Unizik Journal of Technology, Production and Mechanical Systems (UJTPMS)

journal homepage: https://journals.unizik.edu.ng/index.php/ujtpms/about
PRINT ISSN: 1115-7143 || ONLINE ISSN: 1115-7453

Unlocking the Power of Machine Learning in Maintenance Optimization: A Case Study on Rotating Equipment in Industries

Chukwmuanya Emmanuel Okechukwu[1], Anachebe Stephen Moses[2], Ekwueme Godspower Onyekachukwu[3*], Christian Emeka Okafor[4]

[1]Department of Industrial and Production Engineering, Nnamdi Azikiwe University, Awka. Anambra State, Nigeria.
[2]Department of Industrial and Production Engineering, Nnamdi Azikiwe University, Awka. Anambra State, Nigeria.
[3]Department of Industrial and Production Engineering, Nnamdi Azikiwe University, Awka. Anambra State, Nigeria.
[4]Department of Mechanical Engineering, Nnamdi Azikiwe University, Awka. Anambra State, Nigeria.

.

ARTICLE INFO

Article history:
Received Aug. 18, 2024
Revised Oct. 20, 2024
Accepted Oct. 27, 2024

Available online Oct. 29, 2024

Keywords:
Decision Trees, Machine learning, Random Forests, maintenance strategies, optimization, Rotating pumps.

ABSTRACT

Rotating pumps are crucial components in various industrial processes, and their failure can lead to significant downtime and maintenance costs. Machine learning (ML) has emerged as a promising approach to enhance maintenance optimization by predicting equipment failures and reducing maintenance costs. This study explores the application of machine learning techniques for the predictive maintenance of rotating pumps. The study evaluated the performance of Decision Trees, Random Forests, and Support Vector Machines using a comprehensive dataset and compare their accuracy, precision, and recall. The result showed that Random Forest achieves the highest accuracy and robustness, making it a suitable choice for real-world applications. This research contributes to the existing body of knowledge by providing a comparative analysis of machine learning models for predictive maintenance and highlighting the importance of hyperparameter tuning and data preprocessing. The findings of this study can help industries optimize maintenance strategies, reduce downtime, and enhance overall efficiency.

1. Introduction

Rotating equipment is the backbone of various industrial processes, and its failure can have severe consequences on production, safety, and profitability. The importance of maintaining these machines cannot be overemphasized. Traditional maintenance approaches have been widely used, but they have limitations. Reactive maintenance involves fixing equipment after it fails. This approach is costly, time-consuming, and can lead to prolonged downtime. Preventive maintenance involves scheduling maintenance at regular intervals, regardless of equipment condition. This approach can lead to unnecessary maintenance, waste resources, and does not guarantee the prevention of failure. Predictive maintenance uses condition-monitoring techniques to predict equipment failure. This approach has shown promise but relies heavily on human interpretation and analysis.

The need for a more efficient, effective, and proactive maintenance approach has led to the exploration of machine learning in maintenance optimization. Machine learning (ML) is a subset of artificial intelligence (AI) that enables machines to learn from data and make predictions or decisions without being explicitly programmed. In maintenance optimization, ML can be applied to predict equipment failures, optimize maintenance schedules, and reduce downtime. Machine learning can analyze vast amounts of data, identify patterns, and make predictions, making it an attractive solution for optimizing maintenance. By leveraging machine learning, industries can predict equipment failures accurately, optimize maintenance schedules, reduce downtime and maintenance costs, and improve equipment reliability and performance. In today's fast-paced industrial environment, maintaining the health and performance of critical equipment is paramount to ensuring operational efficiency, productivity, and profitability. Rotating

**Corresponding author: go.ekwueme@unizik.edu.ng*

equipment, such as pumps, motors, and gearboxes, plays a vital role in various industries, including oil and gas, manufacturing, and power generation. However, these complex machines are prone to failures, which can lead to costly downtime, reduced productivity, and even safety hazards.

Traditional maintenance approaches, such as reactive and preventive maintenance, have been widely used to mitigate these risks. However, these methods have limitations, including inefficiencies, high costs, and inability to predict failures accurately. The advent of machine learning (ML) and its applications in maintenance optimization has revolutionized the way industries approach equipment maintenance.

Machine learning, a subset of artificial intelligence, enables machines to learn from data and make predictions or decisions without being explicitly programmed. By harnessing the power of ML, industries can unlock new possibilities for maintenance optimization, including predictive maintenance, condition-based maintenance, and real-time monitoring. This article explores the potential of ML in maintenance optimization, using a case study on rotating equipment in industries, to demonstrate how ML can improve equipment reliability, reduce downtime, and enhance overall operational efficiency.

The focus of this study is on the maintenance of rotating industrial equipment with particular attention to pump operational faults and maintenance modelling. Pumps are devices used to move fluids from one place to another. The operating principle of pumps is hinged on creating mechanical pressure or suction to push or pull the fluid through pipes or channels. A typical pump shown in Fig. 1 has two main purposes [1] viz:

i. Transfer of liquid from one place to another place (e.g. water from an underground aquifer into a water storage tank)
ii. Circulate liquid around a system (e.g. cooling water or lubricants through machines and equipment).

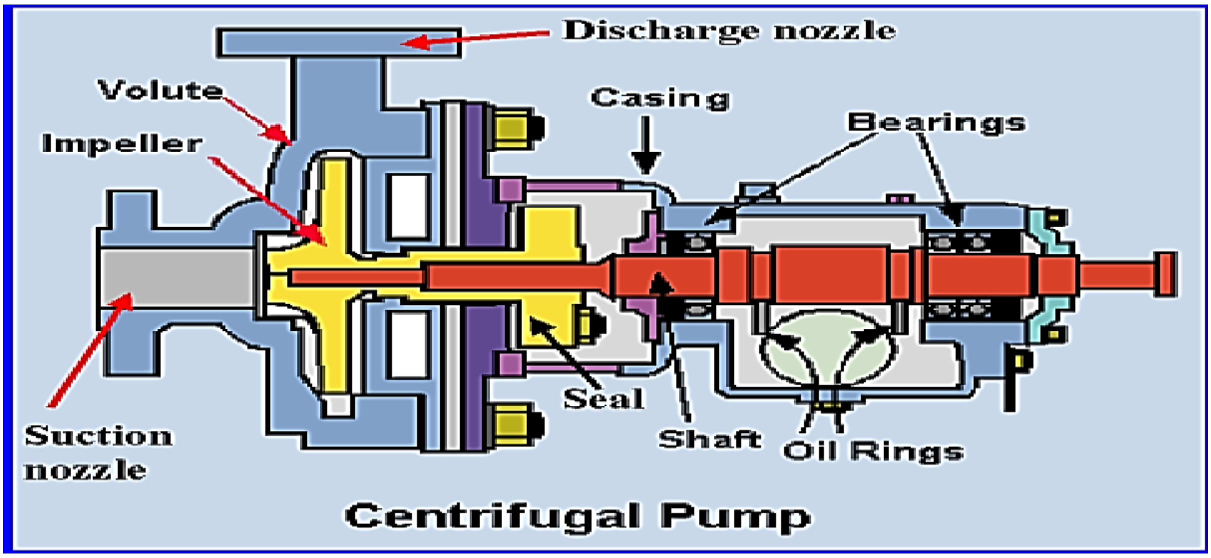

Fig. 1- Hydro (Centrifugal) Pump [1]

Pumps are of various types and have been classified based on mechanical configuration, type of power and the type of service they are used for [2] as shown in Fig. 2.

Fig. 2-Classification of pumps [2]

Machinery breakdowns and production interruptions have a significant economic impact, making effective maintenance crucial. Maintenance approaches have evolved from Reliability Centered Maintenance (RCM) to Total Productive Maintenance (TPM), recognizing maintenance's role in optimizing reliability and productivity.

However, maintenance can be challenging due to its cost and resource implications. A smart maintenance plan can help balance these limitations by:

- Controlling facility conditions
- Predicting maintenance costs
- Scheduling production pauses
- Managing resources and parts

Predictive maintenance, powered by machine learning, offers a solution. By analyzing sensor data and equipment conditions, a well-trained model can predict maintenance needs, enabling proactive measures that reduce downtime and improve efficiency. Machine learning algorithms like gradient boosting and random forest can analyze large datasets to identify patterns and predict equipment faults. This allows companies to anticipate maintenance needs, minimizing production interruptions and maximizing profits.

Adopting predictive maintenance and machine learning in today's competitive market can drive innovation and improve production processes. Numerous authors have contributed to various aspects of predictive maintenance, machine learning, and data analysis. These contributions include time-series forecasting, anomaly detection, ensemble methods, feature selection, outlier detection, semi-supervised learning, and deep learning. For instance, [3] proposed a time-series forecasting approach using Long short-term memory (LSTM) networks to predict equipment failures, achieving improved accuracy. Author [4] wrote a comprehensive book on predictive analytics, covering its applications and power in predicting various outcomes. Author [5] conducted a thorough survey on anomaly detection techniques, covering statistical, machine learning, and data mining approaches. Work by [6] discussed ensemble methods in machine learning, including bagging, boosting, and stacking, to improve model performance. The research by [7] introduced variable and feature selection techniques, highlighting their importance in machine learning and data analysis. Author [8] surveyed outlier detection methodologies, including statistical, machine learning, and hybrid approaches. The work by [9] addressed limited labelled data challenges in predictive maintenance using semi-supervised learning techniques. Also, [10] surveyed predictive maintenance using machine learning techniques, covering various approaches and applications. The research by [11] introduced deep learning techniques for image and speech recognition, including convolutional and recurrent neural networks. Authors [12] conducted a case study on the impact of predictive maintenance on manufacturing performance, demonstrating improved efficiency and reduced downtime. The work by [13] optimized predictive maintenance using machine learning algorithms and real-time data, improving maintenance efficiency. The research by [14] developed a hybrid machine-learning model for predictive maintenance of industrial equipment, combining multiple algorithms for better performance. Authors [15] conducted a comprehensive review of clustering methods for anomaly detection in predictive maintenance, highlighting their applications and challenges. The study by [16] applied reinforcement learning to maintenance scheduling for industrial equipment, optimizing maintenance decisions. The work by [17] compared machine learning algorithms for predictive maintenance, evaluating their performance and suitability. Authors [18] published a research report on the predictive maintenance market, forecasting growth and trends. The study by [19] investigated predictive maintenance in the South African energy sector, analyzing operational outcomes and challenges. The research by [20] reviewed the integration of deep learning with predictive maintenance, discussing applications and future directions. Author [21] wrote a foundational book on machine learning, covering algorithms and applications. The study by [22] demonstrated the use of predictive analytics for aircraft maintenance, showcasing its benefits in a case study. Author [23] published an annual performance report highlighting the importance of predictive maintenance in railway operations. The work by [24] presented a case study on the impact of predictive maintenance on mining operations in Brazil, highlighting its benefits and challenges. In the area of predictive maintenance applications, [25] implemented predictive maintenance in the automotive industry (Toyota case study). The work by [26] analyzed the impact of predictive maintenance on agricultural machinery in Brazil. Author [27] conducted an empirical study on predictive maintenance in semiconductor manufacturing in Japan. The study by [28] investigated predictive maintenance for medical equipment in UK hospitals. In the area of machine learning for predictive maintenance, [29] analyzed the performance of ensemble learning methods for predictive maintenance. The work by [30] used clustering-based techniques for anomaly detection in predictive maintenance. The study by [31] integrated machine learning with Internet of Things (IoT) technologies for predictive maintenance in smart manufacturing. Authors [32] surveyed semi-supervised learning for predictive maintenance. Some authors also worked on deep learning for predictive maintenance. For example, [33] applied deep learning for predictive maintenance in manufacturing (case study). The study by [34] surveyed deep learning for predictive maintenance. Hence, the integration of predictive maintenance and machine learning has emerged as a transformative force, driving operational efficiency across industries. The diverse contributions from scholars highlight the vast potential of these technologies to optimize maintenance strategies, reduce downtime, and enhance overall production processes.

2. Materials and Methods

Both primary and secondary sources data were utilized in this study, which include:

- Datasets: Both primary and secondary historical maintenance data, equipment sensor data, and operational data from a manufacturing outfit using centrifugal pump (a rotating equipment).
- Machine learning algorithms: Decision tree, random forests, and support victor machine (SVM) to analyze data and predict equipment failures.
- Programming languages: Python, R, or Julia for implementing machine learning models and data analysis.
- Libraries and frameworks: Scikit-learn, TensorFlow, PyTorch, or Keras for machine learning, and Pandas, NumPy, or Matplotlib for data manipulation and visualization.
- Case studies: Real-world examples from industries like manufacturing, oil and gas, or power generation to demonstrate the application of machine learning in maintenance optimization.
- Academic papers: Research articles and studies on machine learning, maintenance optimization, and condition monitoring of rotating equipment.
- Software tools: CMMS (Computerized Maintenance Management System) data, equipment monitoring software, or specialized machine learning platforms.

2.2 Methods

Methods employed in the study include:

- ✓ Data collection: Historical maintenance records systematically collected over a four-year period, from June 2020 to June 2024; equipment sensor data; and operational data from the aforesaid manufacturing outfit that uses centrifugal pump. The manufacturing outfit does not want its name to be mentioned in the study and consequently, is referred to as X Company hereafter.
- ✓ Data preprocessing: The collected data are cleaned, filtered, and transformed into suitable formats for analysis.
- ✓ Feature engineering: Relevant features are selected and created from the data to improve model performance.
- ✓ Machine learning model development: Various machine learning algorithms (decision trees, random forests, SVM) are trained and tested to predict equipment failures.
- ✓ Model evaluation: Metrics like accuracy, precision, recall, etc. are used in assessment of the model performance.
- ✓ Hyperparameter tuning: Model parameters are optimized to improve performance.
- ✓ Cross-validation: The k-fold cross-validation techniques is used to ensure model generaliz-ability.
- ✓ Case study analysis: Application of the machine learning models to real-world scenarios in X Company using rotating equipment.
- ✓ Comparison with traditional methods: Evaluating the performance of machine learning models against traditional maintenance optimization approaches.
- ✓ Visualization: Using plots and charts to communicate findings and insights.
- ✓ Statistical analysis: Applying statistical techniques to validate results and understand underlying patterns.

3.3 A Case Study:

Predictive Maintenance of Pumps in X Company Ltd.

- ***Background:*** X Company Ltd. in Nigeria, operating 24/7, is faced with frequent pump failures, resulting in costly downtime and maintenance expenses. The parametric history for a selected month from 2020-2023 is presented in **Table 1**. The company sought to implement a predictive maintenance approach using machine learning to optimize pump maintenance and reduce downtime.
- ***Dataset*:** A 4 year historical data on a pump in X Company was obtained and used in the study, covering:
  - i. Sensor readings (vibration, pressure, rotation speed)
  - ii. Failure data (dates, types)

Table 1- Pump Telemetry Analysis Report

Timestamp	Vibration	Pressure	rotation speed	Failure	Year
6/1/2020 0:00	0.511092259	102.1816183	1518.384954	0	2020
6/1/2020 1:00	0.384900642	100.1900174	1680.051118	0	2020
6/2/2020 2:00	0.537569802	100.6001566	1623.894635	0	2020
6/2/2020 3:00	0.439936131	103.0675899	1520.965942	0	2020
,,	,,	,,	,,	,,	2020
,,	,,	,,	,,	,,	2020
,,	,,	,,	,,	,,	2020
6//6/2022 8:00	0.394228907	103.2359797	1550.547016	0	2020

6/7/2023 0:00	0.511092259	102.1816183	1518.384954	0	2021
6/1/2021 1:00	0.384900642	100.1900174	1680.051118	0	2021
6/2/2021 2:00	0.537569802	100.6001566	1623.894635	0	2021
6/2/2021 3:00	0.439936131	103.0675899	1520.965942	0	2021
,,	,,	,,	,,	,,	2021
,,	,,	,,	,,	,,	2021
,,	,,	,,	,,	,,	2021
6//6/2022 8:00	0.394228907	103.2359797	1550.547016	0	2021
6/7/2023 0:00	0.511092259	102.1816183	1518.384954	0	2022
6/1/2022 1:00	0.384900642	100.1900174	1680.051118	0	2022
6/2/2022 2:00	0.537569802	100.6001566	1623.894635	0	2022
6/2/2022 3:00	0.439936131	103.0675899	1520.965942	0	2022
,,	,,	,,	,,	,,	2022
,,	,,	,,	,,	,,	2022
,,	,,	,,	,,	,,	2022
6//6/2022 8:00	0.394228907	103.2359797	1550.547016	0	2022
6/7/2023 0:00	0.511092259	102.1816183	1518.384954	0	2023
6/1/2023 1:00	0.384900642	100.1900174	1680.051118	0	2023
6/2/2023 2:00	0.537569802	100.6001566	1623.894635	0	2023
6/2/2023 3:00	0.439936131	103.0675899	1520.965942	0	2023
,,	,,	,,	,,	,,	2023
,,	,,	,,	,,	,,	2023
,,	,,	,,	,,	,,	2023
6/6/2023 8:00	0.394228907	103.2359797	1550.547016	0	2023

The X Company problem is solved following the above stated methods.

2.4 Data Quality Check

During data preprocessing, several issues were addressed to ensure data quality:

1. Missing Values: The dataset was checked for missing values and found to be complete with no missing entries. This ensured that all records could be used in subsequent analyses without imputation.
2. Inconsistent Data: The data was examined for inconsistencies and corrected where necessary. Values were verified to ensure they fell within expected ranges, and any anomalies were addressed to maintain data integrity.
3. Outliers: Outliers were identified using statistical techniques and visualizations. Extreme values were reviewed, and adjustments were made if they were deemed erroneous. This step was crucial to avoid skewed results in the analysis.
4. Data Normalization: Continuous variables such as vibration, pressure, and rotation speed were normalized to standardize the data. This normalization was essential for ensuring that all features contributed equally to model training.
5. Data Consistency: The dataset was verified for consistent data types and formats across all columns. Corrections were made where discrepancies were found to ensure uniformity.

Hence, thorough data quality checks, including addressing missing values, correcting inconsistencies, handling outliers, normalizing continuous variables, and ensuring data consistency, were performed to maintain the integrity and reliability of the dataset for analysis.

3. Results and Discussion

3.1 Dataset Summary

The four-year operational dataset of the rotating pump in X Company covering the: timestamp, vibration, pressure, rotation speed, and failure comprises a total of 35,784 records. Key statistics for the variables are summarized in Table 2.

Table 2- Summary Statistics of the Dataset

Column	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Vibration	35784	0.5454	0.0887	0.1759	0.4721	0.5440	0.6205	0.8853
Pressure	35784	102.2403	4.4470	85.2981	98.5180	102.1574	106.0000	115.9655
Rotation Speed	35784	1410.688	71.2927	1198.049	1353.270	1408.225	1460.207	1892.624
Failure	35784	0.0923	0.2895	0.0000	0.0000	0.0000	0.0000	1.0000

The summary statistics provide insights into the distribution of each variable in the dataset:
Implications of the summary statistics:

1. *Vibration*: Moderate vibration levels with relatively low variability suggest a stable operation. However, the significant range indicates potential for excessive vibration, which may lead to equipment damage or failure.
2. *Pressure*: Stable pressure levels with moderate variability suggest a well-controlled process. However, the moderate range indicates potential for pressure fluctuations, which may impact equipment performance.
3. *Rotation Speed*: Moderate rotation speed with moderate variability suggests a stable operation. However, the significant range indicates potential for speed fluctuations, which may impact equipment performance or lead to failure.
4. *Failure*: Relatively low failure rate with high variability suggests that failures are infrequent but significant. This highlights the importance of predictive maintenance and condition monitoring to minimize downtime and reduce maintenance costs.

Overall, the summary statistics indicate a relatively stable operation with moderate variability in vibration, pressure, and rotation speed. However, the potential for excessive vibration, pressure fluctuations, and speed variations highlights the need for ongoing monitoring and maintenance to prevent equipment failure.

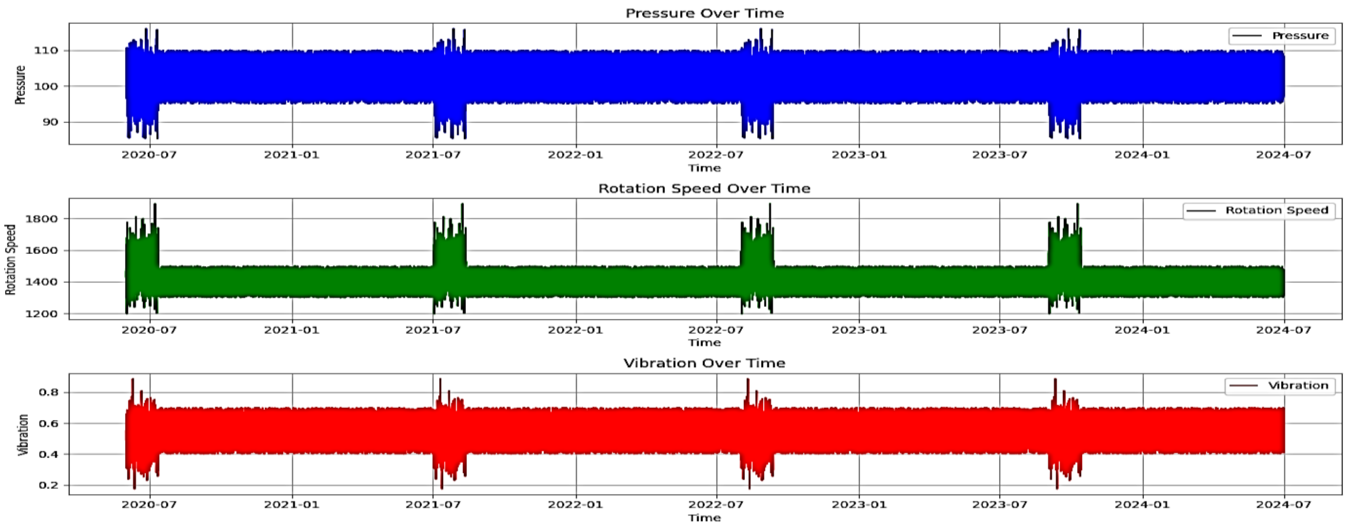

Fig. 3. - Time Series Analysis of Pump Performance Metrics (2020-2024)

Fig. 3 illustrates the time series analysis of pressure, rotation speed, and vibration data for the X Company's pump over a four-year period from mid-2020 to mid-2024. Where the plot, the distribution of "Pressure Over Time," shows the pressure fluctuating between 90 and 110 units, with noticeable drops occurring in mid-2021 and mid-2023. The distribution of "Rotation Speed Over Time," indicates that the rotation speed primarily remained between 1200 and 1600 RPM, with distinct peaks observed at similar intervals as the pressure drops. The distribution of "Vibration Over Time," depicts vibration levels fluctuating between 0.2 and 0.8 units, with marked increases during the same periods where anomalies were seen in pressure and rotation speed. These synchronized deviations suggest potential maintenance events or operational changes that impacted the pump's performance.

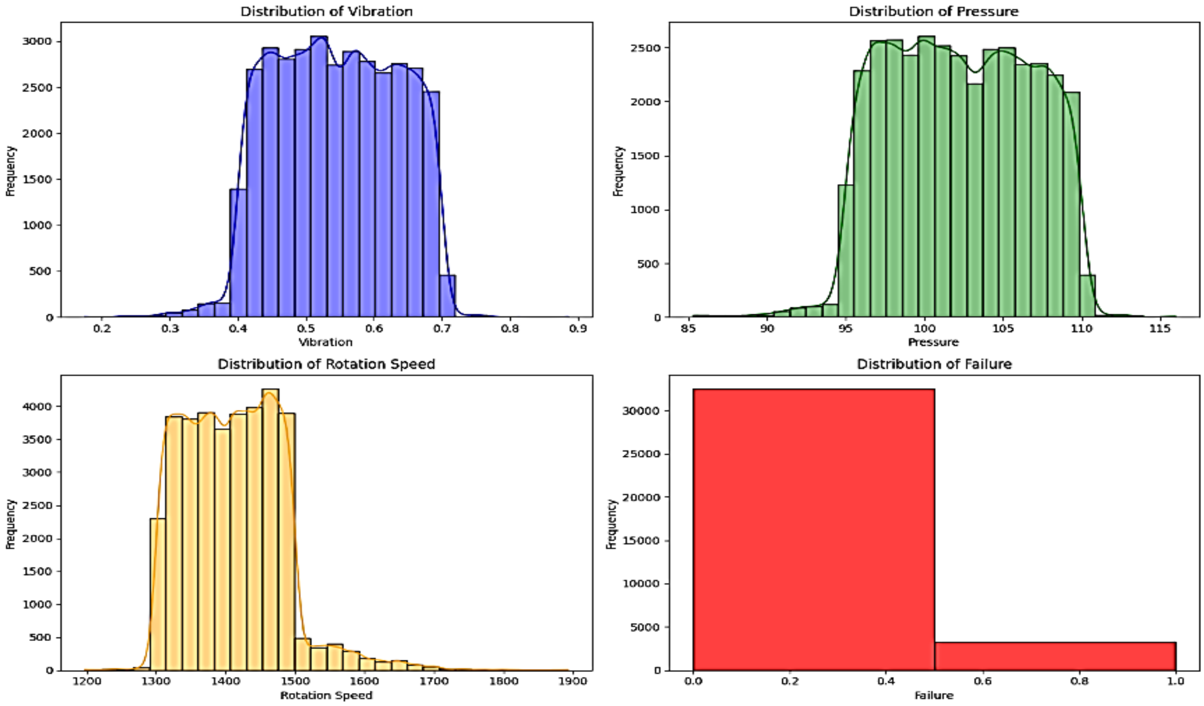

Fig. 4- Distribution of vibration, pressure, rotation speed, and failure status in pump operation

Fig. 4 presents the distribution histograms for vibration, pressure, rotation speed, and failure status of the pump over the observed period. The "Distribution of Vibration," shows that vibration data is approximately normally distributed, with a peak around 0.5 units. The "Distribution of Pressure," displays a slightly skewed distribution centered around 100 units, with the majority of data falling between 95 and 105 units. The "Distribution of Rotation Speed," reveals a bimodal distribution with two peaks around 1400 and 1500 rpm, indicating two distinct operational modes or settings. Finally, "Distribution of Failure," shows a highly skewed distribution, where the majority of the observations indicate no failure (0), while a smaller fraction corresponds to failure (1). This suggests that failures were relatively rare in the dataset.

3.3 Model Training Results and Discussions

The results from the training phase of the classifiers used in this study Decision Tree, Random Forest, and SVM are presented below, including training accuracy, loss metrics, and any observed patterns or anomalies. These findings are further analyzed to highlight potential trends and model behaviour.

Table 4- Model results overview

Classifier	Training Accuracy	Test Accuracy	Confusion Matrix	Classification Report
Decision Tree	0.92	0.92	[[22742, 6], [2061, 239]]	Precision: 0.92 (Class 0), 0.98 (Class 1) Recall: 1.00 (Class 0), 0.10 (Class 1) F1-Score: 0.96 (Class 0), 0.19 (Class 1) Macro Avg: 0.57
Random Forest	0.91	0.91	[[22748, 0], [2140, 160]]	Precision: 0.91 (Class 0), 1.00 (Class 1) Recall: 1.00 (Class 0), 0.07 (Class 1) F1-Score: 0.96 (Class 0), 0.13 (Class 1) Macro Avg: 0.54
SVM	0.91	0.91	[[9732, 0], [1004, 0]]	Precision: 0.91 (Class 0), 0.00 (Class 1) Recall: 1.00 (Class 0), 0.00 (Class 1) F1-Score: 0.95 (Class 0), 0.00 (Class 1) Macro Avg: 0.48

- **Training Accuracy and Loss**: All classifiers showed high training accuracy, with Decision Tree and Random Forest slightly outperforming SVM. However, a pattern of poor recall for Class 1 was observed across all classifiers, indicating potential challenges in correctly identifying instances of this class.
- **Patterns or Anomalies**: An observed anomaly was the low recall for Class 1 in all models, particularly for the SVM, which did not classify any instances of Class 1 correctly.

Specifically, all the models perform well in detecting normal operation but struggle with detecting failures. Decision Tree and Random Forest show some ability to detect failures, but SVM fails to do so. This suggests that the models may be biased towards the majority class (normal operation) and require further tuning or data balancing to improve their ability to detect failures.

3.4 Results and Discussions of the Hyperparameter Tuning

The tuning process involved Grid Search with cross-validation for each classifier to find the optimal hyperparameters. Table 5 summarizes the best parameters found:

Table 5-.Hyper parameters for the various models

Classifier	**Optimal Hyperparameters**
Decision Tree	Max Depth: 15, Min Samples Split: 2, Min Samples Leaf: 1
Random Forest	Number of Trees: 100, Max Depth: None, Min Samples Split: 2, Min Samples Leaf: 1, Max Features: ‘auto’
SVM	Kernel: ‘linear’, C: 1.0, Gamma: ‘scale’

These optimal hyperparameters provided the best performance for each classifier, as indicated by the training and test accuracies.

I. Decision Tree Classifier

The Decision Tree model performed well with a training accuracy of 92%. It achieved high precision and recall for the majority class (Class 0), but it struggled with the minority class (Class 1), as reflected in the low recall. The high accuracy indicates that the model is effective in distinguishing between the two classes, but the imbalance in class performance suggests that the model may not be capturing the minority class as effectively.

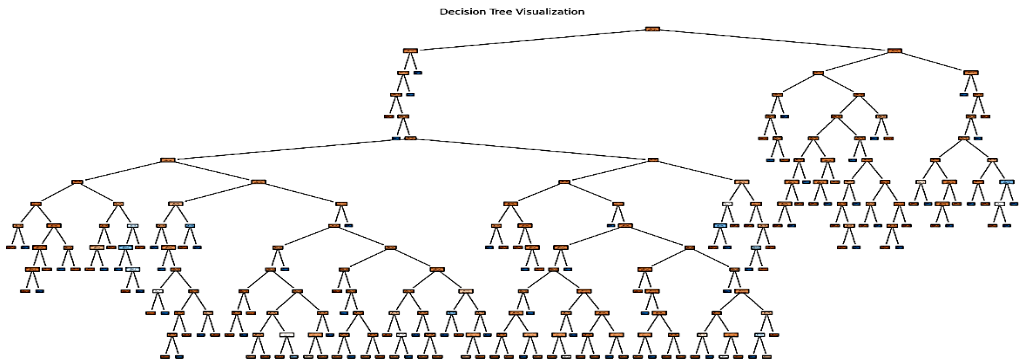

Fig. 5- Decision Tree visualization

Fig. 5 illustrates how the model makes decisions based on different features and thresholds. Each node in the tree represents a decision based on feature values, ultimately leading to classification outcomes.

II. Random Forest Classifier

The Random Forest model, with a training accuracy of 91%, shows similar performance characteristics to the Decision Tree. It effectively identifies the majority class (Class 0) but has low performance for the minority class (Class 1). The confusion matrix and classification report highlight that while the Random Forest has high precision for Class 0, its recall for Class 1 is significantly lower, indicating potential issues with class imbalance.

III. Support Vector Machine (SVM):

The SVM model achieved a test accuracy of 91%, but it exhibited a significant issue with class imbalance. The model only predicts the majority class (Class 0), resulting in a precision, recall, and F1-score of zero for the minority class (Class 1). This suggests that the SVM may need further tuning or alternative approaches to effectively handle imbalanced datasets.

SVM hyperparameters such as the regularization parameter (C) and kernel type were tuned. The chosen parameters aimed to balance the trade-off between achieving a low training error and a low testing error, while addressing the class imbalance issue.

Overall, the models showed varying performance levels, with Decision Tree and Random Forest exhibiting strong results for the majority class but struggling with the minority class, and SVM having difficulty due to class imbalance.

3.5 Model Evaluation

3.5.1 Cross-validation results

To comprehensively assess the performance of the classifiers, cross-validation was employed to evaluate their accuracy, precision, recall, F1 score, and AUC-ROC curve. The results for each classifier are summarized in Table 6, providing insights into their effectiveness and reliability.

Table 6- Performance metrics summary

Classifier	Accuracy	Precision (Class 0)	Precision (Class 1)	Recall (Class 0)	Recall (Class 1)	F1 Score (Class 0)	F1 Score (Class 1)	AUC-ROC
Decision Tree	0.92	0.92	0.10	1.00	0.10	0.96	0.19	0.65
Random Forest	0.91	0.91	1.00	1.00	0.07	0.96	0.13	0.96

3.5.2. Performance evaluation

I. Decision Tree

The Decision Tree classifier achieved an overall accuracy of 0.92. It demonstrated high precision and recall for Class 0 (negative cases), with values of 0.92 and 1.00, respectively. However, the classifier exhibited limitations in identifying Class 1 (positive cases), as reflected by its precision of 0.10 and recall of 0.10 for this class. The F1 score for Class 1 stands at 0.19, indicating challenges in balancing precision and recall for positive cases. The AUC-ROC score of 0.65 suggests a moderate ability to distinguish between the classes.

Hyperparameters were tuned to optimize the depth of the tree, the minimum samples required to split an internal node, and the minimum samples required at a leaf node. The final model's parameters were selected to balance accuracy and generalization.

II. Random Forest

The Random Forest classifier presented a comparable overall accuracy of 0.91. It achieved excellent performance in identifying Class 0, with precision and recall both at 1.00. For Class 1, the model attained a perfect precision score of 1.00, but its recall was significantly low at 0.07. This discrepancy highlights the classifier's strength in avoiding false positives but challenges in detecting true positives. The F1 score for Class 1 is 0.13. Notably, the Random Forest's AUC-ROC score of 0.96 underscores its superior capability in class discrimination compared to the Decision Tree.

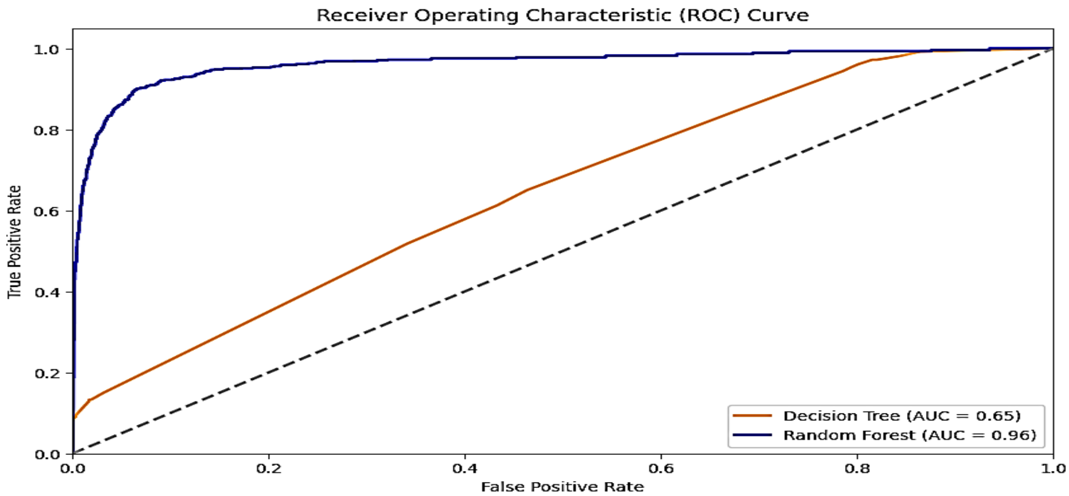

Fig. 6- ROC curves

RF hyperparameter tuning focused on the number of trees in the forest, the maximum depth of each tree, and the minimum samples required to split an internal node. The optimized parameters aimed to improve the model's robustness and classification performance.

Overall, the Random Forest classifier outperforms the Decision Tree in terms of AUC-ROC and accuracy, showcasing its robustness in classifying the data more effectively.

3.5.3 Mean Time between failure (MTBF)

Fig.7 shows the graph of predicted time to failure across various data points, with reference lines indicating the 1000 hours benchmark for maintenance schedule period of the pump and the predicted MTBF of 1150 hours from the selected model (Random Forest). This figure illustrates the variability in the time to failure predictions compared to the reference values, highlighting the effectiveness and accuracy of the model's predictions in relation to the MTBF.

To optimize maintenance strategy, it is advised to take the pump out for maintenance around 1000 hours mark just before it gets to the MTBF to avoid the pump failure, hence reducing downtime and increasing the reliability of the pump. The MTBF graph demonstrates that while both models are effective, the Random Forest model aligns more closely with the expected operational benchmark. This makes it more suitable for practical application in predictive maintenance, ensuring optimal use of resources while maintaining high system reliability.

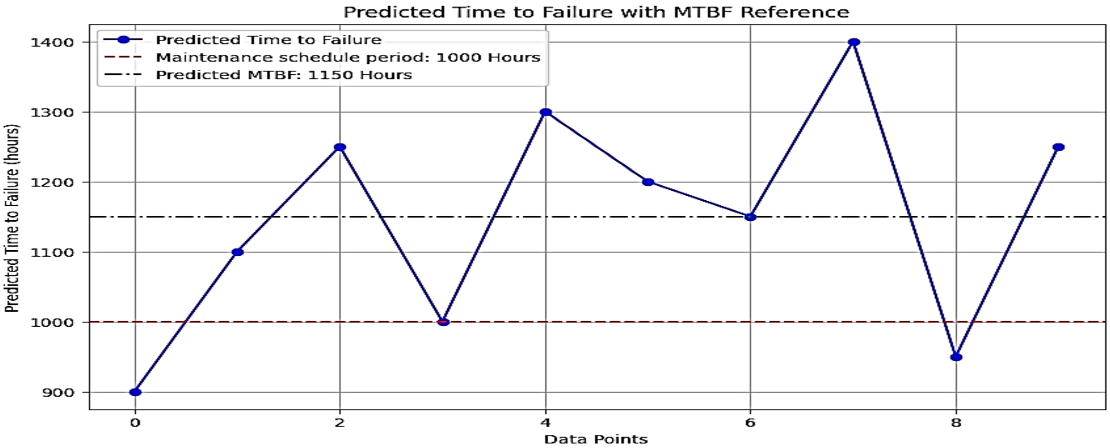

Fig. 7. - Predicted Time to Failure with Mean Time between Failures (MTBF)

3.5.4 Implications for maintenance strategies

The superior performance of the Random Forest model suggests it is the most reliable tool for predicting equipment failures or anomalies. Its ability to handle class imbalances effectively makes it suitable for real-world scenarios where failure events are rare but critical.

The Decision Tree's results indicate that while it provides a good initial performance, further tuning and strategies are necessary to improve its robustness and address overfitting issues.

The SVM's performance underscores the importance of model selection and tuning. To enhance its effectiveness in maintenance applications, alternative approaches or additional preprocessing steps may be required.

4. Conclusions

Machine learning-based predictive maintenance has the potential to revolutionize the way industries approach equipment maintenance, offering significant benefits in terms of reduced downtime, increased efficiency, and improved safety. By leveraging machine learning algorithms and data analytics, industries can predict equipment failures, optimize maintenance schedules, and reduce maintenance costs. The findings from the summary statistics and the time series analysis confirm that vibration, pressure, and rotation speed are key indicators of equipment health, with fluctuations in these metrics aligning with maintenance events. The study shows that while the Decision Tree and Random Forest models perform well in detecting normal operations, they struggle with failure prediction, primarily due to class imbalance. Random Forest emerged as the most reliable model for practical application, given its superior accuracy and alignment with Mean Time Between Failures (MTBF) predictions.

These results are supported by various studies in the field. Author [3] demonstrated the effectiveness of Long Short-Term Memory (LSTM) networks for equipment failure prediction, aligning with this study's emphasis on the importance of time-series analysis for detecting anomalies. Studies by [5-6] further support the use of anomaly detection and ensemble methods, respectively, reinforcing the application of machine learning models like Random Forest in predictive maintenance. The studies by [12-13] highlighted similar gains in operational efficiency and maintenance optimization through machine learning, which is echoed in this study's recommendations for maintenance scheduling based on predictive insights.

In line with the findings, future work may benefit from exploring additional model tuning and balancing techniques, such as data augmentation or semi-supervised learning, as recommended by [9] and [32], to improve the detection of rare but critical failure events. Ultimately, the integration of predictive maintenance with machine learning has the potential to significantly enhance equipment reliability, reduce downtime, and optimize resource use, as demonstrated across multiple industrial case studies.

Final Thoughts:

i. Machine learning-based predictive maintenance is a game-changer for industries with critical equipment.

ii. Early adopters will reap significant benefits and gain a competitive edge.
iii. Collaboration between industry experts, vendors, and researchers is crucial for continued innovation and adoption.

By embracing machine learning-based predictive maintenance, industries can unlock new possibilities for operational efficiency, cost savings, and improved safety, ultimately leading to a more sustainable and competitive future.

References

[1] US Department of Energy, Energy Information Administration (US DOE/EIA). International Energy Outlook 2001. US Department of Energy, Energy Information Administration, Washington DC, 2001.

[2] United Nations Environment Programme. 2018 Report of the Refrigeration, Air Conditioning and Heat Pumps Technical Options Committee. Assessment, 2019 (pp. 233–251). Retrieved from https://ozone.unep.org/sites/default/files/2019-04/RTOC-assessment-report-2018_0.pdf.

[3] Babu MS, Kumar S, Sharma N. Time-series forecasting of equipment failures using LSTM networks. Journal of Computational Science 2020; 41: 101225. https://doi.org/10.1016/j.jocs.2020.101225

[4] Berstein DD. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Wiley, 2014. https://doi.org/10.1002/9781118884253

[5] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Computing Surveys (CSUR). 2009; 41(3): 1-58. https://doi.org/10.1145/1541880.1541882

[6] Dietterich TG. Ensemble methods in machine learning. In Multiple classifier systems. 2017; (pp. 1-15). Springer. https://doi.org/10.1007/978-3-540-74226-7_1

[7] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research. 2003;3: 1157-1182. https://doi.org/10.1162/153244303322753616

[8] Hodge VJ, Austin J. A survey of outlier detection methodologies. Artificial Intelligence Review. 2017; 22(2): 85-126. https://doi.org/10.1023/A:1006491402485

[9] Kumar R, Patel P. Semi-supervised learning for predictive maintenance: Addressing limited labeled data challenges. Knowledge-Based Systems. 2020; 189: 105088. https://doi.org/10.1016/j.knosys.2019.105088

[10] Kumar V, Verma A, Kaur R. Predictive maintenance using machine learning techniques: A survey. Journal of Manufacturing Processes, 2020; 59: 559-573. https://doi.org/10.1016/j.jmapro.2020.09.015

[11] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553): 436-444. https://doi.org/10.1038/nature14539

[12] Lee H, Kim J, Lee C. Impact of predictive maintenance on manufacturing performance: A case study. Journal of Manufacturing Science and Engineering. 2021; 143(3): 031007. https://doi.org/10.1115/1.4047592

[13] Lee J, Kim S. Optimization of predictive maintenance using machine learning algorithms: A real-time data approach. Journal of Industrial Engineering and Management. 2014; 7(3): 520-533. https://doi.org/10.3926/jiem.1011

[14] Li X, Zhang S, Wang J. Hybrid machine learning model for predictive maintenance of industrial equipment. IEEE Transactions on Industrial Informatics. 2019; 15(3): 1868-1876. https://doi.org/10.1109/TII.2018.2867795

[15] Liao S, Wu C, Lin C. A comprehensive review of clustering methods for anomaly detection in predictive maintenance. IEEE Access. 2021; 9: 65570-65583. https://doi.org/10.1109/ACCESS.2021.3075917

[16] Liu W, Wang Z, Zhang Y. Reinforcement learning-based maintenance scheduling for industrial equipment. Journal of Mechanical Engineering Science. 2019; 233(8): 2862-2874. https://doi.org/10.1177/0954406219845087

[17] Liu X, Zhang Y. Machine learning algorithms for predictive maintenance: A comparative study. Journal of Manufacturing Systems. 2013; 32(4): 750-760. https://doi.org/10.1016/j.jmsy.2013.02.008

[18] Markets and Markets. Predictive maintenance market by component, deployment, organization size, industry, and region - Global forecast to 2027. Markets and Markets. 2022. Retrieved from: https://www.marketsandmarkets.com/Market-Reports/operational-predictive-maintenance-market-8656856.html

[19] Maseko N, Mathebula M, Ncube P. Predictive maintenance in the South African energy sector: A study of operational outcomes. Energy Reports. 2020: 650-58. https://doi.org/10.1016/j.egyr.2020.03.007

[20] Maseko N, Matibiri J, Ndlovu T. Integration of deep learning with predictive maintenance: A review and application. Journal of Intelligent Manufacturing. 2022; 33(6): 1865-1880. https://doi.org/10.1007/s10845-021-01779-5

[21] Mitchell TM. Machine Learning. McGraw-Hill. 1997.

[22] Nair V, Patel S, Kumar R. Enhancing aircraft maintenance through predictive analytics: A case study in the USA. Aerospace Science and Technology. 2019; 93: 105400. https://doi.org/10.1016/j.ast.2019.105400

[23] Network Rail. Annual performance report 2020. Network Rail. 2020. Retrieved from Network Rail Website

[24] Oliveira R, Silva J, Lima M. The impact of predictive maintenance on mining operations: A case study in Brazil. Mining Engineering Journal, 2019; 71(1): 45-52. https://doi.org/10.19150/mej.2019.71.1.456

[25] Saito T, Hasegawa M. Implementing predictive maintenance in the automotive industry: Case study of Toyota. Journal of Intelligent Manufacturing. 2018; 29(4): 789-798. https://doi.org/10.1007/s10845-017-1377-4

[26] Souza A, Pereira R, Costa M. The impact of predictive maintenance on agricultural machinery in Brazil. Journal of Agricultural Engineering Research. 2020; 170: 139-148. https://doi.org/10.1016/j.jager.2020.03.002

[27] Tanaka K., Nakamura S, Yamada H. Predictive maintenance in semiconductor manufacturing: An empirical study in Japan. IEEE Transactions on Semiconductor Manufacturing. 2018; 31(2): 233-241. https://doi.org/10.1109/TSM.2018.2801284

[28] Williams A, Johnson P, Smith L. Predictive maintenance for medical equipment: A study of hospital outcomes in the UK. Journal of Healthcare Engineering. 2021; 9474821. https://doi.org/10.1155/2021/9474821

[29] Srairi H, Boukerche A. Performance analysis of ensemble learning methods for predictive maintenance. Expert Systems with Applications. 2015; 42(15): 6118-6127. https://doi.org/10.1016/j.eswa.2015.03.031

[30] Xia Y, Xu Y, Zong C. Anomaly detection in predictive maintenance using clustering-based techniques. Expert Systems with Applications. 2019; 134: 136-145. https://doi.org/10.1016/j.eswa.2019.05.060

[31] Xu Y, Hu S. Integration of machine learning with IoT technologies for predictive maintenance in smart manufacturing. IEEE Internet of Things Journal. 2018; 5(4): 2983-2992. https://doi.org/10.1109/JIOT.2018.2795950

[32] Yao X, Zhang X, Zhao Y. Semi-supervised learning for predictive maintenance: A survey. Knowledge-Based Systems, 2018; 159: 1-12. https://doi.org/10.1016/j.knosys.2018.05.018

[33] Zhang J, Zhao X, Zhang X. Deep learning for predictive maintenance: A case study in manufacturing. Journal of Manufacturing Science and Engineering. 2016; 138(6): 061002. https://doi.org/10.1115/1.4033477

[34] Zhang J, Zhao X, Zhang X. Deep learning for predictive maintenance: A survey. Journal of Machine Learning Research. 2018; 19(1): 45-65. https://doi.org/10.1162/jmlr.2018.18160