particularly effective in identifying potential failure modes and calculating RPN values. Supervised learning algorithms, such as Decision Trees, Random Forests, and Support Vector Machines (SVMs), trained on historical failure data, have demonstrated strong capabilities in detecting failure patterns [5, 18].

Additionally, ML models trained on historical datasets can estimate critical parameters such as:

- Consequence of Failure
- Likelihood of Failure
- Likelihood of Detection

These estimated values are then used to compute the RPN, significantly streamlining the risk assessment process [16, 18].

Leveraging of Large Language Models (LLMs) for FMEA involves integration of advanced AI techniques with traditional risk assessment methodologies. Recent research demonstrates that combining LLMs with frameworks like Retrieval-Augmented Generation (RAG), multi-agent systems, and domain-specific knowledge bases enables automated extraction, analysis, and synthesis of failure data from diverse sources. These approaches significantly reduce manual effort and improve risk prediction accuracy across industrial sectors, from manufacturing and coal mine equipment maintenance to broader industrial applications. Fine-tuning LLMs with high-quality, domain-specific datasets has enhanced failure mode classification

performance, while multi-modal integrations and knowledge graph enhancements further empower these systems to generate context-aware, actionable recommendations. Collectively, these innovations are revolutionizing FMEA by streamlining processes, augmenting human expertise, and setting a new standard for efficient, data-driven risk management [3, 6, 17, 19, 20, 21, 23].

These innovations reflect the growing shift toward automation and intelligence in FMEA, enhancing its role in proactive risk management and system reliability.

2 Motivation and Aim of Research

2.1 Motivation

The preparation of FMEA/FMECA is a manual, time-intensive, and resource-heavy process due to the complexity of analysing multiple failure modes across various systems. Each analysis demands a detailed evaluation of multiple factors, including structural design, operational requirements, and environmental conditions.

Given these complexities, manual FMEA/FMECA preparation requires a substantial investment of time, personnel, and expertise. This makes the process costly, labour-intensive, and susceptible to human error, potentially leading to inconsistencies in risk assessments and decision-making.

Recent advancements in Large Language Models (LLMs), combined with Retrieval-Augmented Generation (RAG), present a transformative opportunity to automate the FMEA and FMECA processes. By leveraging AI-driven Natural Language Processing (NLP), these techniques can analyze vast amounts of reliable and trustworthy technical literature, industry standards, and historical failure data to generate FMEA/FMECA tables with minimal human intervention.

2.2 Aim of the Research

Aim of an ongoing project is to develop a framework that leverages Large Language Models (LLMs), integrated with Retrieval-Augmented Generation (RAG), to automate the generation of FMEA/FMECA tables with minimal human intervention.

This approach aims to enhance efficiency, accuracy, and scalability in risk assessment and failure mode analysis.

2.3 Scientific Novelty and Importance of Research

Publicly available software tools, such as ChatGPT, can generate FMEA/FMECA tables; however, they do not specify the sources used in their analysis, making it difficult to verify the reliability of the generated information.

The proposed framework will address this limitation by exclusively utilizing trustworthy, pre-vetted documents stored in a designated directory. This ensures that all extracted failure mode data, risk assessments, and mitigation strategies are based on credible, industry-accepted sources.

Additionally, the framework will provide explicit references for each piece of information included in the FMEA/FMECA table. By citing the original documents, users can manually cross-check the authenticity of the analysis, ensuring transparency, accuracy, and traceability in risk assessment.

The proposed methodology aims to enhance traditional FMEA/FMECA approaches by leveraging Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) to develop an automated, intelligent, and scalable risk assessment framework. This innovation is expected to provide several key advantages:

1. *Reduced Resource Consumption*: Automating data collection, analysis, and risk prioritization minimizes the labor-intensive nature of FMEA/FMECA, allowing personnel to focus on higher-value tasks.
2. *Faster Analysis*: AI-driven systems can rapidly process large datasets, enabling more frequent, dynamic, and real-time risk assessments compared to traditional manual methods.
3. *Use of Reliable and Trustworthy Information Sources*: The framework ensures that only pre-vetted, credible documents—including equipment-specific inspection and maintenance reports—are used, leading to more accurate and reliable risk assessments.
4. *Facilitates Manual Verification*: By citing original sources, the framework ensures that users can manually cross-check the authenticity of the information, improving transparency and accountability in risk assessment.

2.4 Limitations

The development of a reliable automated FMEA/FMECA system is a highly complex task, requiring a deep understanding of advanced algorithms, artificial intelligence (AI) techniques, and risk assessment methodologies. The effectiveness of such a system depends on multiple factors, including the level of automation required, the application domain, and available computational resources.

The project is currently at the proof-of-concept (PoC) stage, focusing on validating initial models and simulations to assess feasibility and accuracy. This phase includes:

- Testing AI-based FMEA/FMECA table generation using pre-vetted document sources.

- Evaluating accuracy of suggested failure mode, and failure mechanism by comparing AI-generated outputs with manually created FMEA/FMECA reports.
- Identifying computational and data quality challenges that may impact large-scale deployment.

As the project advances, future efforts will be directed toward:

- Refining and optimizing models
- Improving presentation and usability
- Ensuring regulatory and industry compliance
- Scaling for real-world applications

3 Methods

3.1 Data Collection

In this study, the Failure Mode and Effects Analysis (FMEA) task will be conducted by a large language model (LLM). The model's knowledge base will comprise a dataset containing previously conducted FMEA studies across various types of equipment. In addition to these empirical studies, the dataset will incorporate research focusing on FMEA ontologies, further enhancing the model's understanding of the FMEA process. Furthermore, to ensure alignment with industry standards and best practices, the dataset will include relevant ISO standards associated with performing FMEA.

3.2 Description of Retrieval-Augmented Generation (RAG) Framework

In this study, a *Retrieval-Augmented Generation (RAG)* framework is employed to enhance the performance of the language model by integrating domain-specific knowledge. RAG is a powerful approach that improves the accuracy, relevance, and contextual understanding of responses by allowing language models to retrieve external knowledge before generating output. This is particularly useful for technical domains like Failure Mode and Effects Analysis (FMEA), where precise and standardized information is crucial.

The RAG process consists of three primary stages: *Text Preprocessing*, *Information Retrieval*, and *Large Language Model (LLM) Integration.* Each of these stages plays a critical role in ensuring that the final output is both relevant and accurate.

Text Preprocessing: The first step in the framework involves Text Preprocessing, where all available textual data is converted into vector representations. This transformation is achieved using techniques such as word embeddings (e.g., Word2Vec,

FastText) or transformer-based embeddings (e.g., BERT, OpenAI's embeddings). These vector representations encapsulate the semantic meaning of the text, allowing for efficient and meaningful comparisons. Once the text is vectorized, it is stored in a vector database, which enables rapid and accurate retrieval of information when needed. The purpose of this stage is to create a structured and searchable knowledge base that can later be queried to retrieve relevant information. Figure 1 illustrates the steps involved in the Text Preprocessing stage.

Figure 2 outlines the comprehensive process of preparing documents and building a model for analyzing Failure Modes and Effects (FMEA) within the Retrieval-Augmented Generation (RAG) framework. Initially, all necessary documents are collected and organized in a designated folder. Key tools and libraries, such as LangChain for workflow management, document loaders (using Unstructured and PyPDF), and a vector database (Chroma DB) for storing vector representations, are then installed and configured.

The model can be built using either API key models (such as OpenAI, GROQ, or Mistral) or locally run models (like Ollama or those available through Hugging Face).

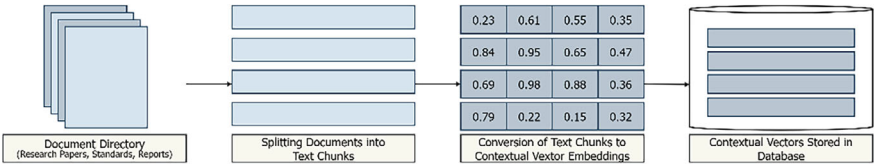


Fig. 1 llustration of the Text Preprocessing stage, where available textual data is transformed into vector representations and stored in a vector database

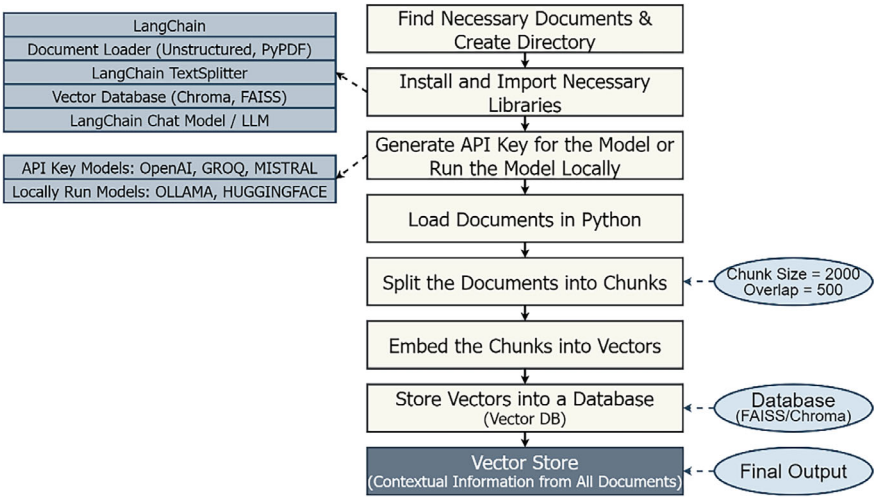


Fig. 2 Code structure for text pre-processing part of the RAG implementation

In this study, the Llama 3.1-8B model—derived from "GROQ"—has been employed, although both GROQ and Ollama are open-source alternatives. Additionally, while Chroma DB is used as the vector database in this setup, other options like FAISS (Facebook AI Similarity Search) can serve similar purposes. Overall, this structured setup under the RAG framework ensures that the model generates answers with precise and relevant context drawn from the prepared documents.

Once the setup is complete, the documents are loaded into Python and segmented into smaller "chunks"—each with a size of 2000 tokens and an overlap of 500 tokens—to ensure that contextual integrity is maintained. Each of these chunks is then converted into a vector representation, a numerical format that encapsulates the semantic meaning of the text. These vectors are stored in the vector database, forming a "vector store" that is essential for efficient and accurate retrieval.

Within the RAG framework, this vector store is later utilized during the Information Retrieval phase. When a user query is submitted, it is also transformed into a vector, and cosine similarity is used to identify the most relevant stored vectors. These retrieved vectors, together with the original query, are provided as input to the language model, which then generates a coherent and contextually rich response.

Information Retrieval: The second stage, Information Retrieval, is where the system processes a user's query by converting it into a vector representation, similar to how the initial dataset was transformed. This query vector is then compared with the stored vectors in the database using cosine similarity or other distance-based measures (e.g., Euclidean distance, dot product similarity). Cosine similarity is particularly effective as it measures the angular distance between two vectors, ensuring that semantically similar concepts are identified even if the exact wording differs.

Once the similarity scores are computed, the system selects the most relevant vectors from the database—those that closely match the user's query. This retrieval step ensures that the model has access to highly relevant external knowledge before proceeding to generate the final response. The efficiency and accuracy of this stage are critical, as retrieving incorrect or irrelevant data could compromise the reliability of the generated output.

Figure 3 depicts the Information Retrieval process within the RAG framework. When a user submits a query, it is first converted into a vector representation using the same embedding process applied during text preprocessing. This query vector is then compared against the stored vectors in the vector database via cosine similarity, which helps identify the most relevant matches. The top matching vectors, together with the original query, are then fed into the Large Language Model. The LLM processes this combined input to generate a coherent and contextually relevant response. Essentially, the LLM plays a dual role by both converting the text corpus and the query into vectors and generating the final output. Additionally, the framework leverages the LangChain library to efficiently manage tasks such as vector retrieval and text splitting.

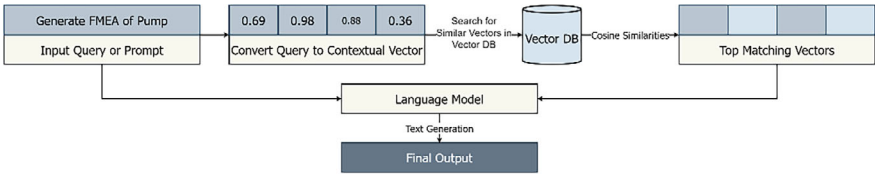


Fig. 3 Retrieval of information and generation of response based on user query

Large Language Model (LLM) Integration: The final stage, LLM Integration, occurs after the relevant vectors have been retrieved from the database. At this point, the Large Language Model (LLM) processes the retrieved information and generates a coherent, well-structured textual response. This is where the model integrates both its pre-trained knowledge and the retrieved external knowledge to formulate an informed and contextually appropriate output.

By combining retrieval-based knowledge with generative capabilities, the LLM ensures that responses are more precise, factually grounded, and tailored to the specific query. This hybrid approach significantly enhances the model's ability to perform FMEA-related tasks, as it can leverage historical FMEA studies, industry standards (such as ISO guidelines), and structured ontological data to produce accurate and insightful analyses.

In summary, the RAG framework optimally integrates structured retrieval and generative AI to enhance the accuracy and contextuality of the model's outputs. This approach not only improves the reliability of automated FMEA processes but also facilitates better decision-making and risk assessment, ensuring that the analysis aligns with established best practices.

3.3 Application of Retrieval-Augmented Generation (RAG) Framework

Figure 4 presents a comprehensive pipeline for establishing a question-and-answer system that leverages a Large Language Model within a Retrieval-Augmented Generation (RAG) framework to generate FMEA insights. In this setup, the process begins by configuring a retriever that efficiently queries a vectorized store of FMEA-related documents. This retriever is a key component of the RAG framework, ensuring that when a query is made, the most relevant contextual information is promptly accessible.

Next, the language model is imported and initialized, equipping it to interpret user queries and generate coherent responses. To seamlessly integrate the retrieval and generation components, a processing chain is constructed using the LangChain library. This chain effectively links the question-answering functionality with the retrieval process, so that the language model can incorporate the retrieved data to craft an informed answer.

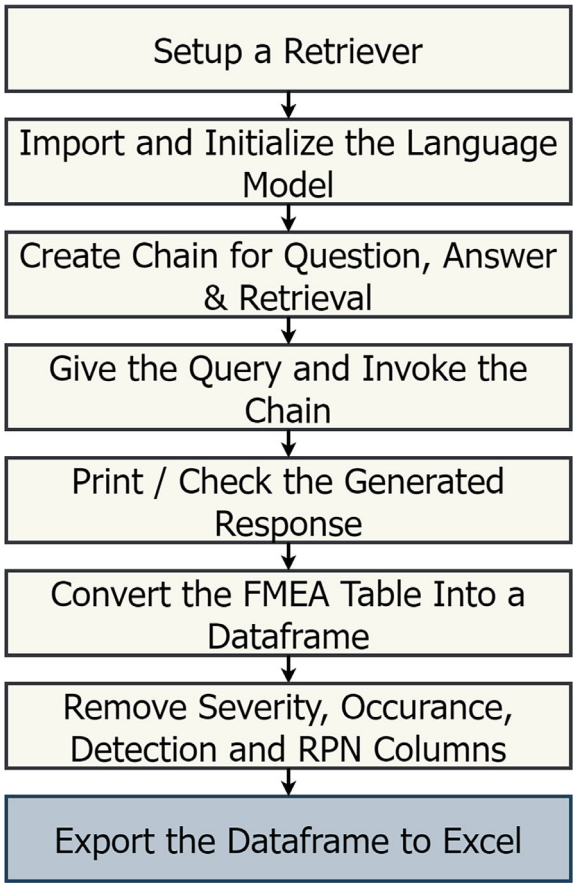


Fig. 4 Retrieval and output generation process

Once the system is set up, a user's query activates the chain. The retriever identifies and supplies the pertinent vectors from the database, and the language model uses both the original query and the retrieved information to generate a relevant response. The resulting output is then printed or checked for accuracy, ensuring the quality of the generated answer.

Subsequently, the FMEA data is transformed into a DataFrame—a structured format that facilitates easier manipulation and analysis. During this stage, certain columns such as Severity, Occurrence, Detection, and the Risk Priority Number (RPN) are removed to account for the variability that should instead be determined by a maintenance expert, considering the specific industry or organizational context.

Finally, the refined DataFrame is exported to Excel, allowing for further use and sharing of the processed FMEA data. Overall, this RAG-based workflow streamlines the retrieval, processing, and dissemination of critical FMEA information, ensuring that domain experts receive accurate and contextually relevant insights.

4 Results and Discussion

4.1 Implementation of the RAG System for FMEA Generation

The Retrieval-Augmented Generation (RAG) system was implemented within the Google Colab environment to facilitate automated Failure Mode and Effects Analysis (FMEA) generation. The process began with the input of a structured query—*Generate FMEA of pump*—which prompted the system to retrieve relevant information from a pre-processed knowledge base and generate a tailored FMEA specifically for pump operations. Leveraging the retriever and LLM pipeline, the system identified key potential causes, effects, and corrective actions related to pump failure, ensuring that the generated analysis aligned with industry best practices.

4.2 Output Refinement and Formatting

Once the initial FMEA output was generated, it was exported to an Excel file to enhance readability and organization. Python scripts were used for further data structuring, formatting, and optimization, ensuring that the FMEA table was clear and structured in a way that allowed for easy interpretation. The formatting process focused on improving text alignment, column spacing, and logical flow, ultimately enhancing usability for maintenance experts and engineers.

To further refine the output quality, an iterative prompting approach was employed. Multiple iterations of the query were tested, allowing for assessment of the system's responsiveness, accuracy, and consistency in generating FMEA documentation. Additionally, the temperature parameter of the model was adjusted at various levels to explore its impact on the clarity, relevance, and specificity of the generated content. Lower temperature settings were tested to encourage more deterministic and precise outputs, while higher temperatures introduced more variation and creativity in the generated responses. This experimentation provided valuable insights into optimizing the model's configuration for generating high-quality, structured FMEA reports.

4.3 Enhancements to the FMEA Table Structure

During the initial iterations, the generated FMEA table included essential columns typically found in Failure Mode and Effects Analysis reports. These columns consisted of:

- *Cause*: Identifying potential reasons for failure
- *Effect*: Describing the consequences of failure
- *Severity (S)*: Measuring the impact of failure
- *Occurrence (O)*: Estimating the likelihood of failure
- *Detection (D)*: Assessing how easily the failure can be detected
- *Risk Priority Number (RPN)*: A calculated metric (S × O × D) used for prioritizing risks
- *Corrective Action*: Outlining recommended measures to mitigate failure risks

To further enhance clarity and usability, additional refinements were applied after exporting the data to Excel. Python-based automation was used to adjust column formatting, standardize terminology, and improve table legibility. Additionally, a legend was incorporated to define the significance of the numeric ratings for Severity (S), Occurrence (O), and Detection (D). This legend provided a standardized reference, ensuring that users could correctly interpret the risk assessment metrics.

Figure 5 displays the initially generated FMEA table for pump systems, produced using the open-source language model Llama 3.1 within the RAG framework. In its original form, the table includes all available columns and features, providing a comprehensive analysis of potential pump system failure causes, effects, and corresponding corrective actions. The table is organized into two main sections, which together offer a detailed overview of the system's vulnerabilities.

However, after reviewing the initial output, specific columns—namely, Severity, Occurrence, Detection, and RPN—were removed. These metrics, while important, are highly subjective and can vary widely based on factors such as equipment condition, breakdown characteristics, operating environment, maintenance history, and the expertise of the evaluating personnel. By omitting these elements, the FMEA process is streamlined and becomes more adaptable across various industrial contexts. This refinement enables the analysis to be customized based on context-specific insights, ensuring that the evaluation remains both practical and aligned with real-world operational variability.

Figure 6 presents a simplified Failure Modes and Effects Analysis (FMEA) for a pump system, designed to highlight the most common issues affecting pump performance while outlining effective corrective actions. This streamlined analysis helps maintenance teams and engineers quickly identify potential failure causes, assess their impact, and implement targeted solutions to mitigate risks.

Failure Mode	Cause	Effect	Severity (S)	Occurrence (O)	Detection (D)	RPN	Corrective Action
1.1	Pump suction pipe not completely filled with liquid	Decrease in pump efficiency	4	6	3	72	Ensure pump suction pipe is filled with liquid before starting the pump
1.2	Insufficient available NPSH	Decrease in pump efficiency	4	4	3	48	Increase NPSH available to the pump
1.3	Selection of pump with too high a suction specific speed	Decrease in pump efficiency	4	2	6	24	Select a pump with a suitable suction specific speed
2.1	Impeller selection with abnormally high head coefficient	Decrease in pump efficiency	4	6	3	48	Select an impeller with a suitable head coefficient
3.1	Pump vibration at low flows	Decrease in pump efficiency	7	6	3	126	Balance the pump and ensure proper installation
4.1	Running the pump against a closed discharge valve without opening a by-pass	Damage to pump	10	2	6	120	Ensure the discharge valve is fully open before starting the pump
4.2	Operating pump below recommended minimum flow	Damage to pump	10	4	3	120	Ensure the pump is operating at or above the recommended minimum flow
5.1	Low flows of oil failure mode	Decrease in pump efficiency	4	4	3	48	Regularly check and maintain the oil levels in the pump
6.1	High temperature failure mode	Decrease in pump efficiency	7	6	3	126	Regularly check and maintain the temperature of the pump
7.1	Electrical failure mode	Damage to pump	10	2	6	120	Regularly check and maintain the electrical connections to the pump
8.1	Mechanical failure mode	Damage to pump	10	4	3	120	Regularly check and maintain the mechanical components of the pump
9.1	Blockage of pump	Damage to pump	10	4	3	120	Regularly check and maintain the pump for blockages

Fig. 5 Unformatted initial output from the RAG implementation, displaying the raw FMEA table for pump systems generated by the system

Failure Mode	Cause	Effect	Corrective Action
1.1	Pump suction pipe not completely filled with liquid	Decrease in pump efficiency	Ensure pump suction pipe is filled with liquid before starting the pump
1.2	Insufficient available NPSH	Decrease in pump efficiency	Increase NPSH available to the pump
1.3	Selection of pump with too high a suction specific speed	Decrease in pump efficiency	Select a pump with a suitable suction specific speed
2.1	Impeller selection with abnormally high head coefficient	Decrease in pump efficiency	Select an impeller with a suitable head coefficient
3.1	Pump vibration at low flows	Decrease in pump efficiency	Balance the pump and ensure proper installation
4.1	Running the pump against a closed discharge valve without opening a by-pass	Damage to pump	Ensure the discharge valve is fully open before starting the pump
4.2	Operating pump below recommended minimum flow	Damage to pump	Ensure the pump is operating at or above the recommended minimum flow
5.1	Low flows of oil failure mode	Decrease in pump efficiency	Regularly check and maintain the oil levels in the pump
6.1	High temperature failure mode	Decrease in pump efficiency	Regularly check and maintain the temperature of the pump

Fig. 6 Simplified pump FMEA

The table is systematically structured into four key columns—Cause, Effect, and Corrective Action—as previously described. Each row within the table represents a specific failure scenario, providing a logical breakdown of how a particular failure mode arises, its consequences on pump functionality, and the necessary steps to prevent or resolve the issue.

5 Conclusions

The paper outlines a proposed framework and shares preliminary results from an ongoing project aimed at developing a methodology for generating FMEA/FMECA tables solely from pre-vetted, credible documents. Leveraging the open-source language model Llama 3.1 within a Retrieval-Augmented Generation (RAG) framework, the system has already demonstrated encouraging performance in automating failure analysis. However, the research has also identified several areas for further enhancement to improve the overall FMEA generation system:

- Benchmarking against human-generated FMEA documents to better align the system with industry standards.
- Expanding the system's capabilities through multi-modal integration with CNNs, enabling the detection of failures in image and video data.
- Enhancing conversational consistency by incorporating memory buffers and chat-prompt templates, which would improve user interaction.
- Refining the RAG framework to independently process table and text data, thereby increasing output precision.

Together, these advancements promise to create a more robust, reliable, and efficient tool for failure analysis and risk management.

References

1. American Petroleum Institute (API): API 581: Risk-Based Inspection Methodology. API, Washington, DC (2016)
2. Bindingsbø, O.T., Singh, M., Øvsthus, K., Keprate, A.: Fault detection of a wind turbine generator bearing using interpretable machine learning. Front. Energy Res. **11**, 1284676 (2023). https://doi.org/10.3389/fenrg.2023.1284676
3. Cao, X., Xu, W., Zhao, J., Duan, Y., Yang, X.: Research on large language model for coal mine equipment maintenance based on multi-source text. Appl. Sci. **14**, 2946 (2024). https://doi.org/10.3390/app14072946
4. Cardiel-Ortega, J.J., Baeza-Serrato, R.: Failure mode and effect analysis with a fuzzy logic approach. Systems **11**(7), 348 (2023). https://doi.org/10.3390/systems11070348
5. Hassan, F.U., Nguyen, T., Le, T., Le, C.: Automated prioritization of construction project requirements using machine learning and fuzzy failure mode and effects analysis (FMEA). Automat. Construct. **154**, 105013 (2023). https://doi.org/10.1016/j.autcon.2023.105013

6. Hassani, I.E., Masrour, T., Kourouma, N., Motte, D., Tavčar, J.: Integrating large language models for improved failure mode and effects analysis (FMEA): a framework and case study. In: Proceedings of the Design Society, Vol. 4: DESIGN 2024, 2019–2028 (2024). https://doi.org/10.1017/pds.2024.204
7. International Organization for Standardization.: ISO 17359:2018—Condition Monitoring and Diagnostics of Machines—General Guidelines. ISO (2018)
8. International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC).: ISO/IEC Guide 73:2009—Risk Management—Vocabulary. Geneva, Switzerland, ISO (2009)
9. International Organization for Standardization (ISO).: ISO 55000:2014—Asset Management—Overview, Principles, and Terminology. Geneva, Switzerland, ISO (2014)
10. International Organization for Standardization (ISO).: ISO 55002:2018—Asset Management—Management Systems—Guidelines for the Application of ISO 55001. Geneva, Switzerland, ISO (2018)
11. International Organization for Standardization (ISO).: ISO 31000:2018—Risk management—Guidelines. Geneva, Switzerland, ISO (2018)
12. International Organization for Standardization (ISO).: ISO 31010:2019—Risk Management—Risk Assessment Techniques. Geneva, Switzerland, ISO (2019)
13. Ivančan, J., Lisjak, D.: New FMEA risks ranking approach utilizing four fuzzy logic systems. Machines **9**(11), 292 (2021). https://doi.org/10.3390/machines9110292
14. Kolios, A.J., Umofia, A., Shafiee, M.: Failure mode and effects analysis using a fuzzy-TOPSIS method: a case study of subsea control module. Int. J. Multicriteria Decis. Making **7**(1), 29–53 (2017)
15. Norwegian Petroleum Industry (NORSOK).: In: NORSOK Z-013: Risk and Emergency Preparedness Analysis. Oslo, Norway, Standards Norway (2010)
16. Peddi, S., Lanka, K., Gopal, P.R.C.: Modified FMEA using machine learning for food supply chain. Mater. Today: Proc. (2023). https://doi.org/10.1016/j.matpr.2023.04.353
17. Ray, P.P.: Transforming industrial reliability: the intersection of failure modes and effects analysis with artificial intelligence. J. Failure Anal. Prevent. **23**, 1383–1384 (2023). https://doi.org/10.1007/s11668-023-01697-6
18. Sader, S., Husti, I., Daróczi, M.: Enhancing failure mode and effects analysis using auto machine learning: a case study of the agricultural machinery industry. Processes **8**(2), 224 (2020). https://doi.org/10.3390/pr8020224
19. Stewart, M., Hodkiewicz, M., Li, S.: In: Large Language Models for Failure Mode Classification: An Investigation, pp. 1–8. (2023). ArXiV. https://doi.org/10.48550/arXiv.2309.08181
20. Thomas, D.: Revolutionizing failure modes and effects analysis with ChatGPT: unleashing the power of AI language models. J. Failure Anal. Prevent. **23**, 911–913 (2023). https://doi.org/10.1007/s11668-023-01659-y
21. Wang, H., Li, Y.: Large language model empowered by domain-specific knowledge base for industrial equipment operation and maintenance. In: Proceedings of the 5th International Conference on System Reliability and Safety Engineering (SRSE), pp. 474–479. (2023)
22. Wang, L., Sun, L., Kang, J., Wang, Y., Wang, H.: Risk identification of FPSO oil and gas processing system based on an improved FMEA approach. Appl. Sci. **11**(2), 567 (2021). https://doi.org/10.3390/app11020567
23. Xia, Y., Jazdi, N., Weyrich, M.: Enhance FMEA with large language models for assisted risk management in technical processes and products. In: Proceedings of the International Conference on Emerging Technologies and Factory Automation, 10–13 September 2024, Padova, Italy (2024)
24. Yucesan, M., Gul, M., Celik, E.: A holistic FMEA approach by fuzzy-based bayesian network and best–worst method. Complex Intell. Syst. **7**, 1547–1564 (2021). https://doi.org/10.1007/s40747-021-00279-z

Enhancing Circularity in SMEs by Incorporating Industry 4.0 Technologies—An Analysis on Gaps and Barriers



Rohit Singh, M. L. Mittal, Rohit Joshi, and Chhavi Solanki

Abstract The competitive, efficient, and circular transition of the current supply chain due to successful interaction between I4.0 and circularity is depicted by the current state of theory and practices. In spite of several proposed theoretical frameworks by academicians and available literature for this complex shift of linear supply chain to CSC, there is still a deficiency of organised, company-oriented, and practical approaches which is capable of transitioning linear supply chains towards circularity. In addition, operational challenges, urges, and constraints hindering CSCs emergence are still overpassed and, therefore, present a vast area of research in this domain. The theoretical gaps or voids in the literature pertaining to CSC-I4.0 are uncovered through a systematic gap analysis, and the potential barriers are investigated through a comprehensive literature review in this paper, with a major focus on SMEs. This two-fold field research area of Circular Economy and Industry 4.0 technologies is more comprehensively studied by extensively identifying the related scholarly articles using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method. This investigation rigorously reviews existing literature, encompassing 45 articles thoroughly analyzed, consolidated, and categorized across various dimensions. The hindrances in the metamorphosis of linear SC to CSC are then grounded into five categories and exhaustively studied, identifying the 7 most emphasized, critical, and recurrent realistic barriers. The study highlights the need for a strategic blueprint to act against CSC intransigent voids and realistic hindrances as in this research and act as cues for the stakeholders like policymakers, researchers, and managers.

R. Singh (✉) · M. L. Mittal
Department of Mechanical Engineering, Malaviya National Institute of Technology, Jaipur, India
e-mail: 2022RME9071@mnit.ac.in

R. Joshi
Operations and Quantitative Techniques, IIM Shillong, Shillong, India

C. Solanki
Department of Civil Engineering, JECRC University, Vidhani, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_44

Keywords Circularity · SME · Industry 4.0 · Circular supply chain · Circular economy · Lifecycle management

1 Introduction

The two main ideas propelling modern supply chains (SCs) toward increased competence and productivity are the circular economy (CE) and industry 4.0 (I4.0) [33, 36]. Various frameworks [12], models [19, 37, 41], methodologies [7, 16, 21] and Key performance indicators [30] focussed in the literature to demonstrate the smooth collaboration between CE, I4.0, and supply chains in adopting a circular transition.

Closed-loop supply chains (CLSCs), as opposed to linear SCs, focus on various waste treatment procedures and management techniques while also incorporating backflows that may generate extra value [3]. Circular supply chains (CSC) have been referred to in the literature by a number of terms, including closed-loop supply chain (CLSC), reverse supply chain, and green SC. CLSCs involve reusing recovered materials and incorporate Original Equipment Manufacturers (OEMs).

Despite a number of philosophical categories and studies, there is still a void in understanding the present problems that hinder the growth of CSCs. On one hand, various studies suggested, industrial strategies and Circular Business Models (CBMs) concerning this emerging context are either underway or currently deployed. On the other side, various studies suggested the essentiality of I4.0 technologies, mentioning that the sustainable I4.0 plays a pivotal role in fighting global competition in SME industries [31], or Floyd et al. [8], illuminating the obstructive consequence of digitalization.

This article addresses the need to move Industry 4.0's narrow focus on firms to a more comprehensive and systemic perspective that takes into account social development, governance, fiscal and monetary factors, and public dimensions. The way for investigating the defiant impact of Industry 4.0 technologies from the perspective of the Circular Economy will be paved by a more comprehensive and improved focus along the SC and society, helping to meet the requirements and demands of current and future generations. Indeed, as per Walker et al. [42], there is little exploration of I4.0 technologies from the CE dimension across the entire SC. Although attention on the individual topics of CE, I4.0, and Circular Supply Chain (CSC) has been depicted by various studies, still a deficit of priority assigned to the integrated concept of three dimensions, and its collaborative impact has been detected.

The preceding research question aids the investigator in understanding the main aim of this study. The research aims to accomplish the following:

RQ: To identify the gaps and barriers in employing Circularity in the Supply Chain of SMEs.

The architecture of the paper is as follows. The research approach is presented in Sect. 2. The key findings are presented in Sect. 3. In Sect. 4, limits and future research trends are discussed, along with concluding considerations.

2 Research Methodology

Prevalent CE, I4.0 studies in the literature have been structurally investigated and analysed to broaden the concept of the CSC readiness by identifying the barriers to the efficient implementation of the I4.0-CE model. For this purpose, a structured search string was formed and used to identify literature databases and find relevant and suitable articles for this research. The search strings used were ": (ABS ("Industry 4.0") OR ABS ("IoT") OR ABS ("Blockchain" OR "BT") OR ABS ("Digital technology") OR ABS ("AI" OR "ML") OR ABS ("Cyber Physical System") AND ABS ("Circular Supply chain") AND ABS ("Circularity") AND ABS ("Circular Economy") OR ABS ("Closed Loop") OR ABS ("Reverse Supply Chain")) AND ABS ("SME"). The academic databases, publishers' databases, and research engines that were selected as sources for the literature review were Science Direct, Springer, Web of Science, Google Scholar, Scopus, and Research Gate. These sources allowed access and coverage of a wide range of databases and literature that take into account CE or I4.0 perspectives and frameworks.

Scientific articles published from 2018 up to 2023 were gathered to investigate about half a decade from the initial diffusion of digital technologies. Removing the limitation of field content, to search for contributions, queries were used on titles, abstracts, and keywords. The initial research was done considering the broader field of analysis and related to the query "Supply Chain" AND "Industry 4.0′′ AND "Circular Economy", was constrained to journal articles and conference papers. However, with a desire to explore beneficial and productive contributions, further explorations were restricted only to journal articles. Thus, after removing superfluities, this led to the identification of 139 articles. Further, abstracts, titles, and keywords were reviewed in the first selection of articles from the identified sample of 139 articles, and complete manuscripts were assessed in the second screening of articles. This refining criteria led to the selection of 45 beneficial and productive articles focussing on the major trends in the field.

This study provides a systematic literature review (SLR) using the Prisma framework. PRISMA dispenses an excellent strategy to discuss emergent and sparkling topics and a replicable research approach that illustrates the complete review process. We conducted our investigation using the four stages framework given by Tranfield in 2003 for performing an SLR. The four stages includes: 1. Identification, 2. Screening, 3. Eligibility, and 4. Inclusion. Figure 1 describes the Prisma Framework in detail.

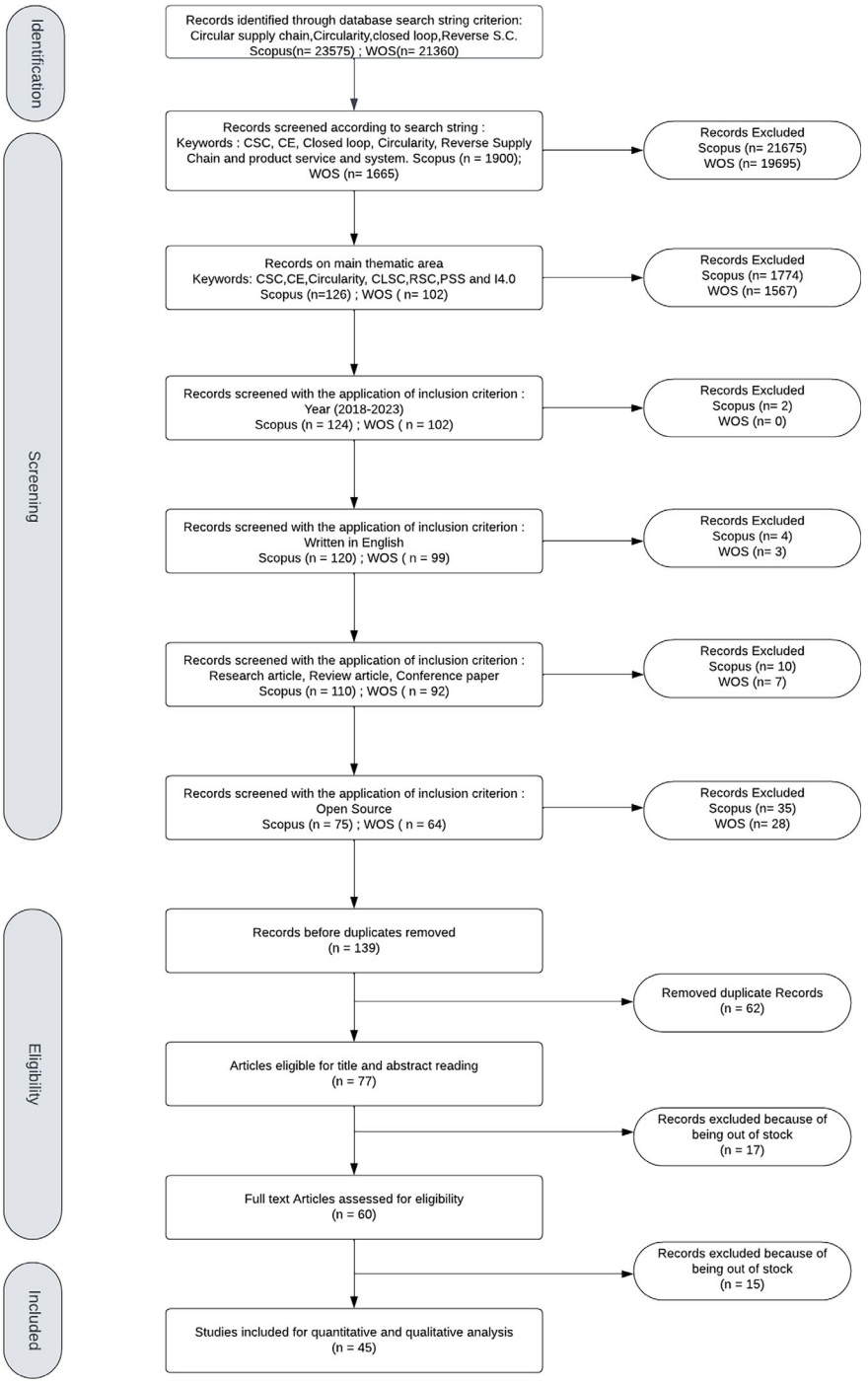


Fig. 1 PRISMA framework for paper selection

3 Findings

The outcomes of this research encompass the intransigent voids and realistic hindrances unfolded by the extensive and exhausted SLR.

3.1 Hypothetical Voids

The hypothetical voids in the realm of CSC, have been gathered in six categories as shown in Table 1, in order of decreasing diffusion rate.

In detail, *Implementation Deficiencies* were considered as insufficiency of directions, inadequate directives, shortage of orientation while administering production in an uncertain environment concerning CLSC (Hahn et al., 2020) and insufficiency of policies or plans to deploy I4.0 technology and increase SC's efficiency [35].

Systematic Integration Challenges, imply few resources in the study field that combine and take into account several factors. Ramirez et al. [34] condemned the lack of incorporation of 3 triple bottom layers into SC management (SCM), while Khandelwal and Barua [25] addressed a broad perspective of encompassing Circular SC Management (CSCM) in the Indian plastic industry.

Concerning the *Unidentified Barriers and Risks*, examples can be seen, as specified by Ozkan-Ozen et al. [32], in a scarcity of definition of concurred integrated hindrances of Cicularity, supply chain and I4.0 technologies, or as specified by Yadav

Table 1 Classification of gaps

Gap	Attributes
Implementation deficiencies	Scarcity of techniques, practices, solutions, mechanisms, and initiatives, to augment the CSC adoption practically
Systematic integration challenges	Inadequate integrated strategy capable to comprehend and embed twofold research context (CE-I4.0)
Unidentified barriers and risks	Undiscovered barriers, unidentified risks and mysterious threats in the CSC domain
Strategic limitations	Scarcity of strategical approaches and plan, inadequate managerial discussion, inefficient systemic management
Measurement and evaluation gaps	Absence of significant metrics and indicators, unaccounted measurable effects in the scope of CSC
Standardization and formalization issues	Insufficiency of formal definitions and identifications, lack of standardization. Inadequate classifications

et al. [45] about barriers to integrate Sustainable Supply Chain (SSC) decision making policies.

Strategical Limitations, as mentioned by Zhang et al. [46], are indicated by a lack of forums and conversation around managerial choices, regulatory regulations, and business models that either facilitate or impede the use of suitable technologies. Furthermore, the lack of a tactical goal necessary to integrate the CLSC's approach into traditional company operations was noted.

The *Measurement and Evaluation Gaps* category highlights a lack of quantifiable indicators and measures. Indeed, as suggested by Xie et al. [44] about scarcity of performance evaluation indicators of intelligent SC and as suggested by Jain et al. [23] about deficiency of sufficient qualitative and quantitative parameters to measure key performance of CSCs. Singh et al. [38] emphasized the paucity of research currently available for creating an index to evaluate how well SC cooperation with I4.0 practices works.

Lastly, major *Standardization and Formalization Issues* are referred as the systematization of SC parameters and indexes [30], an amalgamated widely recognised definition of Industry 4.0 in discourse of SC management [38] and as signified from [18] about an organised and structured foundation of prototypes helping the creation of Cyber Physical System concerning SC processes.

3.2 CSC Practical Barriers

The real-time obstacles were identified and grouped into 5 main classes. The 17 barriers were determined and classified into 5 major categories.

Administrative/senatorial barriers

- Scarcity of fiscal policies and stimuli: Kumar et al. [26] threw light on government's reluctancy to accelerate the development and as mentioned by Khandelwal and Barua [25] about the deficiency of policies pertaining to fiscal relief to encourage circularity in supply chain [28] about the deficient stimuli and rewards for greener and sustainable activities to advance circular models.
- Fragile and poorly enforced environmental laws: Khandelwal et al. [25], condemned the scarcity in standardised policies and enforcement rules for the environment.
- Scarcity of universal guidelines and performance indicators: ineffective evaluation of productivity and inefficient tracking of supply chain activities because of a fragile framework for measuring performance [26].

Economic barriers

- Restricted monetary resources and assistance: Horv´ath and Szab´o [20] and Dey et al. [5] highlighted the scarcity of monetary resources and financial support,
- Significant expenditures and implementation expenses: as suggested by Ozkan-Ozen et al. [32], transforming SCs into CSCs, current SCs need to be remodelled

concurrently with the acquisition of technologies that facilitate I4.0 and as signified by Sharma et al [39], that this transition requires high expertise like technical training, equipments, and consultancy which however, leads to significant expenditures and implementation expenses.
- Unsureness about financial yield: Kazancoglu et al. [24] discussed the expectations and propensity of hesitant enterprises to make investments due to the reason mentioned by Ozkan-Ozen et al. [32] of the uncertain return of investment and risk enlargement due to the unpredictable nature of circular flows.

Social/Civil Barriers

- Fear and reluctance against revolutionary transitions: As stated by Yadav et al. [45], reluctance to reform and, as quoted by Kumar et al. [26], Employees' dread of automation-related layoffs lead to a challenging environment for the acquisition.
- Lack of skilled expertise: Sharma [39] pointed the necessity of enhanced expertise required in order to oversee the latest I4.0 technologies, which may hamper the outcome of CSC adoption, and as mentioned by Kazancoglu et al. [24], about the weakened intellect of the workforce about CE concepts and practices.
- Scantness of awareness and knowledge: A purposeful scarcity of CSCM schemes and initiatives was criticized from a market perspective [25] as well as from an organizational perspective, which led to a decreased level of determination [26].

Institutional/Organizational Barriers

- Inadequacy of collaboration and coordination in the SC architecture: Members of the SC are unwilling to assist, plan, and support CSC activities [6]. The differing opinions, fear of helplessness and chaos, or lack of faith leads to an unsupportive attitude [24].
- Inadequate management endorsement and assistance: As commercial potential and challenges are not fully understood, deficient financial provisions, and blurred vision, the management is generally resistant in aiding initiatives for sustainable executions [1]
- Deficiency of relevant training and reasonable educational programmes: Mangla et al. [28] pointed out the proper and relevant training and pertinent development programmes for stakeholders.
- Ineffective strategic blueprint: Mangla et al. [28] discussed the scantness of planning, and administration of CSC management principles and concepts, and Khandelwal and Barua [25] firmly confirmed this.
- Deficient organization proactiveness, trust, and readiness: The SC network must be remodeled in order to move from linearity towards circularity while also implementing sustainability and an I4.0 transition. In fact, organizations lack creative foresight and are uncertain about this new construct [32, 39]

Technical barriers

- Insufficient information systems and data management: There is a larger need for data management skills as in CSC the complexity increases tremendously [32].

As such, inadequate information systems and data management proves to be an important barrier.

- Inadequate improvised technology, facilities, infrastructure, and assets: Sharma et al. [39] mentioned degraded internet connectivity and the inadequacy of desired infrastructures as crucial hindrances to a sustainable I4.0-driven work environment.
- Meagre compatibility and unification of technical platforms: Synergy is necessary for the methodical blending of new and traditional methods [39].

Particularly in an evolving economy setting with particular reference to SMEs, the challenges and obstacles associated with the adoption of CSCs bordered by Industry 4.0 technologies are persistent and hence require a greater focus.

4 Conclusion

This study aided in CSC research domain (two fold perspective (I4.0, CSC)) with two main results of identifying 6 hypothetical voids, 17 realistic hindrances grounded in 5 categories.

An empirical research methodology stimulated to the PRISMA method has been developed, and a systematic literature review was executed to diagnose an assortment of intransigent gaps and realistic hindrances pertinent to CSC.

Implementation Deficiencies and Systematic Integration Challenges are found to be dominant in Hypothetical voids. This extendedly emphasized literature research scrutinizing tangible outcomes of deployment of CSC with a strategic, comprehensive, analytical, and integrated approach. The 17 identified barriers were grounded into 5 major groups and further analysis led to showcase 7 most emphasized barriers: Scarcity of fiscal policies and stimuli, Fragile and poorly enforced environmental laws, restricted monetary resources and assistance, significant expenditures and implementation expenses, Meagre compatibility and unification of technical platforms, Inadequacy of collaboration and coordination among the SC members, and Inadequate technological advancement, facilities, infrastructure and assets.

This work might be viewed as a prelude to forthcoming insights and innovations that are consistent with the study that is being presented and are desired to fill the voids that have been identified. As a result, the researchers, approaching institutions and organisations will benefit from the vast reservoir of knowledge that is empowered with profound understanding and deeper insight.

Finally, the hypothetical gaps or challenges specified in the article showcased the requirement to detect and recommend innovative and improved solutions, models, and practical approaches to elevate the CSCs' operationalisation and improvise their outreach. Alternatively, from a philosophical perspective, explicit, unambiguous, and meticulous sustainability appraisal and assessment methods should be evolved, analysed, and suggested, and CSC interactions and mechanisms, as well as rigorous benchmarks, must be faced in order to monitor their operations.

References

1. Cezarino, L.O., Liboni, L.B., Oliveira Stefanelli, N., Oliveira, B.G., Stocco, L.C.: Diving into emerging economies bottleneck: industry 4.0 and implications for circular economy. Managem. Decision (2019)
2. Chauhan, C., Singh, A.: A review of industry 4.0 in supply chain management studies. J. Manuf. Technol. Managem. **31**(5), 863–886 (2019)
3. Colangelo, G., Facchini, F., Ranieri, L., Starace, G., Vitti, M.: Assessment of carbon emissions' effects on the investments in conventional and innovative waste-to-energy treatments. J. Cleaner Prod. **388** (2023)
4. De Angelis, R., Howard, M., Miemczyk, J.: Supply chain management and the circular economy: towards the circular supply chain. Prod. Planning Control **29**(6), 425–437 (2018)
5. Dey, P.K., Malesios, C., De, D., Budhwar, P., Chowdhury, S., Cheffi, W.: Circular economy to enhance sustainability of small and medium-sized enterprises. Business Strategy Environ. **29**(6), 2145–2169 (2020)
6. Farooque, M., Zhang, A., Liu, Y.: Barriers to circular food supply chains in China. Supply Chain Managem. **24**(5), 677–696 (2019)
7. Fern´andez-Caram´es, T.M., Blanco-Novoa, O., Froiz-Míguez, I., Fraga-Lamas, P.: Towards an autonomous industry 4.0 warehouse: a UAV and blockchain-based system for inventory and traceability applications in big data-driven supply chain management. Sensors (Basel. Switzerland) **19**(10) (2019)
8. Floyd, J.A., D'Adamo, I., Wamba, S.F., Gastaldi, M.F.: Competitiveness and sustainability in the paper industry: the valorisation of human resources as an enabling factor. Comput. Indus. Eng. 110035 (2024)
9. Ehie, I., Ferreira, L.M.D.F.: Conceptual development of supply chain digitalization framework. IFAC-PapersOnLine **52**(13), 2338–2342 (2019)
10. Garay-Rondero, C.L., Martinez-Flores, J.L., Smith, N.R., Caballero Morales, S.O., Aldrette-Malacara, A.: Digital supply chain model in industry 4.0. J. Manuf. Technol. Managem. **31**(5), 887–933 (2019)
11. Garrido-Hidalgo, C., Olivares, T., Ramirez, F.J., Roda-Sanchez, L.: An end-to-end internet of things solution for reverse supply chain management in industry 4.0. Comput. Indus. **112**, Article 103127 (2019)
12. Garrido-Hidalgo, C., Ramirez, F.J., Olivares, T., Roda-Sanchez, L.: The adoption of internet of things in a circular supply chain framework for the recovery of WEEE: the case of lithium-ion electric vehicle battery packs. Waste Managem. **103**, 32–44 (2020)
13. Geissdoerfer, M., Naomi, S., Monteiro, M., Carvalho, D., Evans, S.: Business models and supply chains for the circular economy. **190**, 712–721 (2018)
14. Genovese, A., Acquaye, A.A., Figueroa, A., Koh, S.C.L.: Sustainable supply chain management and the transition towards a circular economy: evidence and some applications. Omega (United Kingdom) **66**, 344–357 (2017)
15. Gonz´alez-S´anchez, R., Settembre-Blundo, D., Ferrari, A.M., García-Mui˜na, F.E.: Main dimensions in the building of the circular supply chain: a literature review. Sustainability (Switzerland) **12**(6), 1–25 (2020)
16. Gonz´alez Rodríguez, G., Gonzalez-Cava, J.M., M´endez P´erez, J.A.: An intelligent decision support system for production planning based on machine learning. J. Intell. Manuf. **31**(5), 1257–1273 (2020)
17. Hahn, G.J.: Industry 4.0: a supply chain innovation perspective. Int. J. Prod. Res. **58**(5), 1425–1441 (2020)
18. Hetterscheid, E., Schlüter, F.: Morphology for designing cyber-physical systems in supply chain planning and control processes. Logist. J. 1–11 (2019)
19. Hoffa-Dabrowska, P., Grzybowska, K.: Simulation modeling of the sustainable supply chain. Sustainability (Switzerland) **12**(15) (2020)

20. Horv´ath, D., Szab´ o, R.Z.: Driving forces and barriers of industry 4.0: do multinational and small and medium-sized companies have equal opportunities? Technol. Forecast. Soc. Change **146**, 119–132 (2019)
21. Hussain, M., Malik, M.: Organizational enablers for circular economy in the context of sustainable supply chain management. J. Cleaner Prod. **256**, Article 120375 (2020)
22. Jia, F., Yin, S., Chen, L., Chen, X.: The circular economy in the textile and apparel industry: a systematic literature review. J. Cleaner Prod. **259**, Article 120728 (2020)
23. Jain, S., Jain, N.K., Metri, B.: Strategic framework towards measuring a circular supply chain management. Benchmarking **25**(8), 3238–3252 (2018)
24. Kazancoglu, I., Kazancoglu, Y., Yarimoglu, E., Kahraman, A.: A conceptual framework for barriers of circular supply chains for sustainability in the textile industry. Sustain. Developm. **28**(5), 1477–1492 (2020)
25. Khandelwal, C., Barua, M.K.: Prioritizing circular supply chain management barriers using fuzzy AHP: case of the Indian plastic industry. Global Business Rev. (2020)
26. Kumar, P., Singh, R.K., Kumar, V.: Managing supply chains for sustainable operations in the era of industry 4.0 and circular economy: analysis of barriers. Resources Conserv. Recycling **164**, 105215 (2021)
27. Luthra, S., Kumar, A., Zavadskas, E.K., Mangla, S.K., Garza-Reyes, J.A.: Industry 4.0 as an enabler of sustainability diffusion in supply chain: an analysis of influential strength of drivers in an emerging economy. Int. J. Prod. Res. **58**(5), 1505–1521 (2020)
28. Mangla, S.K., Luthra, S., Mishra, N., Singh, A., Rana, N.P., Dora, M., Dwivedi, Y.: Barriers to effective circular supply chain management in a developing country context. Prod. Plann. Control **29**(6), 551–569 (2018)
29. Masi, D., Day, S., Godsell, J.: Supply chain configurations in the circular economy: a systematic literature review. Sustainability **9**(9), 1602 (2017)
30. Morella, P., Lamb´an, M.P., Royo, J., S´anchez, J.C., Corrales, L.D.C.N.: Development of a new green indicator and its implementation in a cyber–physical system for a green supply chain. Sustainability (Switzerland), **12**(20), 1–19 (2020)
31. Narwane, V.S., Raut, R.D., Yadav, V.S., Singh, A.R.: Barriers in sustainable industry 4.0: a case study of the footwear industry. Int. J. Sustain. Eng. **14**(3), 175–189 (2021)
32. Ozkan-Ozen, Y.D., Kazancoglu, Y., Kumar Mangla, S.: Synchronized barriers for circular supply chains in industry 3.5/industry 4.0 transition for sustainable resource management. Resour. Conservat. Recycling **161**, Article 104986 (2020)
33. Rajput, S., Singh, S.P.: Connecting circular economy and industry 4.0. Int. J. Inform. Managem. **49**, 98–113 (2019)
34. Ramirez-Pe˜na, M., Fraga, F.J.A., Salguero, J., Batista, M.: Assessing sustainability in the shipbuilding supply chain 4.0: a systematic review. Sustainability (Switzerland) **12**(16) (2020)
35. Raut, R.D., Gotmare, A., Narkhede, B.E., Govindarajan, U.H., Bokade, S.U.: Enabling Technologies for Industry 4.0 manufacturing and supply chain: concepts, current status, and adoption challenges. IEEE Eng. Managem. Rev. **48**(2), 83–102 (2020)
36. Rosa, P., Sassanelli, C., Urbinati, A., Chiaroni, D., Terzi, S.: Assessing relations between circular economy and industry 4.0: a systematic literature review. Int. J. Prod. Res. **58**(6), 1662–1687 (2020)
37. Safiullin, M.R., Savelichev, M.V., Elshin, L.A., Moiseev, V.O.: Increasing stability of economy through supply chain management and the circular economy. Int. J. Supply Chain Managem. **9**(4), 721–725 (2020)
38. Singh, R.K., Kumar, P., Chand, M.: Evaluation of supply chain coordination index in context to industry 4.0 environment. Benchmarking (2019)
39. Sharma, M., Kamble, S., Mani, V., Sehrawat, R., Belhadi, A., Sharma, V.: Industry 4.0 adoption for sustainability in multi-tier manufacturing supply chain in emerging economies. J. Cleaner Prod. **40**, Article 125013 (2020)
40. Taddei, E., Sassanelli, C., Rosa, P., Terzi, S.: Circular supply chains in the era of industry 4.0: a systematic literature review. Comput. Indus. Eng. **170** (108268) (2022)

41. Tripathi, S., Gupta, M.: A framework for procurement process re-engineering in industry 4.0. Business Process Managem. J. (2020)
42. Walker, A.M., Vermeulen, W.J.V., Simboli, A., Raggi, A.: Sustainability assessment in circular inter-firm networks: an integrated framework of industrial ecology and circular supply chain management approaches. J. Cleaner Prod. **286**, 125457, Elsevier Ltd.
43. Wei, F., Liu, S., Yin, L., Li, W., Yu, Z.: Research on performance evaluation system for green supply chain management based on the context of recycled economy—taking Guangxi's manufacturing industry as example. J. Grey Syst. **26**(2), 177–187 (2014)
44. Xie, Y., Yin, Y., Xue, W., Shi, H., Chong, D.: Intelligent supply chain performance measurement in industry 4.0. Syst. Res. Behav. Sci. **37**(4), 711–718 (2020)
45. Yadav, S., Garg, D., Luthra, S.: Analysing challenges for internet of things adoption in agriculture supply chain management. Int. J. Indus. Syst. Eng. **36**(1), 73–97 (2020)
46. Zhang, A., Venkatesh, V.G., Liu, Y., Wan, M., Qu, T., Huisingh, D.: Barriers to smart waste management for a circular economy in China. J. Cleaner Prod. **240**, Article 118198 (2019)

Modern Tools for Understanding Causality: A Comparative Review of Inference Methods



Ayushi Rajpoot, Paras Garg, Gunjan Soni, and A. P. S. Rathore

Abstract Causal inference is important for discovering cause-and-effect relationships across disciplines, enabling insights beyond simple correlations. This paper reviews techniques like Graph Neural Networks (GNNs), Causal Forests, DEMATEL, (ISM) Interpretive Structural Modelling, (FCM) Fuzzy Cognitive Mapping, Bayesian Networks, Structural Equation Modelling (SEM), Directed Acyclic Graphs (DAGs), Granger Causality Analysis, and Markov Chains. These methods are compared across metrics such as conceptual framework, applicability, data requirements, model complexity, interpretability, accuracy, assumptions, and ease of use. Traditional techniques like SEM and Bayesian Networks excel in modeling latent variables, while DAGs clarify confounders. Machine learning-based methods like GNNs and Causal Forests handle high-dimensional data and heterogeneous effects. DEMATEL, ISM, and FCM are valuable for decision-making in systems with uncertainty. This review highlights the evolution from statistical to advanced computational approaches, emphasizing their growing importance in decision-making, policy evaluation, and complex system analysis across healthcare, economics, and AI.

Keywords Causality · Causal inferences · Interdependence modelling · Inter-causal relationship techniques

1 Introduction

In this paper, we analysis the expanding research on causal interdependence. A statistical method for establishing cause-and-effect correlations, causal inference is crucial in many fields. The goal of causal analysis is to deduce not only probabilities or beliefs beneath static settings, but too the changing aspects of those beliefs under dynamic situations, such as those induced by new policies, therapies, or other external

A. Rajpoot · P. Garg (✉) · G. Soni · A. P. S. Rathore
Malaviya National Institute of Technology Jaipur, Jaipur, India
e-mail: 2021rme9075@mnit.ac.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_45

interventions [1]. Causal inference is influential tool for understanding the effects of interventions and explaining complex phenomena. It goes beyond basic correlation, allowing us to estimate the consequences of changing our behaviour or acting in a different way in the past [2]. Although correlation and causation have fairly different meanings, they are sometimes used interchangeably. A broad link is indicated by correlation; two variables are associated if they show an upward or downward trend. When a cause and an effect are partially dependent on one another and the cause is partially to blame for the result, this is known as causation. The procedure of defining a causal relationship based on the circumstances surrounding the occurrence of a result. Causal inference differs from correlation inference primarily in that it examines how the effect variable changes when the cause is altered. "Correlation does not imply causation," as is widely known. For instance, a study found that girls who eat breakfast typically weigh less than those who don't, supporting the idea that eating breakfast can aid in weight loss. However, rather than being causally related, these two events can only be correlated. Perhaps the females who have breakfast every day lead healthier lives, fetching in regular exercise, getting sufficient sleep, and upholding a stable diet, all of which donate to their ultimate weight loss. In this instance, eating breakfast and being lightweight are both caused by leading a healthy lifestyle; as such, we can also consider it to be a confounding factor in the relationship between the two [3].

1.1 Importance

For the purpose of estimating the impacts of interventions and guiding policy decisions, causal inference is essential in many domains. In psychiatric epidemiology, causal inference techniques are essential for developing actionable etiologic models and directing prevention efforts [4]. Statistical models, natural experiments, and randomized clinical trials are some of these techniques; each has advantages and disadvantages. In nursing research, causal inference is also essential, especially when figuring out how nurse staffing affects patient outcomes [5]. Moreover, the application of causal inference techniques may improve explainability and impartiality in machine learning models, which will aid in the making of more accountable and transparent AI systems. Since robust causal inference techniques are classically unavailable for randomized controlled experiments, observational studies offer valuable substitutes for deducing causal relationships from gathered data [2].

1.2 Scope of Review

The review paper on causal inference delivers in depth research of the crucial statistical method, which is essential for creating cause-and-effect relationships across various fields. The paper highpoints the importance of causal inference in empirical

sciences by succeeding its development from philosophical foundations to current statistical techniques. It covers fundamental ideas, historical progress, fundamental concepts, and cutting-edge methods used in a variety of arenas. Besides simple correlation, causal inference aids researchers comprehend how beliefs and behaviour change in retort to interventions or environmental shifts. Without taking into version underlying factors like general lifestyle picks, the research highlights that while correlation displays a linking between two variables, causation suggests that one variable directly marks another. It addresses the value of causal inference in fields like machine learning, nursing research, and mental epidemiology, accenting how these methods influence policy choices and improve model elucidation. In order to grasp complex systems and cascading failures—both of which are vital for risk investigation and resilience strategies—the study also discusses interdependency modelling methodologies. The paper highlights many approaches used for causal inference, their aces and ploys, and applicability in real-world situations. It also compares the crucial causal inference strategies based on metrics. In order to ensure application to contemporary research objectives, the study also measures new trends and methodologies that progress causal analysis. talked about their positive and negative features.

1.3 Structure of Review Paper

The review's structure is methodically divided into multiple sections, introductory with an overview of causal inference and its status. The scope of the review is then considered in the paper, along with the many approaches and uses of causal inference. The section on review methodology explains how the literature was accumulated and examined to give a thorough summary. The next sections explore central ideas, such as historical developments and important causal inference principles. Propensity score matching (PSM) and randomized controlled trials (RCTs), two statistical practices that are essential for defining causal correlations in both experimental and observational investigations, are also roofed in the study. Clarity and depth in comprehending the workings of causal inference are ensured by this methodical approach.

2 Foundational Concept of Causal Inferences

This section covers the expansion stages of causal inference, foundational principles of causal model and as well covers the data about various causal inference practices.

2.1 Historical Background

Causal inference, the process of decisive cause-effect relations, has a rich historic background straddling century. The historical basics of causal inference can be drawn back to empiricist logicians like Bacon, Locke, and Hume, who emphasized reflection and inductive perceptive. In the nineteenth century, Koch provided a framework for identifying diseases associated with microorganisms, while the twentieth century saw efforts by Hill and others to systematize causal inference in observational research. Wright's 1918 invention of path analysis, though a milestone, faced initial challenges due to philosophical and conceptual barriers [6]. The formal development of causal inference is primarily a 20th-century phenomenon, with recent advancements in statistical methods and computing tools [7]. The field has evolved from early theories of causation to modern approaches using directed acyclic graphs and novel analytical methods. Key concepts in causal inference, such as potential outcomes and counterfactuals, have emerged over the past few decades, building on traditional statistical methods like estimating equations and semiparametric theory [8].

2.2 Fundamental Principle

Our thoughtful of causal inference has since augmented numerous folds, due chiefly to advances in three areas:

1. Nonparametric structural equations.
2. Graphical model
3. Symbiosis among counterfactual and graphical approaches.

These developments are essential to the experimental sciences since most investigations in the social, behavioural, and health sciences are driven by causal rather than statistical research concerns.

Erstwhile to the validation of causal inference via the latent outcome framework, however, rather fewer severe but significant discussions of causality arose. Causation can be well-defined as the process by which one event leads to the production of another event. In statistical terms, causation is present when the change in one variable results in a change in another variable. For example, working more hours at a job that pays hourly will lead to a larger pay check; here, the number of hours worked directly causes the increase in income [9]. This concept is fundamental in various fields, including statistics, philosophy, and science. A causal method to an examination must first describe the causal research question, and then classically proceeds as per the steps given in Fig. 1:

Causal assumptions (step 5) are frequently expressed in terms of counterfactuals, or possible outcomes. Finding the pertinent variables required to propose a realistic

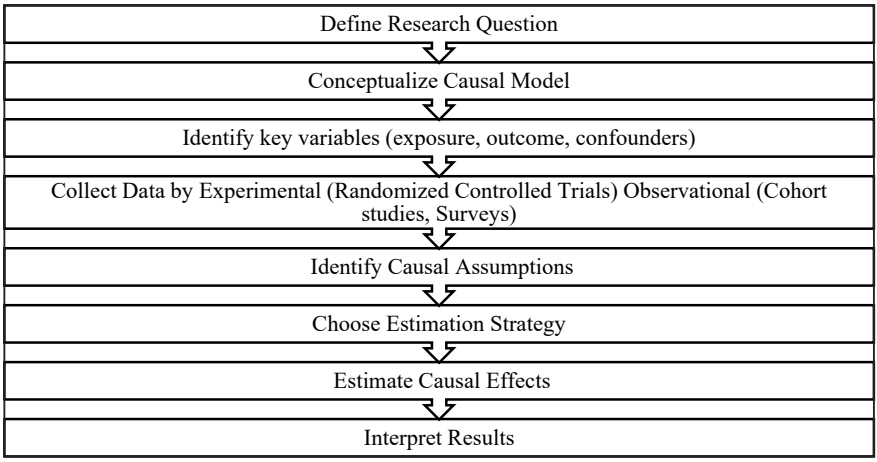


Fig. 1 Steps of causal analysis

model for the data producing structure from current literature or substantive specialists is a crucial part of assessing the effect (step 7). One way to express this is with a directed acyclic graph (DAG) [8].

Crucial Questions in Causal Inference:

Does X cause Y, or is it merely associated with Y?

What is the magnitude of the causal effect of X on Y?

Can we predict the effect of interventions on X?

2.3 Causal Graphs

Description (*Causal Structure*):

A causal construction of a set of variable quantity V is a directed acyclic graph (DAG) in which individually node corresponds to a dissimilar element of V, and each link signifies a direct useful relationship among the consistent variables. [10]

Description (*Directed Acyclic Graphs*):

A tool for visually representing causal traditions made about the data-generating relations being considered in a given practical problem, directed acyclic graphs (DAGs), also known as causal graphs, can help identify variables that may skew or distort the causal effect or effects of interest. DAGs are visual tools used in research to represent causal relationships and guide study design and analysis [1]. DAGs consist of nodes connected by one-directional arrows, depicting the flow of information from cause to effect without cyclic paths. They help researchers identify confounding factors, selection bias, and inappropriate statistical adjustments that

could introduce bias. The nodes in a generalised directed acyclic has following way of alignment of its edges: [11]

(i) A source node: both edges point out,
(ii) A sink node: both edges point in
(iii) A neutral node: one edge points in and the other out.

If there are arrows connecting each inter-variable link, a graph is considered to be directed. An arrow as of one variable to another signifies that changes in the initial variable affect the second. If there are no closed loops in the graph, it is considered acyclic. Lastly, all joint causes of treatment Z and outcome Y (whether measured or not) must be represented in the diagram for a DAG to be considered causative [8].

Take into consideration, for instance, the left panel of Fig. 2. Here, we can observe that M functions as a intermediary in the link between P and B, while A is a common cause and, consequently, a confounding factor in that relationship. DAGs can be expanded to represent treatment classifications, with actions and overruling variables observed at many time-points, as in the right panel of Fig. 1, or they can be expanded to embrace extra covariates, for example, by taking confounders into separate consideration or by adding variables that might affect B but not P. While taking into account any other channels that connect the two variables, the goal of causal interdependence is to measure the "flow" of the influence of an interference on treatment P across all directed trails that attach P to B. In the left panel of Fig. 2, for instance, there are two trails (one direct P → B, the other passing through M) by which P might act on B. In disparity, there is an undirected path (P ← A → B) that can be used to observe a co-variation between P and B until the trail is congested, which is usually accomplished by acclimatizing on A. Conditioning on A eliminates confounding by eliminating the path P ← A → B, which is referred to as a backdoor or confusing path. Figure 2 right panel depicts a more complicated scenario with less evident conditioning techniques to eliminate confusion [8].

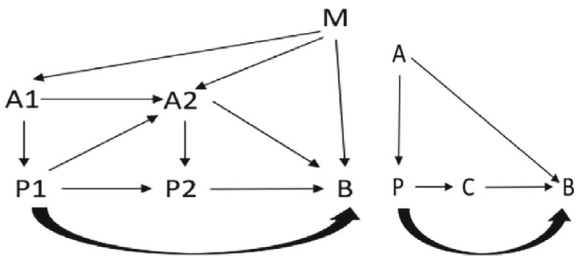


Fig. 2 Instances of directed acyclic graphs. The left panel depicts a straightforward scenario with a sole treatment, P, whose impact on the result, B, is mediated by C and confused by A. The longer, more intricate example on the right side has two sequential treatments, P1 and P2, each of which is temporally preceded by covariates A1 and A2, and a final result, B. A1, A2, and B are causally impacted by a potentially unseen variable M

3 Comparison of Different Causal Techniques

This section covers the details about the comparison of different causal inference techniques, guidelines for applicability of these and their strengths and weaknesses.

3.1 Comparison

Following table covers the comparison of causal inference techniques on the basis of selected metrics which gives the information regarding the its conceptual framework, applicability, data required for the techniques, its complexity, interpretability, accuracy, assumption required and also discussed about the software tool that can be used for the applicability of causal inference techniques (Table 1).

3.2 Pros and Cons of the Causal Inference Techniques

Following section covers the strengths and weaknesses of interdependency techniques which will be useful for the applicability of these techniques (Table 2).

4 Discussion and Conclusion

As discussed in this paper about causal interdependency techniques and comparison among different techniques and its applicability covering the strength and weakness, this section gives the detail about the new emerging causal inference techniques, summary and research implication of the techniques discussed in above sections. This paper has provided a detailed analysis of causal inference methods, emphasizing their significance for understanding cause-and-effect relationships across a wide array of disciplines. Beyond simple correlation examination, causal inference provides more in-depth understanding of how fluctuations in one variable might have a direct impact on another. By exploring numerous techniques and approaches, this review exemplifies how causal inference has evolved into a critical tool for decision-making, estimation of policies, and research in arenas such as healthcare, economics, social sciences, and machine learning. The paper discusses numerous important approaches, such as Granger Causality Analysis, Markov Chains, Directed Acyclic Graphs (DAGs), Bayesian Networks, and Structural Equation Modelling (SEM). For probabilistic perceptive in inexact systems, Bayesian networks function incredibly well, especially when dealing with partial data. SEM, on the other hand, is useful in the social sciences for modelling correlations between latent and observable variables. DAGs make it humbler to identify confounders and mediators

Table 1 Comparison of causal inference techniques

S No	Techniques	Conceptual framework	Applicability	Data requirements	Practical implications	References
1	Bayesian network	Probabilistic reasoning	Risk assessment, diagnostic modelling	Complete data, prior knowledge	Medical diagnosis, supply chain risk	[12–14]
2	Structural equation modelling	Statistical, Graphical models	Social sciences, latent variable modelling	Large sample sizes	Psychological, economic data	[15–17]
3	Directed acyclic graphs (DAGs)	Graph theory, probabilistic	Causal inference, epidemiology, econometrics	Observational data	Health research, policy interventions	[8, 18, 19]
4	Granger causality analysis	Time-series causality	Time-series forecasting, econometrics	Time-series data	Economic forecasting, stock market analysis	[20–22]
5	Markov chain	Stochastic processes	Predictive modelling, decision-making under uncertainty	Transition matrices, probabilities	Inventory systems, queueing theory	[23–25]
6	Graph neural network	Deep learning, Graph theory	Complex network analysis, recommendation systems	Graph-structured data	Social networks, fraud detection, personalized systems	[26–28]
7	Causal forest	Machine Learning, Decision trees	Heterogeneous treatment effects, personalization	Large datasets	Personalized medicine, targeted marketing	[29–31]
8	DEMATEL	Decision-making, Causal mapping	Group decision-making, problem structuring	Expert judgment	Organizational strategy, complex systems analysis	[32–34]
9	ISM (Interpretive Structural Modelling)	Structural modelling	Hierarchical decision processes, system thinking	Expert judgment	Decision analysis, policy making	[35–37]
10	Fuzzy cognitive mapping	Fuzzy logic, Causal maps	Soft systems methodology, fuzzy decision making	Expert-based knowledge	Environmental policy, social systems analysis	[38–40]

Table 2 Pros and cons of interdependency techniques

S. No	Technique	Strengths	Weaknesses	References
1	Bayesian network	Models' uncertainty effectively through probabilistic relationships Handles incomplete data and missing values well Intuitive representation of causal relationships	Computationally expensive for large, complex networks Requires expert knowledge to define causal structures Sensitive to model misspecification	[12–14]
2	Structural equation modelling	Effective for modelling complex relationships between latent and observed variables Allows for hypothesis testing and model fit evaluation	Requires large sample sizes Assumes linear relationships between variables Complexity increases with the number of latent variables	[15–17]
3	Directed Acyclic Graphs (DAGs)	Visualizes and clarifies causal assumptions Identifies confounders and mediators Supports causal inference without requiring large datasets	Assumes the correct causal structure is known Does not handle feedback loops (cyclic relationships) No inherent probabilistic inference	[8, 18, 19]
4	Granger causality analysis	Useful for time-series data Detects predictive relationships in economics and finance Simple to implement	Only applicable to time-series data Assumes stationarity and linearity in relationships Limited to short-term causality (no long-term effects)	[20–22]
5	Markov chain	Simple to understand and apply in sequential state transitions Useful in modelling processes with memoryless properties	Assumes Markov possessions (future state depends lone on current state) Can oversimplify complex systems Difficult to model real-world systems accurately	[23–25]
6	Graph Neural Network	Efficient for large-scale graph-based data Automatically learns relationships between nodes Suitable for high-dimensional, unstructured data	Requires large amounts of data for training Computationally intensive Complex to interpret and explain results compared to traditional models	[26–28]

(continued)

Table 2 (continued)

S. No	Technique	Strengths	Weaknesses	References
7	Causal Forest	Automatically detects treatment heterogeneity Non-parametric, making no assumptions about functional form Flexible with high-dimensional data	Difficult to interpret individual trees within the forest Sensitive to hyperparameter tuning Requires large amounts of training data	[29–31]
8	DEMATEL	Handles complex systems by identifying cause-and-effect relationships Visualizes feedback loops in systems Easy to implement and interpret	Subjective, heavily reliant on expert opinion for determining relationships Does not account for uncertainty or probabilistic relationships	[32–34]
9	ISM (Interpretive Structural Modelling)	Structured approach for modelling complex systems Hierarchically organizes relationships Effective for decision-making in policy and management	Highly reliant on expert judgment for building models Limited to qualitative analysis, lacking quantitative rigor Does not capture feedback loops	[35–37]
10	Fuzzy cognitive mapping	Captures uncertainty in complex systems through fuzzy logic Relaxed to visualize and understand Grips both qualitative and quantitative data	Relies heavily on expert knowledge for defining causal relationships Sensitive to subjective input Challenging to validate or calibrate models	[38–40]

by providing a visual representation of causal linkages. While Markov Chains are used to pretend sequential processes with memoryless features, Granger Causality Analysis is particularly helpful in time-series data for evaluating whether one variable can forecast another. In addition, this review examines more recent progressions in the field of causal inference, including machine learning-based approaches like Causal Forests and Graph Neural Networks (GNNs). More advanced techniques like Counterfactual Generative Models, Invariant Causal Prediction (ICP), Regression Discontinuity Design (RDD) [41, 42] and tools will possibly be used in causal inference in the future, refining our capacity to comprehend and analyse complex systems and ultimately facilitating smarter, data-driven decision-making.

References

1. Pearl, J.: The foundations of causal inference. Sociol. Methodol. **40**(1), 75–149 (2010)
2. Kuang, K., Li, L., Geng, Z., Xu, L., Zhang, K., Liao, B., Jiang, Z.: Causal inference. Engineering **6**(3), 253–263 (2020)
3. Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., Zhang, A.: A survey on causal inference. ACM Trans. Knowledge Discovery From Data (TKDD) **15**(5), 1–46 (2021)
4. Ohlsson, H., Kendler, K.S.: Applying causal inference methods in psychiatric epidemiology: a review. JAMA Psychiat. **77**(6), 637–644 (2020)
5. Costa, D.K., Yakusheva, O.: Why causal inference matters to nurses: the case of nurse staffing and patient outcomes. Online J. Issues Nursing **21**(2) (2016)
6. Dong, Z.: Wright's path analysis. Theoria: An Int. J. Theory History Foundat. Sci. **39**(1), 67–88 (2024)
7. Rubin, D.B.: Essential concepts of causal inference: a remarkable history and an intriguing future. Biostatist. Epidemiol. **3**(1), 140–155 (2019)
8. Moodie, E.E., Stephens, D.A.: Causal inference: critical developments, past and future. Canad. J. Statist. **50**(4), 1299–1320 (2022)
9. Brunet, T.D.: Local causation. Synthese **199**(3), 10885–10908 (2021)
10. Pearl, J.: Statistics and causal inference: a review. Test 12, 281–345 (2003)
11. Evans, T.S., Expert, P., Vasiliauskaite, V.: Cycle analysis of directed acyclic graphs. Physica A: Statist. Mechan. Appl. **596**, 127097 (2022)
12. Villa, S., Rossetti, M.: Learning continuous time Bayesian network classifiers using mapreduce. J. Statist. Softw. **62**, 1–25 (2015)
13. Ghosh, A., Ahmed, S., Khan, F., Rusli, R.: Process safety assessment considering multivariate non-linear dependence among process variables. Process Safety Environ. Protect. **135**, 70–80 (2020)
14. Hossain, N.U.I., El Amrani, S., Jaradat, R., Marufuzzaman, M., Buchanan, R., Rinaudo, C., Hamilton, M.: Modeling and assessing interdependencies between critical infrastructures using Bayesian network: a case study of inland waterway port and surrounding supply chain network. Reliab. Eng. Syst. Safety **198**, 106898 (2020)
15. Irfan, M., Wang, M., Akhtar, N.: Impact of IT capabilities on supply chain capabilities and organizational agility: a dynamic capability view. Operat. Managem. Res. **12**(3), 113–128 (2019)
16. Liao, S.H., Hu, D.C., Ding, L.W.: Assessing the influence of supply chain collaboration value innovation, supply chain capability and competitive advantage in Taiwan's networking communication industry. Int. J. Prod. Econ. **191**, 141–153 (2017)
17. Shukor, A.A.A., Newaz, M.S., Rahman, M.K., Taha, A.Z.: Supply chain integration and its impact on supply chain agility and organizational flexibility in manufacturing firms. Int. J. Emerg. Markets **16**(8), 1721–1744 (2021)
18. Chen, H., Chen, Y., Lin, B., Wang, Y.: Can short selling improve internal control? An empirical study based on the difference-in-differences model. Account. Finan. **58**(5), 1233–1259 (2019)
19. Lipsky, A.M., Greenland, S.: Causal directed acyclic graphs. JAMA **327**(11), 1083–1084 (2022)
20. van Esch, R.J., Shi, S., Bernas, A., Zinger, S., Aldenkamp, A.P., Van den Hof, P.M.: A Bayesian method for inference of effective connectivity in brain networks for detecting the Mozart effect. Comput. Biol. Med. **127**, 104055 (2020)
21. Hui, L., Huiling, L., Liwei, L., Jia, Z., Huaijun, R.: Causal relationship analysis of high-dimensional time series based on quantile factor model. Knowledge-Based Syst. **284**, 111263 (2024)
22. Vasile, V., Ştefan, D., Comes, C.A., Bunduchi, E., Ştefan, A.B.: Granger causality. Roman. J. Econ. Forecast. **23**(4), 131 (2020)
23. Sheu, J.B., Kundu, T.: Forecasting time-varying logistics distribution flows in the one belt-one road strategic context. Transport. Res. Part E: Logist. Transport. Rev. **117**, 5–22 (2018)
24. Yu, P., Wang, P., Wang, Z., Wang, J.: Supply chain risk diffusion model considering multi-factor influences under hypernetwork vision. Sustainability **14**(14), 8420 (2022)

25. Tafakkori, K., Jolai, F., Tavakkoli-Moghaddam, R.: Disruption-resilient supply chain entities with decentralized robust-stochastic capacity planning.7. Reliab. Eng. Syst Safety **238**, 10944 (2023)
26. Tao, J.: Optimization of dynamic graphic packaging design scheme based on graph neural network. Mobile Inform. Syst. 5075884 (2022)
27. Liu, F., Wang, J., Yang, J.: A graph neural network recommendation method integrating multi head attention mechanism and improved gated recurrent unit algorithm. IEEE Access (2023)
28. Kang, S.: K-nearest neighbor learning with graph neural networks. Mathematics **9**(8), 830 (2021)
29. Li, Z.F., Zhou, Q., Chen, M., Liu, Q.: The impact of COVID-19 on industry-related characteristics and risk contagion. Finan. Res. Lett. **39**, 101931 (2021)
30. Credit, K., Lehnert, M.: A structured comparison of causal machine learning methods to assess heterogeneous treatment effects in spatial data. J. Geograph. Syst. 1–28 (2023)
31. Wager, S., Athey, S.: Estimation and inference of heterogeneous treatment effects using random forests. J. Amer. Statist. Assoc. **113**(523), 1228–1242 (2018)
32. Patanjal, K., Kumar, M.S., Yigit, K., Ali, E.: A decision framework for incorporating the coordination and behavioural issues in sustainable supply chains in digital economy. Annals of Operat. Res. 721–749 (2023)
33. Fakhrzad, M.B., Firozpour, M.R., Nasab, H.H.: A fuzzy DEMATEL approach for evaluating the risk factors-(Case study: the instruction center of Iran's national oil company). RAIRO-Operat. Res. **55**(4), 2525–2543 (2021)
34. Chang, B., Chang, C.W., Wu, C.H.: Fuzzy DEMATEL method for developing supplier selection criteria. Expert Syst. Appl. **38**(3), 1850–1858 (2011)
35. Diabat, A., Govindan, K.: An analysis of the drivers affecting the implementation of green supply chain management. Resour. Conserv. Recycling **55**(6), 659–667 (2011)
36. Chowdhury, N.A., Ali, S.M., Mahtab, Z., Rahman, T., Kabir, G., Paul, S.K.: A structural model for investigating the driving and dependence power of supply chain risks in the readymade garment industry. J. Retail. Consum. Serv. **51**, 102–113 (2019)
37. Hashemi Petrudi, S.H., Ahmadi, H.B., Azareh, Y., Liou, J.J.: Developing a structural model for supply chain viability: a case from a developing country. Operat. Managem. Res. **17**(1), 324–339 (2024)
38. Reklitis, P., Sakas, D.P., Trivellas, P., Tsoulfas, G.T.: Performance implications of aligning supply chain practices with competitive advantage: empirical evidence from the agri-food sector. Sustainability **13**(16), 8734 (2021)
39. Shojaei, P., Haeri, S.A.S.: Development of supply chain risk management approaches for construction projects: a grounded theory approach. Comput. Indus. Eng. **128**, 837–850 (2019)
40. Dolatabad, A.H., Mahdiraji, H.A., Babgohari, A.Z., Garza-Reyes, J.A., Ai, A.: Analyzing the key performance indicators of circular supply chains by hybrid fuzzy cognitive mapping and fuzzy DEMATEL: evidence from healthcare sector. In: Environment, Development and Sustainability, pp. 1–27. (2022)
41. Louizos, C., Shalit, U., Mooij, J.M., Sontag, D., Zemel, R., Welling, M.: Causal effect inference with deep latent-variable models. In: Advances in Neural Information Processing Systems, pp. 30. (2017)
42. Imbens, G.W., Lemieux, T.: Regression discontinuity designs: a guide to practice. J. Econometrics **142**(2), 615–635 (2008)
43. Rinaldi, S.M., Peerenboom, J.P., Kelly, T.K.: Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Syst. Magazine **21**(6), 11–25 (2001)
44. Ouyang, M.: Review of modeling and simulating interdependent critical infrastructure systems. Reliab. Eng. Syst. Safety **121**, 43–60 (2014)
45. Aven, T., Zio, E.: Some considerations on the treatment of uncertainties in risk assessment for practical decision-making. Reliab. Eng. Syst. Safety **96**(1), 64–74 (2011)
46. Mottahedi, A., Sereshki, F., Ataei, M., Qarahasanlou, A.N., Barabadi, A.: Resilience estimation of critical infrastructure systems: application of expert judgment. Reliabil. Eng. Syst. Safety **215**, 107849 (2021)

47. Liakos, A., Pagkalidou, E., Karagiannis, T., Malandris, K., Avgerinos, I., Gigi, E., Tsapas, A.: A simple guide to randomized controlled trials. Int. J. Lower Extremity Wounds (2024)
48. Wang, A., Hu, S., Zhu, M., Wu, M.: Customer contagion effects of voluntary environmental regulation: a supplier green innovation perspective. Energy Econ. **132**, 107446 (2024)
49. Blanco-Encomienda, F.J., Rosillo-Díaz, E.: Quantitative evaluation of the production and trends in research applying the structural equation modelling method. Scientometrics **126**(2), 1599–1617 (2021)
50. Peters, J., Bühlmann, P., Meinshausen, N.: Causal inference by using invariant prediction: identification and confidence intervals. J. Royal Statist. Soc. Ser. B: Statist. Methodol. **78**(5), 947–1012 (2016)
51. Egami, N., Tchetgen Tchetgen, E.J. Identification and estimation of causal peer effects using double negative controls for unmeasured network confounding. J. Royal Statist. Soc. Ser. B: Statist. 02. **86**(2), 487–51 (2024)

Machine Learning-Driven Clustering Based Environmental, Social, and Governance Performance Prediction Model



Gorkem Sariyer, Mert Sarioglu, and Bharti Ramtiyal

Abstract This paper integrates machine learning (ML) to propose a cluster-based ESG performance prediction model. Using data from 1235 companies across 8 sectors, we first clustered companies using their environmental, social, and governance performances and predicted their grouping using financial parameters such weighted average cost of capital (WACC), systematic risk, and market risk premium. The k-means ++ algorithm was used to cluster companies, revealing notable differences in their ESG pillar scores. To predict ESG performance group, we employed Decision Tree, Random Forest, and XGBoost comparatively. Among these the best-performing algorithm, Random Forest, had 97.57% accuracy. In the next stage, feature engineering was applied to identify key financial parameters that influence ESG prediction, where weighted average cost of capital equity and market risk premium emerging as the most significant factors. These findings demonstrate the significance of financial parameters in predicting ESG performance and present implications for investors, analysts and companies to ensure alignment with sustainability and ESG goals. The proposed model demonstrates the effective integration of ML and financial data for ESG analysis and prediction.

Keywords ESG performance · Sustainability · Financial performance · Machine learning · Clustering

G. Sariyer (✉) · M. Sarioglu
Yasar University, Business Administration, Bornova, Türkiye
e-mail: gorkem.ataman@yasar.edu.tr

B. Ramtiyal
Faculty of Management and Commerce, Poornima University, Jaipur, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

M. Singh et al. (eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences*, Mechanisms and Machine Science 185,
https://doi.org/10.1007/978-3-031-95963-9_46

1 Introduction

A wide range of factors are critical to the financial valuation and performance of companies. These factors are known as financial parameters where they affect the decision-making processes of managers, investors, and shareholders. Although financial parameters are of great significance, sustainability related concerns reshape the business environment, and leads to a set of changes in the valuation of financial information. Therefore, the integration of non-financial factors, namely Environmental, Social, and Governance (ESG) performances of companies, which affects the investment decision [30]. On this basis, the combination of financial and ESG information is even demanded by investors through reporting practices of companies [4]. Moreover, while ESG performances are critical for meeting stakeholder expectations [11], it also contributes to corporate and managerial reputation [8, 10]. In short, ESG concerns have been regarded as an essential part of business activities, which is critical to companies, users of information, and humanity.

The changing dynamics of the business environment lead increasing number of companies to embrace ESG. At this point, it is extremely critical to realize the main purpose behind the ESG concept. Given humanity's common problems, such as global warming, environmental destruction and pollution, hunger and poverty, companies must ensure that their business practices are not only aligned with profitability, but also with people and planet [2, 3]. Therefore, environmental and social awareness should be located at the heart of governance practices, leading to the reduction of environmental and social problems while increasing profitability and value [3, 29]. Moreover, the evidence shows that financial outcomes are also centered around governance structure, corporate strategies, culture, and technology, which are directly part of ESG activities [36]. In a nutshell, addressing all these issues together is a determining factor of firms' sustainability performance. In such a context, measuring companies' ESG performance is a critical indicator reflecting how well ESG is integrated into business. One of the ways to measure ESG performance is to score companies' ESG activities. Accordingly, various third-party agencies namely Fitch, MSCI, S&P, LSEG, and Sustainalytics score ESG performance of companies based on its each pillar considering different methodological approach. In terms of the existing literature, these ESG scores are recognized to be among the key indicators of companies' ESG performance. [5, 39], where it affects firm's financial outlook [32, 41] . Therefore, companies having high ESG performance are expected to perform better in financial parameters. In other words, the degree of achievement of ESG goals set by companies and third-party organizations can be evaluated through ESG scores. It can thus shape the decision-making process of investors and shareholders, which in turn may affect financial parameters. In conclusion, ensuring satisfactory ESG practices benefits companies in the capital market, which in turn impacts investors' financial decisions and therefore leads to increased shareholder value through better corporate finance and management performance [19, 21]. Investor demand for sustainable companies, combined with strong ESG scores, means that the value of companies increases. Therefore, it provides the financial incentives for companies to promote

better ESG practices in the future. This suggests a significant association between ESG and financial parameters, from investors, shareholders, and companies viewpoint. However, there is an ongoing debate as to which financial parameters are more dominant and material in this relationship.

Considering the existing literature, ESG practices and financial parameters have been the main interest of different studies. These studies have mostly focused on well-known financial parameters. Accordingly, Vitale et al. [41] presented disclosing ESG information significantly and positively affects the OROA (Operating Return on Assets), ROS (Return on Sales), and the ROE (Return on Equity) ratios of the sampled international companies. A similar positive effect is also shown by Chininga et al. [15], in which ESG ratings are positively significant for firm performance when ROE and Tobin's Q ratios are considered in South African companies. Moreover, Han et al. [23] considered ESG scores to predict financial measures in Korean stock market. Another evidence presented by Naeem et al. [31] reveals positive and significant roles of ESG scores on Tobin'Q and ROA ratios in emerging countries. Likewise, high association between sustainability scores and ROA is documented by De Lucia et al. [28]. The high relation between ESG scores on Tobins'Q and ROA is similarly addressed by Aydoğmuş et al. [7] using BIST Sustainability indexed firms in Turkey. The study that is conducted by Arslan-Çetin et al. [6] considering the BIST Sustainability indexed companies in Turkey, indicates same results. Therefore, strong association is stated between sustainability scores and ROA, ROE, net profit margin, and asset growth rate. The relationship between ESG performance and financial performance on the FTSE350 indexed companies is analyzed in the study by Ahmad et al. [1]. While sustainability scores have positive effects on financial outcomes, some ESG components are found to affect financial results differently. Correspondingly, these metrics are recognized as the most widely discussed measures in researches to address the association between companies' ESG performance and financial outlook. However, as the literature review shows numerous studies reproduce the similar results through considering popular and frequently used metrics. Therefore, the potential impact of sustainability performance on financial performance metrics can easily be estimated, where the different points need to be emphasized to bring novelty to the literature considering the association between ESG and financial scoring of companies.

Given the existing literature, a remarkable point is that mentioned studies in the previous paragraphs are mostly based on firm value and financial performance metrics namely Tobins'Q, ROA and ROE. However, there are other financial parameters regarding the components of corporate finance such as cost of capital and systematic risk (beta) level of companies. Accordingly, ESG performance and score can be evaluated in the context of corporate finance, investment analysis, and valuation. On this basis, a company's ESG practice also affects its Weighted Average Cost of Capital (WACC). Therefore, firms having strong ESG performance are able to get capital in a cheaper way as they are considered less risky. The results of Priem and Gabellone [35] suggest that higher sustainability performance lower the cost of capital, particularly in countries having light legal systems. In these countries, social and environmental factors lower costs, but governance mechanisms can raise

borrowing costs by enabling more debt. The mentioned study is also sampled 600 firms that listed in STOXX Europe 600. Extending previous studies, Ernst and Woithe [20] find that firms in the S&P 500 index having strong sustainability performance have lower costs of debt and equity capital. Similarly, Chen et al. [13] conclude that ESG practice of companies in China lead to lower the cost of equity capital in both direct and indirect ways by mitigating market risk and improving the diversification of equity capital for companies. Considering the companies form India, along with the moderating role of independent assurance, ESG is related to financial performance and cost of debt [19]. Apart from cost of capital, systematic risk is also associated with ESG performance. In this context, Gupta and Aggarwal's [22] analysis based on ESG pillars indicate that the cost of capital is associated to only the governance pillar of ESG, while all ESG components are negatively associated with systematic risk. Accordingly, systematic risk (beta) assesses a company's market risk separately from its capital structure, where it is expected to decrease for companies with strong ESG performance because they are less vulnerable to external shocks and market volatility. The findings of existing studies also support this idea. According to Ciciretti et al. [17], higher ESG scores lead to lower expected returns, where it is probably a consequence of the low level of risk. In a similar way, improved ESG performance results in lowering systematic risk [33]. Strong ESG scoring can decrease the cost of capital, but investors can expect lower expected returns under this situation [18]. The study of Choi et al. [16] on market risk premium and ESG evidences that risk arises from ESG is systematic and priced. Therefore, it is expected to have an impact on expected return and market risk premium respectively. In short, the expected returns of companies with strong ESG performance or scores affects investors' decision and risk. Conversely, poor ESG score can be indicated as riskier, which may increase risk premium and thus its expected return.

Based on the existing literature, it is believed that many factors are influential in investors' decision-making, and ESG performance is not an extension. Putting all this together, instead of focusing on well-known financial performance measures such as Tobins' Q, ROA and ROE, this paper focuses on ESG performance along with cost of capital structure, systematic risk and market risk premium. Previous studies have somehow demonstrated the effects of WACC, systematic risk, expected return and market risk premium on ESG performance. However, their role in predicting groups based on ESG scores across various companies is unknown. Additionally, it is not clear whether the determined financial parameters can be used to predict the grouping of companies according to their ESG scores. Therefore, we aim to answer;

(1) How are the companies homogenously grouped based on their ESG pillar scores?
(2) Can we predict the companies' ESG score grouping efficiently using only their financial parameters?

To answer these research questions, we take advantages of Machine Learning (ML) over traditional methods. ML offers significant benefits in clustering ESG-based performance and predicting companies' ESG score groupings using financial parameters. First, we cluster companies according to their ESG performance using

k-means ++ . In the next step, we estimate the ESG score grouping of companies using the determined financial parameters incorporating tree-based machine learning algorithms. Lastly, the significance of financial parameters are evaluated through Random Forest Regressor. To do so, we analyzed 1235 companies as a final sample size from both emerging and developed markets. The results are also believed to have a number of implications. The proposed ML techniques enable faster and more accurate clustering, which in turn facilitates companies to predict their ESG performance using financial parameters. The proposed techniques can be applied to different subjects. The results also demonstrate implications for decision-makers, investors, shareholders and companies.

2 Methodology

2.1 Data Set

This study uses ML techniques to propose a cluster-based ESG performance prediction model. To reach results for this purpose, our sample contains of 1235 companies over a one-year period in 2020. In terms of sample, we use the companies having full data set on ESG scores and financial information to draw a conclusion. In this context, Fitch is one of the independent organizations that provides different financial and non-financial ratings for companies, countries and markets. They also measure ESG performance of companies regarding its own methodology. Therefore, this study considers companies whose ESG scores are measured by Fitch. In the first step, the data of 1235 companies from emerging and developed markets under 8 main sectors that meet these conditions were accessed. In the next step of data collection process, the financial parameters of sampled companies regarding WACC, WACC Debt, WACC Equity, Systematic Risk (Unlevered Beta), and Market Risk Premium were accessed and collected on the database of Bloomberg. Table 1 shows the characteristic of our sample.

This study also utilizes various variables regarding the ESG score and financial parameters, which are accessed through the information provided by Fitch and Bloomberg. In addition to ESG and determined financial parameters, we considered credit rating impact, industry, market, and region to robust our results. These variables are addressed in the Table 2.

2.2 Used Machine Learning Algorithms

We combined clustering and classification functions of machine learning in this paper. Clustering is defined as putting similar observations into same groups in such a way that points in same clusters are more alike than those in others.

Table 1 Sample distribution

Industry	Developed	Emerging
Food and beverage	37	15
Generic	47	39
Industrials and transportation	92	25
Natural resources	128	105
Real estates and construction	105	86
Retail, consumer, and healthcare	81	20
Services and communication	124	40
Utilities	198	93
Total	**812**	**423**
Final sample size	**1235**	

We used the kmeans ++ algorithm to cluster the firmsbased on their E, S, and G scores. We performed this algorithm in the Scikit-learn tool in Python. After grouping the companies according to their sustainability indexes, we used the classification function of ML to predict the firm's sustainability performance grouping given a set of its financial parameters. For classification, we used tree-based algorithms of Decision Tree (DT), Random Forest (RF) and Extreme Gradient Boosting (XGBoost).

DT uses the splitting rule given the input parameters by forming a tree structure. Input variables are continuously divided to enhance homogeneity in a node until it reaches a level where further partitioning does not enhance the homogeneity [9]. The end node presents the value of the target variable as a class or level in classification models.

RF algorithm is an extension of decision trees, where many trees are trained, and the target of the forest is obtained based on the outcomes of individual trees. This prevents the overfitting problem that may be seen for unique trees [24]. Diversity in decision trees can be satisfied by limiting each tree to be trained from a random subset of input variables [25].

XGBoost presents advanced boosting methodology developed by Chen and Guestrin [14]. XGBoost comprises many forests [42]. While this algorithm uses a second-order Taylor expansion to the loss function, it uses a normalization in the objective function [27].

We partitioned the data set into two as train and test sets to measure the classification power of the applied ML techniques. We randomly selected 80% of the observations for algortihmic learning and evaluated the performances of these algorithms in the remaining 20% of data points. We calculated precision, recall, accuracy and F1 statistics to compare the classification performances.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

Table 2 Identification of variables

Variable	Notation	Scale of measurement	Measure
Environmental pillar score	E	Ratio scale	Environmental pillar score calculated and provided by fitch
Social pillar score	S	Ratio scale	Social pillar score calculated and provided by fitch
Governance pillar score	G	Ratio scale	Governance pillar score calculated and provided by fitch
Total ESG score	ESG	Ratio scale	Sum of E, S, and G pillar
Weighted average cost of capital	WACC	Ratio scale	WACC calculated and provided by Bloomberg
Weighted average cost of capital debt	WACC_Cost_ Debt	Ratio scale	WACC cost of debt calculated and provided by Bloomberg
Weighted average cost of capital equity	WACC_Cost_Equity	Ratio scale	WACC cost of equity calculated and provided by Bloomberg
Systematic risk	Unlevered_Beta	Ratio scale	Unlevered beta calculated and provided by Bloomberg
Market risk premium	Risk_Premium	Ratio scale	market risk premium calculated and provided by Bloomberg
Credit rating impact	Credit rating impact	Nominal scale	credit rating impact of companies
Industry	Sector	Nominal scale	Industry of sampled companies
Market	Market	Nominal scale	Emerging or developed market
Region of company	Region	Nominal scale	America, Asia, or Europe

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$F1 = 2^{*}\frac{precision^{*}recall}{precision + recall} \tag{4}$$

We finally used feature engineering to discuss the importance of the identified financial parameters in predicting companies' ESG clusters. We obtained the importance of the financial parameters on ESG performance grouping by implementing the Random Forest Regressor.

2.3 Proposed Model

We proposed a four-stage model in this research. In the first stage of our model, we collected and processed the study data set. In the second stage, to homogenously group the study firms based on their E, S, and G scores, we used clustering. Next, to efficiently forecast the obtained ESG grouping of the companies given their financial performances, we applied tree-based classification algorithms. As a final stage, to discuss the significance of the financial parameters in predicting companies' ESG performance grouping, we used feature engineering. In Fig. 1, we present the step-by-step methodology of our model.

3 Results

For different values of number of clusters, k, we calculated the within cluster sum squared differences and showed in Fig. 2.

According to Fig. 2, we selected $k = 4$ and grouped the study companies into 4 clusters. Table 3 shows the average scores of E, S, and G pillars of the companies assigned to each cluster.

Although average E, S, and G, and overall ESG performances of the four clusters were close to each other, we still observed significant differences. For example, we observed a difference of 8.06% between the scores of the two clusters (clusters 1 and 3) with the highest and lowest average E column scores. We also observed that since the S and G scores of the firms were very close to each other, the difference between the clusters was less in these pillars. But we still calculated a difference of 3.31% for pillar S (clusters 2 and 0) and 1.83% for the pillar G (clusters 1 and 3). In Table 3, we also analyzed that 146 of the companies had the highest average E, G, and ESG score. On the other hand, 161 of the companies had the lowest E, G, and ESG score. Companies grouped in cluster 2 had also low scores on sustainability indexes. In light of these results, we found that there exists an association between the clustering obtained and the sustainability performances of the firms and that one can make inferences using them.

Since the cluster number contains information regarding the ESG performance of a company, we then aimed to predict this cluster number given the financial performance of a company to answer RQ2. Using the identified financial parameters as the model input variables and the cluster number as the model output, and by implementing DT, RF, and XGBoost algorithms, we compared the prediction performances

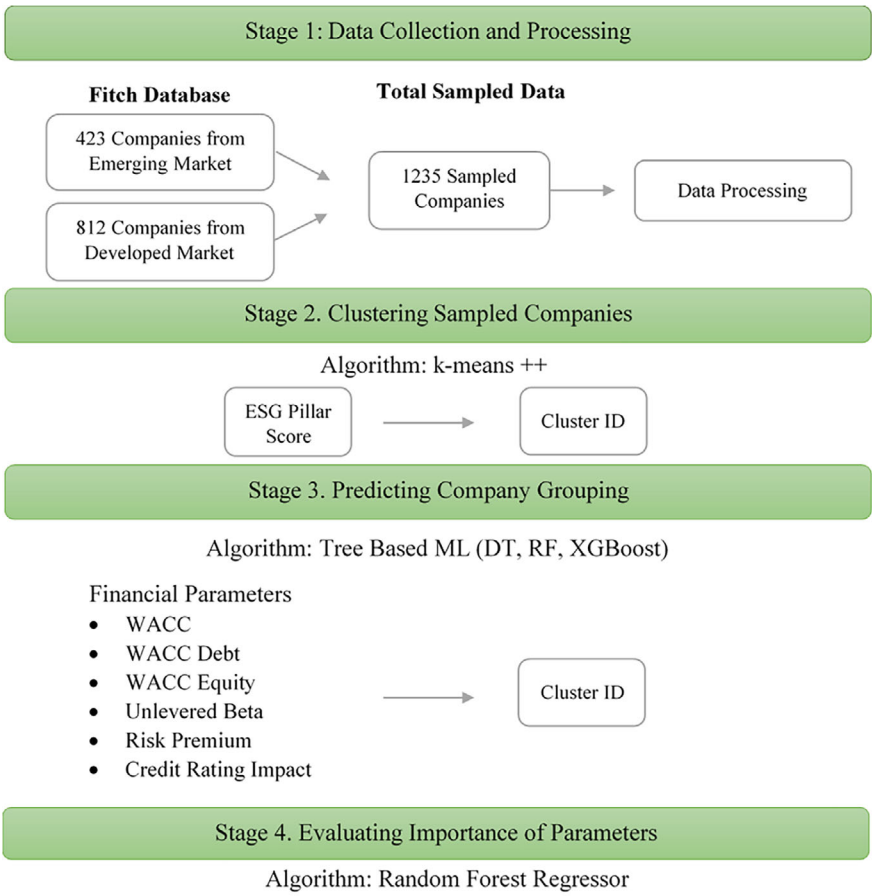


Fig. 1 Flowchart of the research model

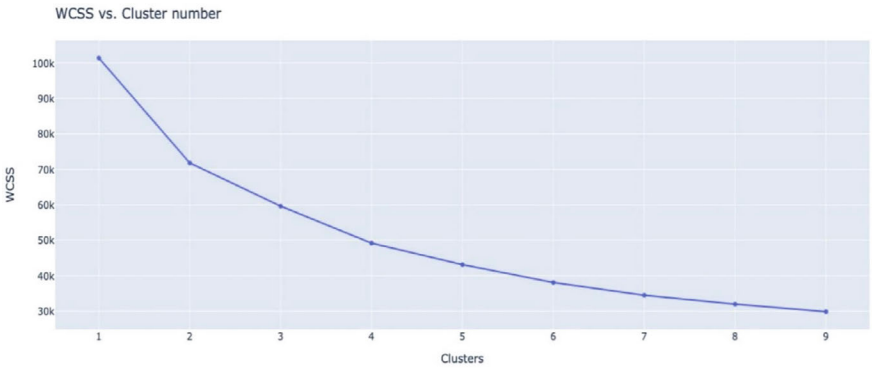


Fig. 2 WCSS values for different k

Table 3 Average E, S, G scores of companies for each cluster

Cluster id	Number of companies	Average E score	Average S score	Average G score	Average ESG score
0	587 (47.5%)	11.31	11.85	12.05	35.21
1	146 (11.8%)	11.80	11.71	12.22	35.73
2	341 (27.6%)	11.56	11.47	12.02	35.04
3	161 (13.0%)	10.92	11.81	12.00	34.73

of these algorithms. Figure 3 presents the average performance of these algorithms comparatively.

In Fig. 3 we observed that the best-performing algorithm is the RF with a 97.57% accuracy. Thus, we selected this algorithm for predicting the ESG cluster of a company in our model. In Fig. 4, we also presented a detailed analysis of the RF algorithm's performance.

In Fig. 4, we saw that for the randomly chosen 247 companies, the RF algorithm had an almost perfect classification performance for each of the clusters. All of the 29 firms assigned to the cluster 2 were also predicted in this cluster. Only 1 out of 124 companies was incorrectly classified in cluster 3, although it was actually in cluster 0. Similarly, only 1 out of 31 companies was classified in cluster 0, although it was grouped in cluster 2. The performance of the RF algorithm was a bit lower for predicting the companies of cluster 3. 4 out of 63 companies were predicted in cluster 0, although they were actually in cluster 3. Thus, we observed that not only the average accuracy statistic, but also precision, recall and F1 statistics of the RF algorithm were high for all of the four clusters.

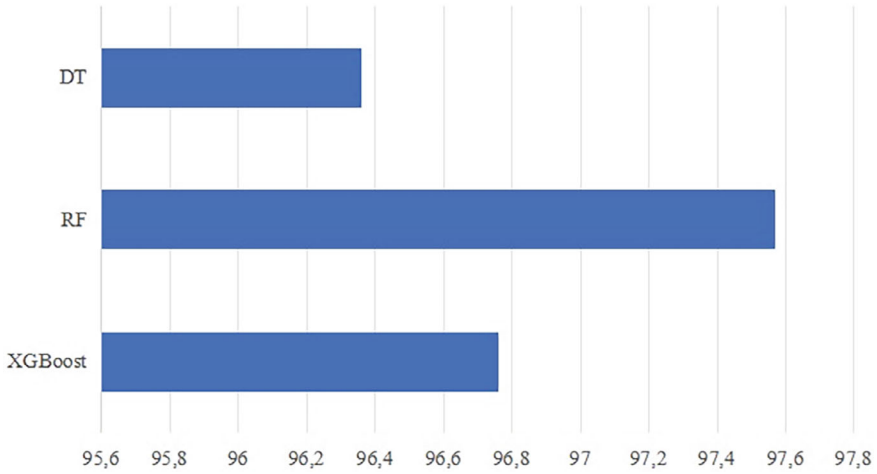
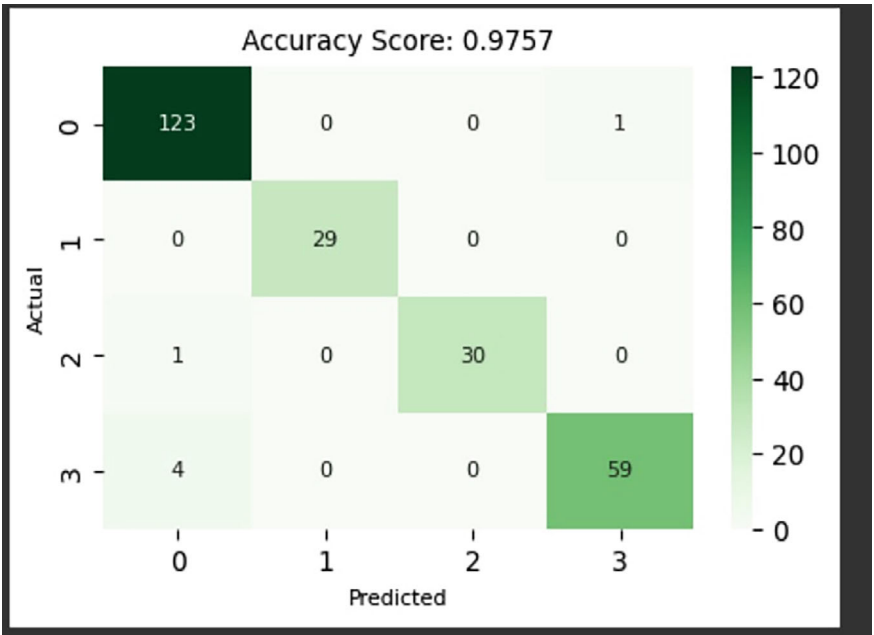


Fig. 3 Accuracy comparison of DT, RF, and XGBoost



	precision	recall	f1-score	support
0	0.96	0.99	0.98	124
1	1.00	1.00	1.00	29
2	1.00	0.97	0.98	31
3	0.98	0.94	0.96	63
accuracy			0.98	247
macro avg	0.99	0.97	0.98	247
weighted avg	0.98	0.98	0.98	247

Fig. 4 Performance of RF algorithm in the test data set

Finally, we aimed to analyze the importance of the identified financial parameters in the ESG group prediction model. We used feature engineering, and presented the ranked ordered importance of the parameters in Fig. 5.

In Fig. 5, we observed that WACC_COST_EQUITY had the highest importance in predicting the ESG performance cluster of a company. Besides, RISK_PREMIUM had the second highest relative importance. While WACC_COST_DEBT and WACC had a similar and high significance in ESG performance cluster prediction, the other identified variables were not significant in the prediction model.

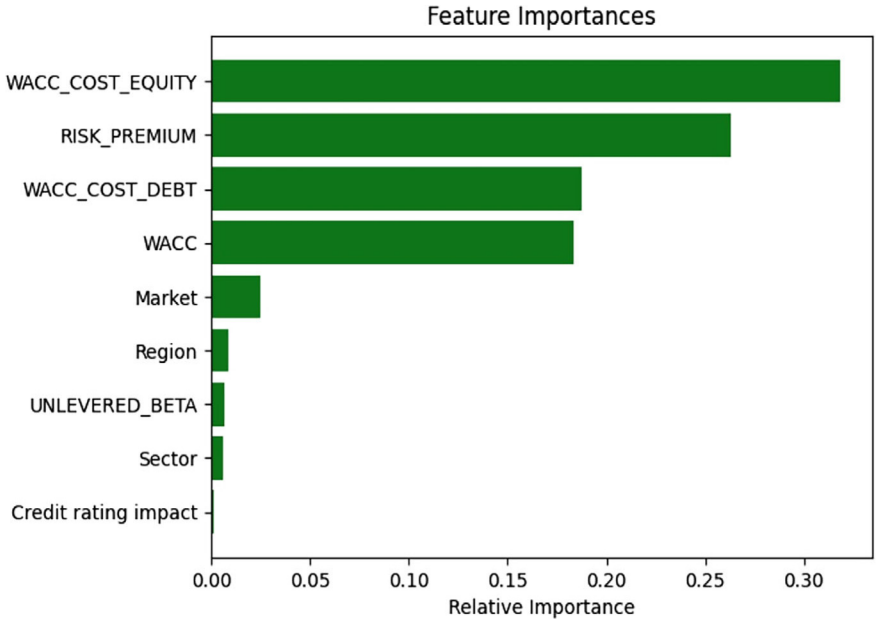


Fig. 5 Significance of the financial parameters

4 Conclusion

In this paper, we used data from 1235 companies in 8 different industries to propose a cluster-based ESG performance prediction model. Therefore, our goal was to group companies based on their ESG pillar score, and then predict their ESG group regarding determined financial parameters. Clustering technique was therefore applied to answer the first research questions. The results show that the ESG pillar scoring for the four clusters varied significantly. Considering the results, differences were noticeable, especially in the Environmental pillar, even though the average ESG scores of groups were comparatively similar. This demonstrates how the E pillar score varied more than the S and G pillars, having a greater impact on the overall ESG grouping. Therefore, the findings of this paper are significant to highlight how pivotal role is taken by the environmental pillar of ESG, particularly when considering a wide range of businesses from both developed and emerging markets under different industries.

We found that financial parameters, including Weighted Average Cost of Capital (WACC), WACC Debt, WACC Equity, and Market Risk Premium, were essential in forecasting ESG groupings in order to answer the second research question. Having 97.57% accuracy, the Random Forest (RF) algorithm was the most effective and accurate model for predicting ESG cluster. The model's efficacy was further confirmed by the high precision, recall, and F1 scores, in which it was presented across all four clusters. Accordingly, one of the findings of this study is that financial parameters

related to capital structure and systematic risk level of companies have an impact on ESG-based clustering. Within different parameters, WACC equity and market risk premium are the indicators that most impact firm's ESG performance grouping. This information is believed to be beneficial for investors, shareholders, financial analysts, and corporate strategists. Investors may recognize a company as riskier if its WACC equity is higher, and this perception may be connected to weaker ESG scoring. Businesses with a higher cost of equity may struggle to raise money for ESG goals or may put immediate financial gain over long-term social or environmental goals. In addition, market risk premium informs about the overall market volatility and risk. Therefore, companies in sectors or markets subject to a higher risk premium may struggle to meet ESG targets due to financial pressures or external uncertainties. In this regard, it gives decision makers the chance to evaluate environmental, social, and governance performance and financial parameters. This could result in better decision-making when choosing companies that prioritize sustainability objectives or investment portfolios with an emphasis on ESG.

The implications of this paper are twofold. Initially, the findings emphasize how crucial it is to combine financial and non-financial metrics when assessing the overall power of businesses for ESG sensitive decision-makers. Therefore, in assessing the cost of capital or systematic risk, these metrics can be used as proxies for a company's adhesion to ESG rules. Secondly, this study highlights the power of ML techniques in and decision making, resulting in improved accuracy along with time and cost efficiency. By leveraging these techniques, investors, shareholders and various stakeholders make more robust predictions about companies' future ESG and financial trends. Additionally, it is possible to apply proposed techniques in different areas in further studies.

In conclusion, this study contributes to ESG and corporate finance literature through proposing ML techniques, which provides a novel approach integrating financial and non-financial data for making decisions. Accordingly, the proposed techniques lead to contribute investment decisions, and policy makers as well as corporate sustainability strategies in point of fostering business practices.

References

1. Ahmad, N., Mobarek, A., Roni, N.N.: Revisiting the impact of ESG on financial performance of FTSE350 UK firms: static and dynamic panel data analysis. Cogent Busin. Managem. **8**(1), 1900500 (2021). https://doi.org/10.1080/23311975.2021.1900500
2. Al Amosh, H., Khatib, S.F.A.: Ownership structure and environmental, social and governance performance disclosure: the moderating role of the board independence. J. Business Soc.-Econ. Developm. **2**(1), 49–66 (2022). https://doi.org/10.1108/JBSED-07-2021-0094
3. Alduais, F.: Unravelling the intertwined nexus of firm performance, ESG practices, and capital cost in the Chinese business landscape. Cogent Econ. Finan. **11**(2), 2254589 (2023). https://doi.org/10.1080/23322039.2023.2254589
4. Alles, M., Palas, R., Vasarhelyi, M.A., Xiong, F.: The case for an app-based financial reporting system. J. Emerg. Technol. Account. **19**(2), 1–21 (2022). https://doi.org/10.2308/jeta-10827

5. Arouri, M., Gomes, M., Pukthuanthong, K.: Corporate social responsibility and M&A uncertainty. J. Corporate Finan. **56**, 176–198 (2019). https://doi.org/10.1016/j.jcorpfin.2019.02.002
6. Aslan Çetin, F., Öztürk, S., Akarsu, O.N.: The effect of ESG data of companies on financial performance: a panel data analysis on the BIST sustainability index. Sosyoekonomi **32**(61), 125–146 (2024). https://doi.org/10.17233/sosyoekonomi.2024.03.07
7. Aydoğmuş, M., Gülay, G., Ergun, K.: Impact of ESG performance on firm value and profitability. Borsa Istanbul Rev. **22**(2), S119–S127 (2022). https://doi.org/10.1016/j.bir.2022.11.006
8. Bifulco, G.M., Savio, R., Paolone, F., Tiscini, R.: The CSR committee as moderator for the ESG score and market value. Corporate Soc. Respons. Environ. Managem. **30**(6), 3231–3241 (2023). https://doi.org/10.1002/csr.2549
9. Breiman, L., Friedman, J., Olshen, R., Stone, C.:. Cart. Classification and regression trees (1984)
10. Broadstock, D.C., Matousek, R., Meyer, M., Tzeremes, N.G.: Does corporate social responsibility impact firms' innovation capacity? The indirect link between environmental and social governance implementation and innovation performance. J. Business Res. **119**, 99–110 (2020). https://doi.org/10.1016/j.jbusres.2019.07.014
11. Camilleri, M.A.: Environmental, social and governance disclosuresin Europe. Sustain. Account. Managem. Policy J. **6**(2), 224–242 (2015). https://doi.org/10.1108/SAMPJ-10-2014-0065
12. Celebi, M.E., Kingravi, H.A., Vela, P.A.: A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst. Appl. **40**(1), 200–210 (2013)
13. Chen, Y., Li, T., Zeng, Q., Zhu, B.: Effect of ESG performance on the cost of equity capital: evidence from China. Int. Rev. Econ. Finan. **83**, 348–364 (2023). https://doi.org/10.1016/j.iref.2022.09.001
14. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August, pp. 785–794. (2016)
15. Chininga, E., Alhassan, A.L., Zeka, B.: ESG ratings and corporate financial performance in South Africa. J. Account. Emerg. Econ. **14**(3), 692–713 (2023). https://doi.org/10.1108/jaee-03-2023-0072
16. Choi, D., Gam, Y.K., Kim, Y.H., Muravyev, D., Shin, H.: Market Risk Premium and ESG Risk (2024). Available at SSRN: https://doi.org/10.2139/ssrn.4762314
17. Ciciretti, R., Dalò, A., Dam, L.: The contributions of betas versus characteristics to the ESG premium. J. Empirical Finan. **71**, 104–124 (2023). https://doi.org/10.1016/j.jempfin.2023.01.004
18. Cornell, B.: ESG preferences, risk and return. Europ. Finan. Managem. **27**(1), 12–19 (2021). https://doi.org/10.1111/eufm.12295
19. Darsono, D., Ratmono, D., Tujori, A., Clarisa, T.Y.: The relationship between ESG, financial performance, and cost of debt: the role of independent assurance. Cogent Business Managem. **12**(1) (2024). https://doi.org/10.1080/23311975.2024.2437137
20. Ernst, D., Woithe, F.: Impact of the environmental, social, and governance rating on the cost of capital: evidence from the S&P 500. J. Risk Finan. Managem. **17**(3), 91 (2024). https://doi.org/10.3390/jrfm17030091
21. Feng, Z., Wu, Z.: ESG disclosure, REIT debt financing and firm value. J. Real Estate Finan. Econ. **67**(3), 388–422 (2023). https://doi.org/10.1007/s11146-021-09857-x
22. Gupta, S., Aggarwal, D.: Shrinking the capital costs and beta risk impediments through ESG: study of an emerging market. Asian Rev. Account. **32**(2), 249–277 (2024). https://doi.org/10.1108/ARA-05-2023-0130
23. Han, J.J., Kim, H.J., Yu, J.: Empirical study on the relationship between corporate social responsibility and financial performance in Korea. Asian J. Sustain. Soc. Responsibil. **1**, 61–76 (2016). https://doi.org/10.1186/s41180-016-0002-3

24. Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: In: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2, pp. 1–758. New York, Springer (2009)
25. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8): 832-844 (1998)
26. Huang, T., Zhang, Q., Wang, Z., Chen, Y., Wang, W.: Research on automatic recognition system of abnormal behavior of big data technology distribution network. In: 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), pp. 1084–1088. IEEE (2022)
27. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25 (2012)
28. De Lucia, C., Pazienza, P., Bartlett, M.: Does good ESG lead to better financial performances by firms? Machine learning and logistic regression models of public enterprises in Europe. Sustainability **12**(13), 5317 (2020). https://doi.org/10.3390/su12135317
29. Malik, N., Kashiramka, S.: Impact of ESG disclosure on firm performance and cost of debt: empirical evidence from India. J. Cleaner Prod. **448**, 141582 (2024). https://doi.org/10.1016/j.jclepro.2024.141582
30. Migliorelli, M.: What do we mean by sustainable finance? Assessing existing frameworks and policy risks. Sustainability **13**, 975 (2021). https://doi.org/10.3390/su13020975
31. Naeem, M.A.., Conlon, T., Cotter, J.: Green bonds and other assets: evidence from extreme risk transmission. J. Environ. Manag. **305**, 114358 (2022)
32. Narula, R., Rao, P., Kumar, S., Matta, R.: ESG scores and firm performance: evidence from emerging markets. Int. Rev. Econ. Finan. **89** (2024). https://doi.org/10.1016/j.iref.2023.08.024
33. Pistolesi, F., Teti, E.: Shedding light on the relationship between ESG ratings and systematic risk. Finan. Res. Lett. **60**, 104882 (2024). https://doi.org/10.1016/j.frl.2023.104882
34. Plotkin, S.S.: Generalization of distance to higher dimensional objects. Proc National Acad. Sci. **104**(38), 14899–14904 (2007)
35. Priem, R., Gabellone, A.: The impact of a firm's ESG score on its cost of capital: can a high ESG score serve as a substitute for a weaker legal environment. Sustain. Account. Managem. Policy J. **15**(3), 676–703 (2024). https://doi.org/10.1108/SAMPJ-05-2023-0254
36. dos Reis Cardillo, M.A., Basso, L.F.C.: Revisiting knowledge on ESG/CSR and financial performance: a bibliometric and systematic review of moderating variables. J. Innov. Knowledge **10**(1), 100648 (2025). https://doi.org/10.1016/j.jik.2024.100648
37. Sariyer, G., Mangla, S.K., Kazancoglu, Y., Xu, L., Tasar, C.O.: Predicting cost of defects for segmented products and customers using ensemble learning. Comput. Indus. Eng. **171**, 108502 (2022)
38. Sariyer, G., Taşkın, D.: Clustering of firms based on environmental, social, and governance ratings: evidence from BIST sustainability index. Borsa Istanbul Rev. **22**, S180–S188 (2022)
39. Sun, Y.: The real effect of innovation in environmental, social, and governance (ESG) disclosures on ESG performance: an integrated reporting perspective. J. Cleaner Prod. **460**, 142592 (2024). https://doi.org/10.1016/j.jclepro.2024.142592
40. Vassilvitskii, S., Artur, A.D.: The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, vol. 8090, pp. 778. (2006). Available online at: http://ilpubs.stanford.edu
41. Vitale, G., Cupertino, S., Riccaboni, A.: The effects of mandatory non-financial reporting on financial performance: a multidimensional investigation on global agri-food companies. British Food J. **125**(13), 99–124 (2023). https://doi.org/10.1108/bfj-06-2022-0545
42. Xu, Y., Zhao, X., Chen, Y., Yang, Z.: Research on a mixed gas classification algorithm based on extreme random tree. Appl. Sci. **9**(9), 1728 (2019)