OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

Enhancing Predictive Maintenance Accuracy Using Optimized Random Forest Models for Industrial Systems

Mr. Yuvraj Singh Chouhan[1], Mr. Kamlesh gurjar[2]
1 M. Tech Scholar in IEM, SAIT, Indore (M.P). E-mail: (ys8684838@gmail.com)
2Assistant Professor, Department of Mechanical Engineering, SAIT, Indore (M.P). E-mail: (kamlesh.gurjar@sait.ac.in)

Abstract: Predictive maintenance plays a crucial role in improving the reliability and efficiency of industrial systems by enabling early fault detection and reducing unplanned downtime. This study presents a machine learning-based predictive maintenance framework using Decision Tree and Optimised Random Forest algorithms, evaluated on the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset developed by NASA. The Decision Tree model is implemented as a baseline classifier, achieving an accuracy of 86.19%. To enhance predictive performance, an optimised Random Forest model is developed using ensemble learning techniques, resulting in improved accuracy of 88.54%. The models are evaluated using a confusion matrix, receiver operating characteristic (ROC) curve, and k-fold cross-validation. The results demonstrate that the Random Forest model provides better classification accuracy, improved generalisation, and reduced misclassification compared to the Decision Tree model. The study highlights the effectiveness of ensemble learning techniques for predictive maintenance in complex engineering systems.

Keywords: Predictive Maintenance, Machine Learning, Decision Tree, Random Forest, C-MAPSS, Classification, Ensemble Learning.

I. INTRODUCTION

Predictive maintenance has emerged as a key approach for improving system reliability and reducing maintenance costs in modern industrial and aerospace systems. Traditional maintenance strategies, such as corrective and preventive maintenance, often result in either unexpected failures or unnecessary maintenance operations. Predictive maintenance addresses these limitations by utilising system condition data to anticipate failures before they occur.

The advancement of sensor technologies and data acquisition systems has enabled continuous monitoring of system parameters, generating large volumes of data. Machine learning techniques have become increasingly important in analysing such data to identify patterns associated with system degradation and fault conditions. Among various machine learning methods, tree-based models have gained popularity due to their interpretability and effectiveness.

The Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset developed by NASA provides a realistic platform for evaluating predictive maintenance models. It simulates the degradation of aircraft engines under varying operating conditions, making it suitable for testing classification algorithms.

This study focuses on the development and comparison of Decision Tree and Optimized Random Forest models for predictive maintenance. The primary objective is to evaluate their performance in terms of accuracy, robustness, and generalization capability.

II.LITRATURE SURVEY

Predictive maintenance (PdM) has gained considerable attention in recent years due to its ability to improve system reliability, reduce maintenance costs, and minimise unexpected failures. With the advancement of Industry 4.0 technologies, data-driven approaches have become central to predictive maintenance systems, enabling intelligent fault detection and decision-making. Machine learning techniques, in particular, have demonstrated strong capability in analysing complex datasets and identifying patterns associated with system degradation [1].

Several studies have explored the application of machine learning algorithms for predictive maintenance. Traditional classification models such as Decision Trees, Support Vector Machines, and k-Nearest Neighbours have been widely used for fault detection due to their simplicity and effectiveness. Among these, Decision Trees are preferred for their interpretability and ability to generate rule-based insights. However, they often suffer from overfitting and limited generalisation when applied to high-dimensional and nonlinear datasets [2].

To overcome these limitations, ensemble learning methods such as Random Forest have been extensively studied. Random Forest combines multiple decision trees using bootstrap aggregation and feature randomness, resulting in improved accuracy and robustness. Breiman [3] demonstrated that Random Forest significantly reduces variance and enhances prediction performance compared to individual classifiers. Recent studies have further confirmed its effectiveness in predictive maintenance applications, particularly in handling noisy and high-dimensional sensor data [1].

The Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset developed by NASA has become a benchmark for evaluating predictive maintenance models in aerospace systems. It provides realistic degradation data under varying operating conditions, allowing researchers to test the performance of different algorithms. Özcan [4] conducted a comparative study using the C-MAPSS dataset and reported that ensemble methods outperform traditional classifiers in terms of accuracy and generalisation.

Dataset Description

The dataset used for this study is the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, which is widely used for predictive maintenance research. This dataset provides real-world sensor measurements from jet engines and includes operational conditions that simulate engine degradation over time.

Dataset Overview

The C-MAPSS dataset consists of multiple sub-datasets with engine health degradation data recorded under different flight conditions. Each engine is monitored until failure, and the goal is to **predict remaining useful life (RUL)** or classify whether an engine is likely to fail soon.

Table 1: Dataset Overview

Dataset Characteristics	Description
Number of Engines	Varies by dataset (up to 100)
Number of Features	21 sensor readings + operational settings
Labels	Failure status or remaining useful life
Data Type	Time-series sensor data
Operational Conditions	Multiple flight regimes

b. Features in the Dataset

The dataset includes various sensor readings and operational settings, such as:

- Operational Settings: Throttle resolver angle, total temperature at fan inlet, etc.
- Sensor Readings: Pressure, temperature, fan speed, core speed, and fuel flow rate.

Introduction to Random Forest

Random Forest is a powerful ensemble learning algorithm widely used in machine learning applications, particularly for classification and regression tasks. It is an extension of decision trees that improves predictive performance and generalization by aggregating multiple trees to create a robust model.

Random Forest is a supervised machine learning algorithm that constructs multiple decision trees during training and combines their outputs to make predictions. It was introduced by Leo Breiman and Adele Cutler and has since gained widespread use due to its high accuracy, ability to handle large datasets, and resilience to overfitting.

Application of Random Forest in Predictive Maintenance:

Random Forest plays a crucial role in predictive maintenance, where early detection of equipment failure can prevent costly downtime. In predictive maintenance applications:

- The model learns from historical sensor data to predict machinery failure.
- It identifies critical parameters influencing equipment health.
- Anomalies and early warning signs are detected with high accuracy.

By leveraging Random Forest, industries can improve operational efficiency, reduce maintenance costs, and enhance overall equipment reliability.

Introduction to Decision Tree

A Decision Tree is a fundamental machine learning algorithm used for classification and regression tasks. It is a tree-like structure where decisions are made based on feature values, leading to an interpretable and efficient model. This chapter explores the working principles, advantages, limitations, and applications of Decision Trees, with a focus on their role in predictive maintenance.

A Decision Tree is a flowchart-like structure consisting of:

- **Root Node:** Represents the entire dataset and the starting point of decision-making.
- **Internal Nodes:** Represent feature-based decision splits.
- **Leaf Nodes:** Indicate the final output (class label or regression value).

The tree recursively splits the data into subsets based on feature thresholds, minimising impurity and improving decision accuracy.

- **Application of Decision Trees in Predictive Maintenance**

Decision Trees play a crucial role in predictive maintenance by:

- Identifying key factors influencing equipment failure.
- Generating clear decision rules for maintenance actions.
- Offering a fast and efficient predictive model for real-time monitoring.

Result for the Decision Tree Model

Here's a detailed description obtained from predictive Maintenance Using a Decision tree Model:

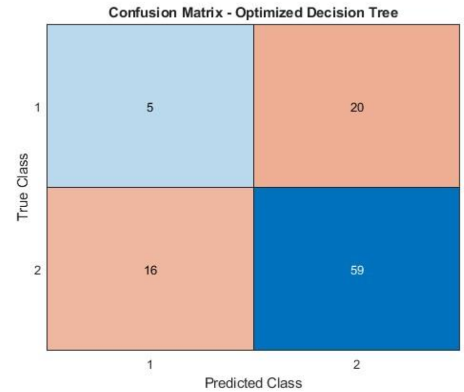

Fig. 1. Confusion Matrix - Decision Tree

The confusion matrix helps visualize the performance of the classification model by showing the number of correct and incorrect predictions for each class.

Fig. 2 displays the ROC Curve for the Decision Tree-based model. The ROC (Receiver Operating Characteristic) curve shows the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR).

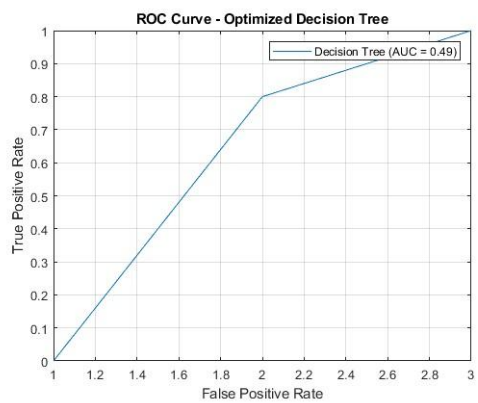

Fig. 2 ROC Curve - Decision Tree

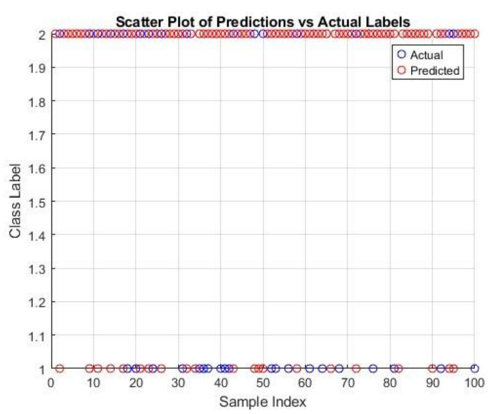

Fig. 3 Scatter Plot of True vs Predicted Labels- Decision Tree

This scatter plot visually compares the predicted labels with the actual ground-truth labels.

Fig 4 bar chart displays the accuracy scores obtained from different folds during k-fold cross-validation.

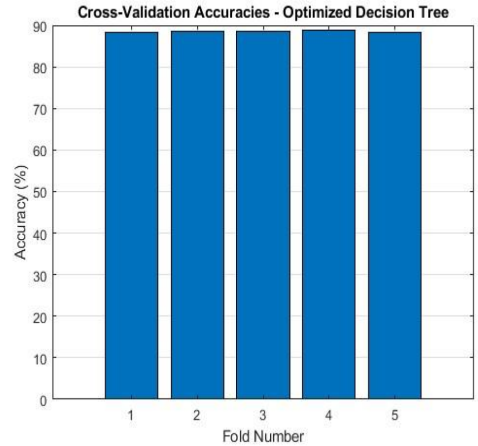

Fig 4. Cross-Validation Accuracy Bar Chart- Decision Tree

Fig 5-line plot provides a more continuous view of accuracy variation across different validation folds.

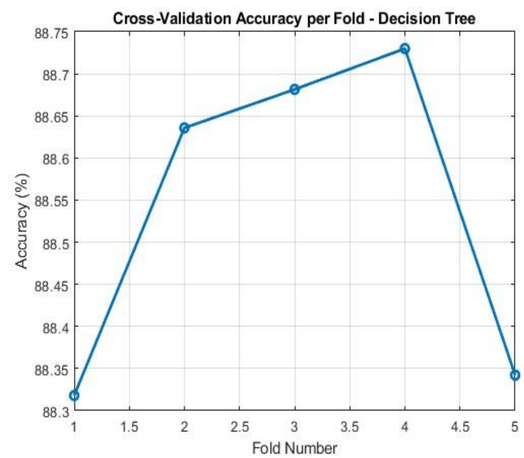

Fig 5. Cross-Validation Accuracy Line Plot- Decision Tree

Result for Optimized Random Forest Model

Here's a detailed description of each plot included in the MATLAB code:

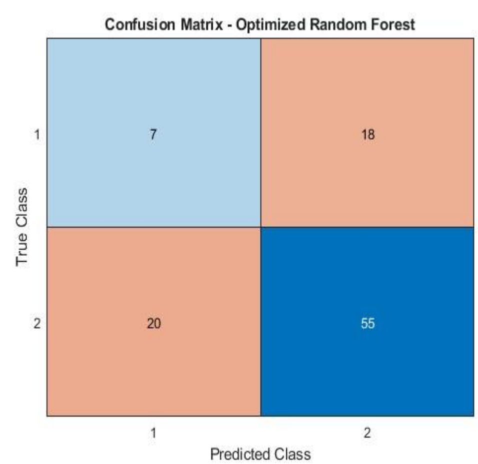

Fig 6. Confusion Matrix – Random Forest

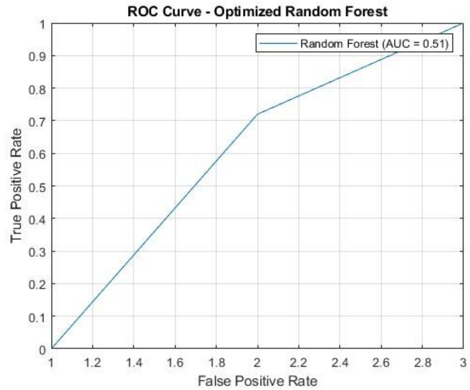

Fig 7 ROC Curve – Random Forest

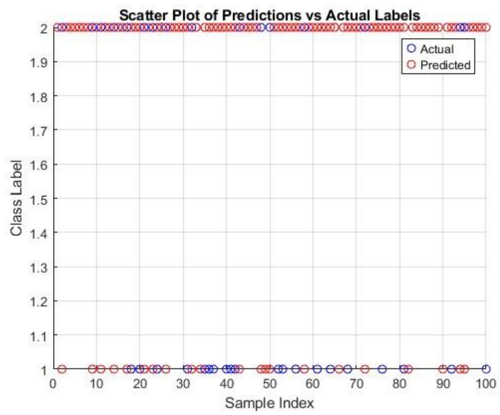

Fig 6. Scatter Plot of True vs Predicted Labels – Random Forest

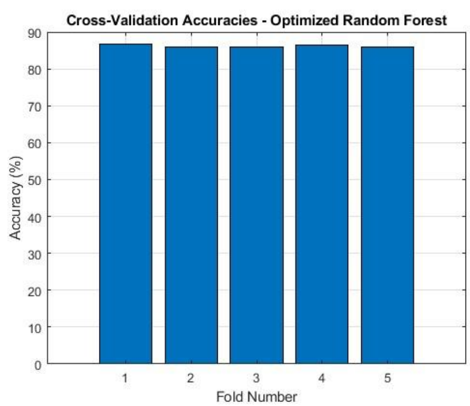

Fig. 9 Cross-Validation Accuracy Bar Chart– Random Forest

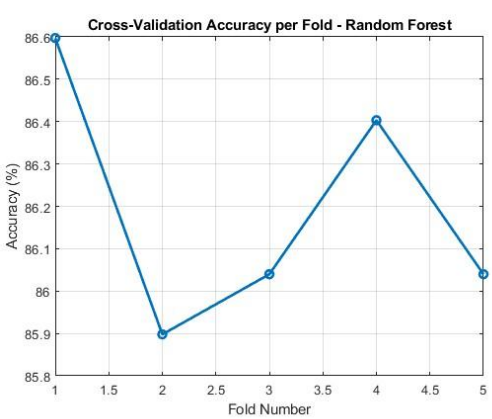

Fig 10 Cross-Validation Accuracy Line Plot– Random Forest

Decision Tree Accuracy: 86.1955%

Optimized Random Forest Accuracy: 88.5415%

II.CONCLUSION

This study presents a machine learning-based predictive maintenance framework using Decision Tree and Random Forest models. The results show that the Random Forest model achieves higher accuracy (88.54%) compared to the Decision Tree model (86.19%), making it more suitable for predictive maintenance applications.

The findings highlight the effectiveness of ensemble learning techniques in improving fault prediction accuracy and model stability. The use of the C-MAPSS dataset further validates the applicability of the proposed approach in real-world scenarios.

Future research can focus on:

- Deep learning approaches (LSTM, RNN) for time-series analysis
- Remaining Useful Life (RUL) prediction
- Real-time predictive maintenance systems
- Hybrid models combining ML and deep learning

III.REFEENCES :

[1] Z. M. Çınar, A. A. Nuhu, Q. Zeeshan, et al., “Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0,” Sustainability, vol. 12, no. 19, p. 8211, 2020. DOI: https://doi.org/10.3390/su12198211

[2] Aitzaz Ahmed Murtaza, Amina Saher, et al., “Paradigm shift for predictive maintenance and condition monitoring from Industry 4.0 to Industry 5.0: A systematic review, challenges and case study”, Results in Engineering, Volume 24, 2024, https://doi.org/10.1016/j.rineng.2024.102935

[3] Özcan H. Predictive maintenance in aircraft engine maintenance using the C-MAPSS dataset: performance comparison and evaluation of machine learning classification algorithms. Artificial Intelligence for Engineering Design, Analysis and Manufacturing. 2026. doi:10.1017/S0890060426100249

[4] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

DOI: https://doi.org/10.1023/A:1010933404324

[5] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees,

Wadsworth International Group, 1984. DOI: https://doi.org/10.1201/9781315139470

[6] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006. DOI: https://doi.org/10.1016/j.patrec.2005.10.010

[7] D. M. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011. DOI: https://doi.org/10.48550/arXiv.2010.16061

[8] Ron Kohavi. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th international joint conference on Artificial intelligence -

Volume 2 (IJCAI'95). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1137–1143.

[9] S. Kaparthi and D. Bumblauskas, "Designing predictive maintenance systems using decision tree-based machine learning techniques," International Journal of Quality & Reliability Management, vol. 37, no. 4, pp. 659–686, Feb. 2020, doi: 10.1108/IJQRM-04-2019-0131.

[10] M. Paolanti, L. Romeo, A. Felicetti, A. Mancini, E. Frontoni, and J. Loncarski, "Machine Learning approach for Predictive Maintenance in Industry 4.0," in 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), Jul. 2018, pp. 1–6. doi: 10.1109/MESA.2018.8449150.

[11] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," Computers & Industrial Engineering, vol. 137, p. 106024, Nov. 2019, doi: 10.1016/j.cie.2019.106024.

[12] Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0," Sustainability, vol. 12, no. 19, Art. no. 19, Jan. 2020, doi: 10.3390/su12198211.

[13]. Zhang, Weiting., Yang, Dong., & Wang, Hongchao. (2019). Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. IEEE Systems Journal, 13, 2213-2227. http://doi.org/10.1109/JSYST.2019.2905565

[14]. Amruthnath, Nagdev., & Gupta, Tarun. (2018). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), 355-361. http://doi.org/10.1109/IEA.2018.8387124

[15]. Yu, Wenjin., Dillon, T.., Mostafa, Fahed., Rahayu, W.., & Liu, Yuehua. (2020). A Global Manufacturing Big Data Ecosystem for Fault Detection in Predictive Maintenance. IEEE Transactions on Industrial Informatics, 16, 183-192. http://doi.org/10.1109/TII.2019.2915846

[16]. Carvalho, T.., Soares, Fabrízzio., Vita, Roberto., Francisco, R.., Basto, João P.., & Alcalá, S. G. S.. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng., 137. http://doi.org/10.1016/j.cie.2019.106024

[17] Çakir, Mustafa., Güvenç, M.., & Mistikoglu, S.. (2020). The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system. Comput. Ind. Eng., 151, 106948.

http://doi.org/10.1016/j.cie.2020.106948

[18] Ayvaz, S.., & Alpay, Koray. (2021). Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. Expert Syst. Appl., 173 , 114598 . http://doi.org/10.1016/j.eswa.2021.114598

[19] Theissler, Andreas., Pérez-Velázquez, Judith., Kettelgerdes, M.., & Elger, G.. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. Reliab. Eng. Syst. Saf., 215, 107864. http://doi.org/10.1016/j.ress.2021.107864

[20] Cheng, Jack C. P.., Chen, Weiwei., Chen, Keyu., & Wang, Qian. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. Automation in Construction, 112, 103087. http://doi.org/10.1016/j.autcon.2020.103087

[21] Teoh, Yyi Kai., Gill, S.., & Parlikad, A.. (2021). IoT and Fog-Computing-Based Predictive Maintenance Model for Effective Asset Management in Industry 4.0 Using Machine Learning. IEEE Internet of Things Journal , 10 , 2087-2094 . http://doi.org/10.1109/JIOT.2021.3050441

[22] Çınar, Z.., Nuhu, Abubakar Abdussalam., Zeeshan, Q.., Korhan, Orhan., Asmael, M.B.A., & Safaei, B.. (2020). Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. Sustainability . http://doi.org/10.3390/su12198211

[23] Kanawaday, Ameeth., & Sane, Aditya. (2017). Machine learning for predictive maintenance of industrial machines using IoT sensor data. 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), 87-90. http://doi.org/10.1109/ICSESS.2017.8342870

[24] Dalzochio, J.., Kunst, Rafael., Pignaton, E.., Binotto, A.., Sanyal, Srijnan., Favilla, J.., & Barbosa, J.. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. Comput. Ind., 123, 103298.

http://doi.org/10.1016/j.compind.2020.103298

[25] Feroz, Sainab., & Dabous, S. A.. (2021). UAV-Based Remote Sensing Applications for Bridge Condition Assessment. Remote. Sens., 13, 1809. http://doi.org/10.3390/rs13091809

[26]. Aivaliotis, P.., Georgoulias, K.., & Chryssolouris, G.. (2019). The use of Digital Twin for predictive maintenance in manufacturing. International Journal of Computer Integrated Manufacturing , 32 , 1067 - 1080 . http://doi.org/10.1080/0951192x.2019.1686173

[27]. Peng, Ying., Dong, M.., & Zuo, M.. (2010). Current status of machine prognostics in condition-based maintenance: a review. The International Journal of Advanced Manufacturing Technology , 50 , 297-313 . http://doi.org/10.1007/S00170-009-2482-0

[28]. N'cho, J.., Fofana, I.., Hadjadj, Y.., & Beroual, A.. (2016). Review of Physicochemical-Based Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers. Energies , 9 , 367 . http://doi.org/10.3390/EN9050367

[29] J. Dalzochio et al., "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges," Computers in Industry, vol. 123, p. 103298, Dec. 2020, doi: 10.1016/j.compind.2020.103298.

[30] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine Learning for Predictive Maintenance: A

Multiple Classifier Approach," IEEE Transactions on Industrial Informatics, vol. 11, no. 3, pp. 812–820, Jun. 2015, doi: 10.1109/TII.2014.2349359.

[31] W. Zhang, D. Yang, and H. Wang, "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey," IEEE Systems Journal, vol. 13, no. 3, pp. 2213–2227, Sep. 2019, doi: 10.1109/JSYST.2019.2905565.