The Original

MACHINE LEARNING-BASED PREDICTIVE MAINTENANCE AND FAULT DIAGNOSIS FOR INTELLIGENT MECHANICAL SYSTEMS

[1]Dr. Abhishek Sharma, [2]Dr. Sujesh Kumar, [3]Ramkrishna Mohan Kambli, [4]Arvinder Singh Channi, [5]Dr. Baijnath Das, [6]Dr. Anjul Rai

[1]Associate Professor, Department of Computer Engineering, Poornima College of Engineering, Jaipur, Rajasthan Technical University, Kota, Email: abhishek.sharma@poornima.org, ORCID ID: 0000-0002-4387-3974
[2]Assistant Professor, Department of Aeronautical Engineering, Mangalore Institute of Technology and Engineering, Moodabidri, Email: kumarsujesh64@gmail.com, ORCID ID: 0000-0001-7675-3138
[3]Ramkrishna Mohan Kambli, Lecturer in Mechanical Engineering, Department of Mechanical Engineering, Specialization in Mechanical Engineering, Government Polytechnic Panaji, Goa, Directorate of Technical Education, Goa, Email: kris.gim@gmail.com, ORCID ID: 0009-0000-2798-7800
[4]Associate Professor, Department of Mechanical Engineering, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, Email: arvindarsinghchanni@gmail.com, ORCID ID:0000-0002-3824-441X
[5]Lecturer, Teerthanker Mahaveer University College of Allied and Healthcare Sciences, Moradabad, Uttar Pradesh, India, Email: baijnathdasbiochemaiims@gmail.com
[6]Assistant Professor, Computer Integrated Manufacturing, Department of Mechanical Engineering, Bhilai Institute of Technology (Chhattisgarh Swami Vivekanand Technical University (CSVTU), Durg Chhattisgarh) Durg Bhilai, Email: anjul.rai@bitdurg.ac.in, ORCID ID: 0000-0002-8915-7586

ABSTRACT
Predictive maintenance has become a critical component of Industry 4.0 by enabling intelligent mechanical systems to minimize unexpected equipment failures, reduce maintenance costs, and improve operational reliability. This study proposes a machine learning-based predictive maintenance framework for machine failure prediction and fault diagnosis using the AI4I 2020 Predictive Maintenance Dataset. The dataset comprises 10,000 machine observations with operational parameters, including air temperature, process temperature, rotational speed, torque, tool wear, and machine type. A comprehensive comparative analysis was conducted using Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Extreme Gradient Boosting (XGBoost). To address the highly imbalanced failure distribution, multiple evaluation metrics, including Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic–Area Under the Curve (ROC–AUC), and Matthews Correlation Coefficient (MCC), were employed. The experimental results demonstrated that XGBoost achieved the best predictive performance, obtaining an accuracy of 98.95%, an F1-score of 82.93%, a ROC–AUC of 98.00%, and an MCC of 0.8289, outperforming the remaining machine learning models. Furthermore, SHapley Additive exPlanations (SHAP) identified torque, torque–speed ratio, and tool wear as the most influential predictors of machine failure, providing transparent model interpretation and supporting informed maintenance decision-making. The proposed framework offers an accurate and interpretable solution for predictive maintenance, contributing to enhanced equipment reliability, reduced downtime, and intelligent maintenance strategies in modern manufacturing environments.

KEYWORDS: Predictive maintenance, Intelligent mechanical systems, Machine learning, XGBoost, Explainable artificial intelligence (SHAP).

1. INTRODUCTION
Intelligent, interconnected, and data-driven production environments have emerged in the field of manufacturing, as a result of the swift developments of the so called "Industry 4.0" integrating technologies like the Industrial Internet of Things (IIoT), Artificial Intelligence (AI), cloud computing and big data analytics. These technologies allow for real-time monitoring, automation and decision making, thus improving manufacturing efficiency, productivity and operational reliability (Xu et al., 2018). This has made intelligent mechanical systems an integral part of modern manufacturing, where sensors, communication technology, and intelligent control algorithms operate in concert to maximize machine performance and help machines operate independently. These systems are commonly used in industrial machinery, automated production lines, and manufacturing facilities to enhance equipment reliability and reduce downtime (David et al., 2021; Kovalevskyy, 2024).

Predictive maintenance is one of the key applications of Industry 4.0 that proves to be a successful method to enhance equipment reliability and decrease the cost of maintenance. Predictive maintenance (PM) is an approach that exploits the operational data and intelligent algorithms to identify the degradation of the machines and forecast failure before they actually happen, as opposed to preventive maintenance (PM) and corrective maintenance (CM) that rely on fixed maintenance schedules and schedules triggered by failures, respectively (Achouch et al., 2022). The proactive maintenance approach helps reduce downtime, prolong equipment life, optimize resource utilization, and boost production efficiency. However, machine failure is one of the major problems that are faced by manufacturing industries as it causes delay in production process, leads to higher cost of maintenance, lower product quality and negative impacts on organizational productivity (Sobaszek et al., 2020).

The recent advances in machine learning have revolutionized predictive maintenance by providing highly accurate analysis of complex operational data gathered from industrial equipment. In the field of machine failure detection, machine learning algorithms have shown potential in detecting patterns of machine failure, categorizing the type of failure, and providing data-driven predictions to assist in intelligent maintenance decisions (Punukollu, 2024; Yadav et al., 2024). In addition, deep learning methods have been employed to autonomously learn complex features from industrial data to boost the accuracy of predictions, fault detection capabilities, etc. (Nguyen & Medjaher, 2019). While these developments have been made, there still are some challenges to be addressed. There are not enough studies to make comprehensive performance comparison of existing algorithms used for the machine learning. Furthermore, many predictive models are "black-box" models, that is, they are not very interpretable, which means that users' trust in the maintenance recommendations provided by this model is not high. But severe class imbalance is also a typical feature of industrial data sets in which failure events are present in relatively small numbers, resulting in skewed classification performance. Moreover, there are very few studies that have simultaneously dealt with binary machine failure prediction and multiclass fault diagnosis in a single machine learning framework with the use of explainable artificial intelligence methods. Hence, in this study, a comprehensive machine learning-based predictive maintenance and fault diagnosis framework for intelligent mechanical systems is presented, which involves comparing various classification algorithms, incorporating SHapley Additive exPlanations (SHAP) into interpret the model, assessing the importance of features, dealing with class imbalance problems through cost-sensitive learning, and providing practical suggestions for intelligent mechanical system maintenance.

2. Research Objectives

1. To develop a machine learning framework for predictive maintenance and fault diagnosis in intelligent mechanical systems.
2. To evaluate and compare the performance of machine learning algorithms for machine failure prediction and fault classification.
3. To improve the interpretability of predictive maintenance decisions using Explainable Artificial Intelligence (SHAP).

2. METHODOLOGY

2.1 Research Design

The study employs quantum research design to create and test a system for predictive maintenance and fault diagnosis for intelligent mechanical systems using machine learning. The proposed framework is systematic, which includes data acquisition, preprocessing, model development, performance evaluation, and model interpretation. Multiple machine learning algorithms are analyzed and compared for development of the best machine failure prediction model and classification.

2.2 Dataset Description and Preprocessing

The AI4I 2020 Predictive Maintenance Dataset is used for research, and it has 10,000 observations of machine condition and failure data. The data set contains features like air temperature, process temperature, rotational speed, torque, tool wear, and machine failure labels (Bansal,2020). Data pre-processing includes dropping irrelevant identifiers, encoding categoricals, normalizing numericals and performing an 80/20 stratified train-test split to ensure balanced class distribution when training a model.

2.3 Machine Learning Model Development

The machine learning algorithms used to predict machine failures and classify fault types are Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, XGBoost, LightGBM and CatBoost. Each model is optimized using five-fold cross validation and grid search to get the best configuration of the models. This comparative method allows us to evaluate various learning methods under the same experimental conditions.

2.4 Model Performance Evaluation

Standard classification performance metrics are used to evaluate the developed models to ensure that a reliable and unbiased comparison can be made. Accuracy, Precision, Recall, F1-score, ROC-AUC and Matthews Correlation Coefficient (MCC) are used for binary machine failure prediction. Macro Precision, Macro Recall, Macro F1-score, Confusion Matrix and Cohen's Kappa are also considered for multiclass fault diagnosis. These are overall metrics of predictive accuracy, overall robustness and class-wise metrics.

2.5 Explainable Artificial Intelligence

The proposed framework incorporates SHapley Additive exPlanations (SHAP) to enhance the interpretation of machine learning predictions. SHAP quantifies the contribution made by each feature of the input to the decision that is made in each individual prediction and from the overall model, making it possible to interpret the decision-making process. SHAP summary plots and feature importance rankings are analyzed to identify the variables which have the greatest impact on machine failure prediction and fault diagnosis, providing reliable and explainable machine maintenance decision making.

3. RESULTS

3.1 Descriptive Statistics of the Dataset

The AI4I 2020 Predictive Maintenance Dataset was made up of 10,000 observations of machines operating under simulated industrial conditions. Data validation confirmed that there were no any missing values in data set and no duplicate values in data set, which are suitable for machine learning analysis. There were 6 operational predictor variables and 1 binary target variable indicating whether the machine failed or not, and 1 multiclass target variable indicating machine failure categories. The average air temperature and process temperature were 300.01K and 310.01K, respectively as presented in Table 1. An average rotational speed of 1538.78 rpm, 39.99 Nm of mean torque and 107.95 min of mean tool wear were obtained by machines. The binary target variable had a mean value of 0.034 which meant that only 3.39% of the observations were machine failures, thus validating the very imbalanced nature of these data (Figure 1).

Table 1. Descriptive statistics of the numerical variables

Variable	Mean	SD	Minimum	Maximum
Air temperature (K)	300.005	2.000	295.3	304.5
Process temperature (K)	310.006	1.484	305.7	313.8
Rotational speed (rpm)	1538.776	179.284	1168	2886
Torque (Nm)	39.987	9.969	3.8	76.6
Tool wear (min)	107.951	63.654	0	253
Machine failure	0.034	0.181	0	1

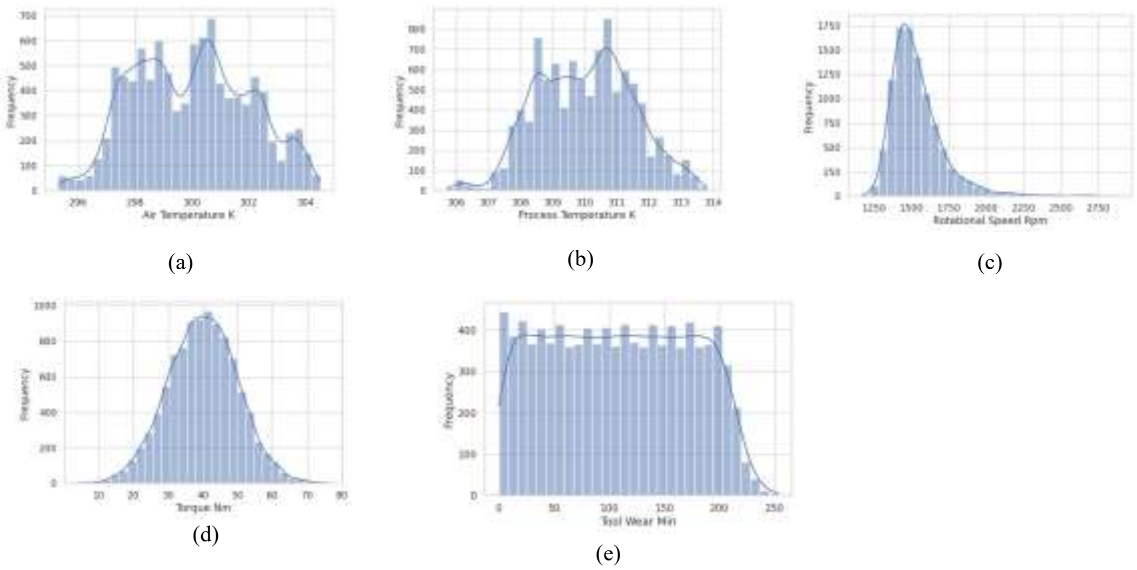


Figure 1. Distribution of Numerical Variables

Figure 2 illustrates the distributions of the numerical predictor variables used for machine failure prediction. Air temperature (a), process temperature (b), and torque (d) exhibit approximately normal distributions, whereas rotational speed (c) is positively skewed and tool wear (e) shows a relatively uniform distribution.

The distribution of machines is given in Table 2. Most observations were collected for Machine Type L (60.00%), with the remaining observations split between Machine Type M (29.97%) and Machine Type H (10.03%) (Figure 2), suggesting that the data is driven by machines with low loads.

Table 2. Distribution of machine types

Machine Type	Frequency	Percentage (%)
L	6000	60.00
M	2997	29.97
H	1003	10.03

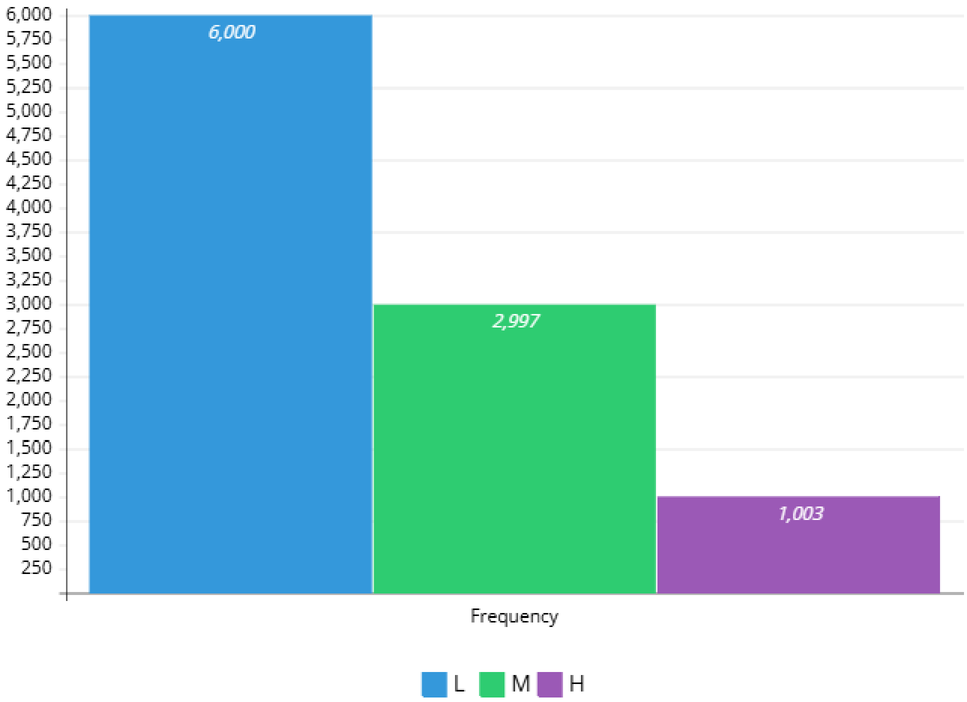


Figure 2. Distribution of Machine Types

Figure 3 presents the distribution of machine types within the dataset. Machine Type L accounts for the largest proportion of observations (6,000), followed by Type M (2,997) and Type H (1,003). This distribution indicates a predominance of low-load machines, providing diverse operational conditions for developing and evaluating predictive maintenance models.

3.2 Distribution of Machine Failures

The summary statistics of the binary machine failure target are shown in Table 3. The 10,000 observations were broken down into 9,661 (96.61%) observations that were associated with normal machine operation and 339 (3.39%) observations that were associated with machine failure events. This imbalance is quite extreme and suggests that metrics other than accuracy would be useful to consider when assessing predictive models to guarantee accurate identification of minority failure cases.

Table 3. Distribution of binary machine failure

Machine Failure	Frequency	Percentage (%)
No (0)	9661	96.61
Yes (1)	339	3.39

Table 4 shows the failure distribution of the multiclass. The largest segment of the data was in the No Failure category (96.52%), with Heat Dissipation Failure, Power Failure, Overstrain Failure, Tool Wear Failure and Random Failures representing only 3.48% of the data. The multiclass classification problem was very difficult since the class of Random Failures was extremely small with only 18 observations/0.18% of the total.

Table 4. Distribution of failure types

Failure Type	Frequency	Percentage (%)
No Failure	9652	96.52
Heat Dissipation Failure	112	1.12
Power Failure	95	0.95
Overstrain Failure	78	0.78
Tool Wear Failure	45	0.45
Random Failures	18	0.18

3.3 Exploratory Data Analysis

Pearson correlation was carried out to examine the correlation between various numerical variables. The results showed that air temperature and process temperature were highly correlated (r = 0.88) as seen in Figure 1 and tabulated in Table 5; this means that as the machine is operating, air temperature will increase as process temperature increases. On the other hand, a strong negative correlation between rotational speed and torque was observed (r = –0.88), indicating that the machines with a higher speed of rotation tend to have lower torque.

The linear correlation of the failure of machines with the predictor variables was relatively weak. The strengths of the positive relationships were 0.19 for machine failure, 0.11 for tool wear, and 0.08 for air temperature, respectively. Process temperature had only a weak positive correlation (r = 0.04), while rotational speed showed a weak negative correlation (r = –0.04). The results show that machine failure is caused by complex nonlinear interactions and is not only determined by one of the operational variables.

The feature distribution plots also showed that air temperature, process temperature, and torque were approximately normally distributed, and that rotational speed was moderately positively skewed. The wear of the tools had a nearly uniform distribution in most of their operating range but dropped at high wear levels. Some extreme observations for torque and rotational speed were indicated in the boxplots and were not excluded because they were believed to be legitimate operating conditions of machines not errors in measurement (Figure 3).

Table 5. Pearson correlation coefficients

Variable Pair	Correlation (r)
Air temperature – Process temperature	0.88
Rotational speed – Torque	–0.88
Torque – Machine failure	0.19
Tool wear – Machine failure	0.11
Air temperature – Machine failure	0.08
Process temperature – Machine failure	0.04
Rotational speed – Machine failure	–0.04

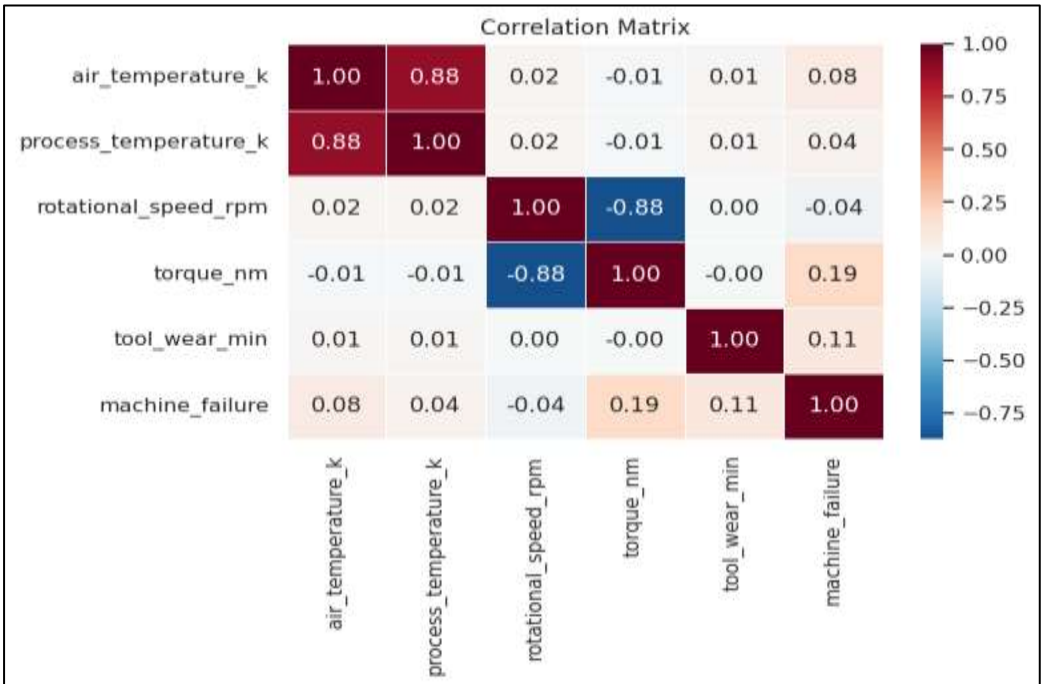


Figure 3. Pearson Correlation Matrix of the Predictor Variables and Machine Failure

Figure 3 illustrates the Pearson correlation matrix among the numerical variables. Air and process temperatures exhibit a strong positive correlation (r = 0.88), while rotational speed and torque show a strong negative correlation (r = −0.88).

3.4 Performance Comparison of Machine Learning Models

Five machine learning algorithms were tested on both five-fold stratified cross validation and independent test set. They were evaluated in terms of Accuracy, Precision, Recall, F1 score, ROC-AUC and Matthews Correlation Coefficient (MCC).

Comparative results are given in Table 6.
The performance of the evaluated algorithms was compared and XGBoost obtained the best overall predictive performance with an accuracy of 98.95%, precision of 92.73%, recall of 75.00%, F1-score of 82.93%, ROC–AUC of 98.00% and an MCC of 0.8289. The results showed that XGBoost can well balance machine failure identification and reduce the number of false positives.
The second-best result was obtained by the Random Forest classifier which obtained an accuracy score of 98.85% and F1 score of 80.67%. Random Forest had the lowest recall (70.59%), with XGBoost having the highest (73.11%), but Random Forest had the best precision (94.12%). The Decision Tree classifier predicted moderately well, while the Support Vector Machine showed the maximum amount of recall (91.18%) with a significantly lower precision (28.97%) meaning that it made many false positive predictions. The overall performance of Logistic Regression was the worst among all the models considered with an F1-score of just 30.18%, while maintaining a decent recall. The results show an overwhelming performance of ensemble learning algorithms over the existing machine learning algorithms under highly imbalanced data scenarios for predictive maintenance.

Table 6. Comparative performance of machine learning models

Model	Accuracy	Precision	Recall	F1-score	ROC–AUC	MCC
XGBoost	0.9895	0.9273	0.7500	0.8293	0.9800	0.8289
Random Forest	0.9885	0.9412	0.7059	0.8067	0.9703	0.8097
Decision Tree	0.9760	0.6515	0.6324	0.6418	0.8102	0.6295
Support Vector Machine	0.9210	0.2897	0.9118	0.4397	0.9682	0.4884
Logistic Regression	0.8635	0.1827	0.8676	0.3018	0.9359	0.3600

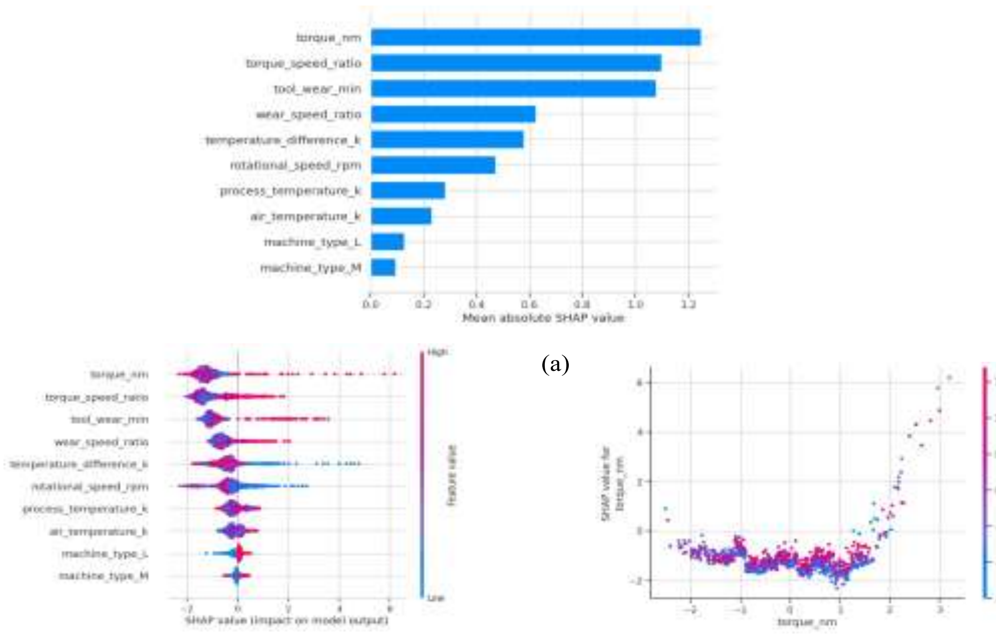
3.5 Explainable Artificial Intelligence Analysis

The best performing XGBoost model was then used to create SHapley Additive exPlanations (SHAP) to increase the transparency of the machine learning prediction. The feature importance rankings for the global models are given in Table 7 and the SHAP summary plots are shown in Figures 4 – 6.
From Table 7, the torque had the highest mean absolute SHAP value of 1.2469, making it the most significant factor affecting the prediction of the machine's failure. The second most significant was the engineered torque speed ratio (1.0977) followed closely by tool wear (1.0786). These results show that the original variables and the engineered variables significantly improved the predictive performance.
The SHAP summary plot also showed that the predicted probability of machine failure increased with higher levels of torque, torque–speed ratio, and tool wear, but decreased with rotational speed in a nonlinear manner. The dependence plot showed that the bigger the torque value, the more the SHAP values rose, which meant that if the mechanical loading on the machine becomes too large, the risk of failure of the machine will increase greatly. Machine variables, on the other hand, showed relatively low SHAP importance, indicating that operational conditions have a significantly higher influence on the failure prediction than the category of the machine.
Overall, the SHAP analysis demonstrates that the proposed XGBoost model achieves high predictive performance, and it also offers a transparent and interpretable decision support system that assists the maintenance engineer in determining the most important operational variables associated with machine failures (Figure 4).

Table 7. SHAP features importance ranking

Rank	Feature	SHAP Importance
1	Torque (Nm)	1.2469
2	Torque–Speed Ratio	1.0977
3	Tool Wear (min)	1.0786
4	Wear–Speed Ratio	0.6234
5	Temperature Difference (K)	0.5793
6	Rotational Speed (rpm)	0.4705

7	Process Temperature (K)	0.2817
8	Air Temperature (K)	0.2311
9	Machine Type (L)	0.1259
10	Machine Type (M)	0.0932



(b) (c)

Figure 5. SHAP-Based Global and Local Interpretability Analysis of the XGBoost Model

Figure 5 presents the SHAP-based interpretation of the XGBoost model, including (a) global feature importance, (b) the SHAP summary plot, and (c) the dependence plot for torque. Torque, torque–speed ratio, and tool wear are identified as the most influential predictors of machine failure, while the dependence plot reveals a nonlinear increase in failure risk at higher torque levels, demonstrating the model's ability to capture complex operational relationships.

4. DISCUSSION

In the present study, we created a predictive maintenance framework for intelligent mechanical systems based on the machine learning technique with the aid of the AI4I 2020 predictive maintenance dataset. XGBoost was the best machine learning algorithm in terms of all the evaluated metrics, with an accuracy of 98.95%, precision of 92.73%, recall of 75.00%, F1-score of 82.93%, ROC–AUC of 98.00% and an MCC of 0.8289. The results prove that ensemble learning algorithms are very powerful for modelling highly complex operational relationships when predicting machine failures especially if the data sets are highly imbalanced in industrial applications.

The advantage of XGBoost over other classifiers is that it uses gradient boosting approach to minimize prediction error over iterations and is also able to capture the nonlinear interactions between the operational variables. XGBoost, unlike classical machine learning algorithms, integrates regularization, efficient tree pruning, and powerful optimization techniques which enhance the prediction accuracy, and generalization capability (Chen & Guestrin, 2016). The present results corroborate previous predictive maintenance research that has demonstrated that the ensemble learning techniques are more effective than single classifiers, since they are able to model complex degradation trends and the variety of operating conditions of machines in the field (Susto et al., 2014; Paolanti et al., 2018). Likewise, Mota et al. (2022) illustrated that improving-based algorithms can offer more reliable predictive maintenance solutions in manufacturing environments by learning properly the nonlinear relationships between operational variables.

Random Forest had similar predictive accuracy, but was lower recall than XGBoost, so there were more true failures that were not detected. In terms of maintenance, failure incidents that have been missed could lead to unexpected equipment failures and production downtime and higher maintenance expenses. Therefore, the maximisation of failure detection and acceptable precision are crucial for predictive maintenance applications (Achouch et al., 2022; Jardine et al., 2006). XGBoost's performance in producing the highest F1-score and MCC demonstrates its powerful performance in both detecting the failure events and reducing false alarms, which makes it a feasible model to be employed in intelligent maintenance systems.

The third method, Support Vector Machine, had the largest recall, and had significantly lower precision and F1-scores due to having too many false positive predictions. This can happen if the dataset is highly imbalanced, and increasing the sensitivity of classifiers will lead to a decrease in specificity (He & Garcia, 2009). Similarly, Logistic Regression had a relatively high Recall value, but low Precision value, which meant that it was not very effective at distinguishing between failure and non-failure observations. These results corroborate the previous investigations that have demonstrated that traditional linear classifiers are not suitable to describe the non-linear relationships that are often found in industries' machine degradation processes (Zhao et al., 2019; Nguyen & Medjaher, 2019).

The data set was severely imbalanced; there were only 3.39% failures for the machines. This imbalance poses a huge problem since traditional accuracy measures can give false high expectations for performance, as they favour the majority class. For this reason, the present study utilized several measures such as Precision, Recall, F1-score, ROC–AUC and Matthews Correlation Coefficient to get a comprehensive picture of the classifiers performance. The use of MCC is very significant because it is a balanced evaluation even if the distribution of classes is very unequal (Chicco & Jurman, 2020). Using several evaluation metrics is in accordance with the previous studies on imbalanced machine learning and predictive maintenance (He & Garcia, 2009).

In addition, the explainable artificial intelligence analysis showed the usefulness of the proposed framework. Three most influential predictors of machine failure were identified using SHAP analysis as torque, torque–speed ratio and tool wear. The results are in line with the theory of machinery diagnostics which states that excessive mechanical loading and progressive mechanical component wear are the main signs of equipment deterioration (Jardine et al., 2006). In addition, the dependence plot showed that there was a nonlinear relationship between the amount of torque required for machine failure, which would not be easily captured by conventional statistical models, with increased torque at higher operating levels. This capability of giving clear explanations is one of the greatest shortcomings of black-box machine learning algorithms, and it makes the system for predictive maintenance more trustworthy (Lundberg et al., 2020; Matzka, 2020).

The engineered variables used in this study, such as the torque–speed ratio, wear–speed ratio, and temperature difference also made significant contributions to predictive performance. These derived variables are those that involve a combination of the operational parameters that are not directly measured in the original measurements. Although they are using ensemble learning algorithms as advanced ones, the high SHAP values in their features proved that feature engineering is still vital for modelling a predictive maintenance system. The same observations are found in the literature for predictive maintenance systems, where engineered operational indicators often provide better classification performance due to better representation of the mechanical behaviour (Calabrese et al., 2020; Paolanti et al., 2018).

The results also have significant practical applications in Industry 4.0 and intelligent manufacturing systems. With the use of continuous monitoring by industrial IoT sensors, it is possible to detect faults at an early stage and schedule maintenance work before a catastrophic fault occurs. A combination of machine learning and real-time sensor networks can enable transition from reactive or time-based maintenance to condition-based maintenance, which decreases downtime, maintenance costs, and enhances production efficiency (Xu et al., 2018; Achouch et al., 2022; Calabrese et al., 2020). This is intelligent maintenance that becomes one of the key elements in the smart manufacturing environment.

However, the results are promising, and a few caveats should be noted. The AI4I 2020 data set does not contain information on actual industrial production data, but rather simulated operating conditions of machines (Ikemura et al., 2021). The data set is widely used as a reference for predictive maintenance research, but is synthetic, which could hinder direct transfer to industrial systems (Matzka, 2020). Second, the distribution of failures in the data set is extremely imbalanced, with a few failures in each class while the Random Failures class has much more. This could impact the results of a multiclass classification. Advanced imbalance handling is also recommended for future research, such as Synthetic Minority Over-sampling Technique (SMOTE) or other cost sensitive learning techniques to enhance the recognition of minority classes (Chawla et al., 2002; He & Garcia, 2009). Third, while XGBoost had an outstanding prediction ability, applying this to real-time industrial data sets from various manufacturing industries to further validate the robustness and applicability of the proposed framework would be beneficial.

The present study shows that the advanced ensemble learning and explainable artificial intelligence are proved to be a good solution for predictive maintenance in an intelligent mechanical system. The proposed framework demonstrated very good performance in predicting and provided clear explanations that can be used for maintenance decision making. The results align with the studies reported in the literature on the subject of Industry 4.0, as they show that interpretable machine learning can significantly improve the reliability of the equipment, mitigate operational risks, and facilitate the implementation of smart maintenance strategies in industry 4.0 settings (Zonta et al., 2020; Achouch et al., 2022).

5.CONCLUSION

This research aimed at developing a machine learning predictive maintenance framework for intelligent mechanical systems using AI4I 2020 predictive maintenance dataset to enhance machine failure prediction and aid proactive maintenance decision making in Industry 4.0 context. The five machine-learning algorithms were compared with each other, and the results showed that XGBoost had the highest overall predictive performance, with an accuracy of 98.95%, an F1 score of 82.93%, a ROC–AUC of 98.00% and an MCC of 0.8289, outperforming Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine. The results validated the effectiveness of ensemble learning methods in dealing with nonlinear relations and large class imbalance generally found in predictive maintenance applications. Moreover, the explainability analysis of the proposed framework using SHAP showed that the torque, torque–speed ratio and tool wear are the most influential factors affecting the prediction of machine failures, providing transparency and interpretability of the obtained results. These results illustrate how high-performance machine learning combined with explainable artificial intelligence can lead to the prediction of accurate and understandable operational insights that help in the effective planning of maintenance while minimising unexpected equipment failures. The study was based on a benchmark set, but the proposed framework offers a solid basis for intelligent maintenance systems and can be expanded to the industrial field. Further research is needed to verify the proposed approach with real-time sensor data, develop more sophisticated imbalance-learning methods and explore hybrid deep learning and digital twin approaches to improve the accuracy of predictions, model adaptability and intelligent maintenance abilities in various manufacturing systems.

REFERENCES

1. Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On predictive maintenance in industry 4.0: Overview, models, and challenges. Applied sciences, 12(16), 8081.
2. Bansal, S. (2020). Machine predictive maintenance classification [Data set]. Kaggle. https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification
3. Calabrese, M., Cimmino, M., Fiume, F., Manfrin, M., Romeo, L., Ceccacci, S., ... & Kapetis, D. (2020). SOPHIA: An event-based IoT and machine learning architecture for predictive maintenance in industry 4.0. Information, 11(4), 202.
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
5. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1), 6.
6. David, S., Anand, R. S., Sheikh, S., Jebapriya, S., Andrew, J., & Xavier, S. B. (2021). A comprehensive overview on intelligent mechanical systems and its applications. Materials Today: Proceedings, 37, 733-736.
7. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.
8. Ikemura, K., Bellin, E., Yagi, Y., Billett, H., Saada, M., Simone, K., ... & Reyes Gil, M. (2021). Using automated machine learning to predict the mortality of patients with COVID-19: prediction model development study. Journal of medical Internet research, 23(2), e23458.
9. Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical systems and signal processing, 20(7), 1483-1510.
10. Kovalevskyy, S. (2024). Intelligent control systems for mechanical engineering technology tasks. Штучний інтелект. Фізико-математичні та технічні науки, міжнар. наук-техн. журнал, 4(101), 218-227.
11. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 56-67.
12. Matzka, S. (2020, September). Explainable artificial intelligence for predictive maintenance applications. In 2020 third international conference on artificial intelligence for industries (ai4i) (pp. 69-74). IEEE.
13. Mota, B., Faria, P., & Ramos, C. (2022, September). Predictive maintenance for maintenance-effective manufacturing using machine learning approaches. In International workshop on soft computing models in industrial and environmental applications (pp. 13-22). Cham: Springer Nature Switzerland.
14. Nguyen, K. T., & Medjaher, K. (2019). A new dynamic predictive maintenance framework using deep learning for failure prognostics. Reliability Engineering & System Safety, 188, 251-262.
15. Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018, July). 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) (pp. 1-6). IEEE.
16. Punukollu, M. (2024). Leveraging Machine Learning for Real-Time Predictive Maintenance in Smart Manufacturing Ecosystems. Los Angeles Journal of Intelligent Systems and Pattern Recognition, 4, 544-585.
17. SobASzek, Ł., GolA, A., & Świć, A. (2020). Time-based machine failure prediction in multi-machine manufacturing systems. Eksploatacja i Niezawodność, 22(1), 52-62.
18. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. IEEE transactions on industrial informatics, 11(3), 812-820.
19. Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: state of the art and future trends. International journal of production research, 56(8), 2941-2962.
20. Yadav, D. K., Kaushik, A., & Yadav, N. (2024). Predicting machine failures using machine learning and deep learning algorithms. Sustainable manufacturing and service economics, 3, 100029.
21. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical systems and signal processing, 115, 213-237.
22. Zonta, T., Da Costa, C. A., da Rosa Righi, R., De Lima, M. J., Da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & industrial engineering, 150, 106889.