EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Predictive maintenance of industrial milling machine using machine learning techniques

G Divya Deepak and Subraya Krishna Bhat*

Department of Mechanical and Industrial Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

Received 14 March 2025
Revised 8 July 2025
Accepted 19 August 2025

Abstract

Condition-based predictive maintenance of industrial machinery is a key area of research in the present world looking towards Industry 4.0. Machine learning (ML) techniques can have tremendous impact in this aspect because of their robust predictive modeling capabilities. The present paper aims to determine the optimized machine learning technique for the predictive maintenance of an industrial milling machine. The data pertaining to the operating parameters and the failure types of the machine is obtained from a public dataset with 10,000 data points. Five of the most popular classification ML algorithms namely, Artificial Neural Network (ANN), Discriminant Analysis (DA), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT) techniques are implemented for the dataset to determine their optimized hyperparameters for an effective prediction of the machine failure type. DT and ANN were found to be the two best techniques with overall accuracy of 99.15% and 98.8%, respectively, and superior performance metrics of Precision, Recall and F-Measure compared to the other models. The results obtained from the present study may be enriched in the future by incorporating deep learning-based models and hybrid ML and intelligent optimization techniques for effective predictive maintenance of various industrial systems. The present approach can thus be employed in real-time factory settings to realize the targets of Smart Manufacturing and Industry 4.0.

Keywords: Artificial neural network, Machine learning, Predictive maintenance, Decision tree, Industry 4.0

1. Introduction

Recently, there has been a huge acceleration in the development of innovative technologies including Internet of things (IoT), cloud computing, data analytics, and augmented reality. These innovative technologies have resulted in a radical shift in industrial maintenance approaches wherein automated systems are introduced with the capacity to forecast machine failure and increase in the longevity of operation [1, 2]. It is imperative to focus on energy optimization especially for production lines in the manufacturing segment as the machine's overall performance depends on its energy utilization efficiency [3, 4]. This has resulted in the research of new maintenance strategies such as condition-based maintenance, prognostic, identifying faults in a manufacturing environment, and overall health management [5]. Predictive Maintenance (PdM) uses historical data to predict machine behavior and is typically applied through prognostic monitoring or condition-based maintenance [6].

In a manufacturing process environment, PdM systems play a pivotal role in detecting and scheduling the maintenance activities required for ensuring product quality while avoiding early maintenance as well as reducing unnecessary costs due to down time. PdM helps in constant monitoring of the machine's performance and integrity, allowing maintenance only when it is essential. Furthermore, PdM systems are built on statistical interpretation methods, integrity variables, historical data, and various engineering approaches for early problem detection [7]. There are a variety of techniques for predicting a systems' health for effective predictive maintenance; for instance, the usage of machine learning (ML) techniques such as, Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naïve Bayes (NB), Discriminant Analysis (DA) and Decision Trees (DT) [8-13], etc.

Furthermore, PdM diagnosis methods can help determine fault type by examining the current machine status [12]. PdM strategy needs improvement in accuracy in line with its explicitness of prediction for ensuring higher efficacy in industrial applications [13]. Industries are rapidly adopting these machine-learning technologies to obtain more precise and accurate predictions required for the maintenance of industrial assets [14]. One of the major areas of IoT implementation is the concept of the Industrial Internet of Things (IIoT) used in PdM. IIoT serves as the bridge between manufacturing as well as production applications and machines for establishing coherent communication between them. These IoT-enabled systems can provide real-time machine data, monitor industrial machines, and differentiate the occurrence of wear and tear damage to prevent machine failures resulting in downtime thus ensuring production continuity [15].

Recently, Heymann and Schmitt [16] developed an ML-based pipeline for effectively predicting the remaining useful lifetime of a polymer 3D printing nozzle by integrating sensor data as well as training of a regression model. De Luca et al. [17] designed as PdM methodology using deep learning approaches for an intelligent maintenance strategy based on real-time sensorics and IoT data obtained from machines in an industrial setup. Karabacak [18] introduced an innovative approach based on ML techniques for tool wear prediction of an industrial milling machine. Multiple ML techniques were compared to identify artificial neural networks as the models

*Corresponding author.
Email address: sk.bhat@manipal.edu
doi: 10.14456/easr.2025.58

which to show the best accuracy. It was observed that choosing the key sensor signals can help develop effective ANN models. Gong and Chen [19] presented a convolutional neural network (CNN) based deep learning model utilizing the IoT and wireless sensor data for the PdM of wind turbine systems. A customized CNN was developed for making robust condition-based monitoring of components such as generators, transformers and gearbox during the wind turbine operation.

Most PdM studies focus on implementing ML models but do not systematically investigate hyperparameter optimization for improved performance [16-19]. Certain studies have focused on deep learning, which can be computationally expensive for real-time industrial applications [17, 19]. While deep learning-based methods such as Transformers and Kernel Attention Networks (KAN) have been introduced in recent years [20-22], classical machine learning models continue to be widely used in predictive maintenance due to their interpretability, lower computational cost, and ease of deployment in industrial settings. This study focuses on optimizing and benchmarking five well-established ML models for predictive maintenance applications. Unlike prior works, this study systematically evaluates and optimizes hyperparameters for five well-established ML models (ANN, DA, NB, SVM, DT) to achieve the best classification performance. The results provide a practical framework for selecting and fine-tuning ML models for PdM applications in Industry 4.0 settings.

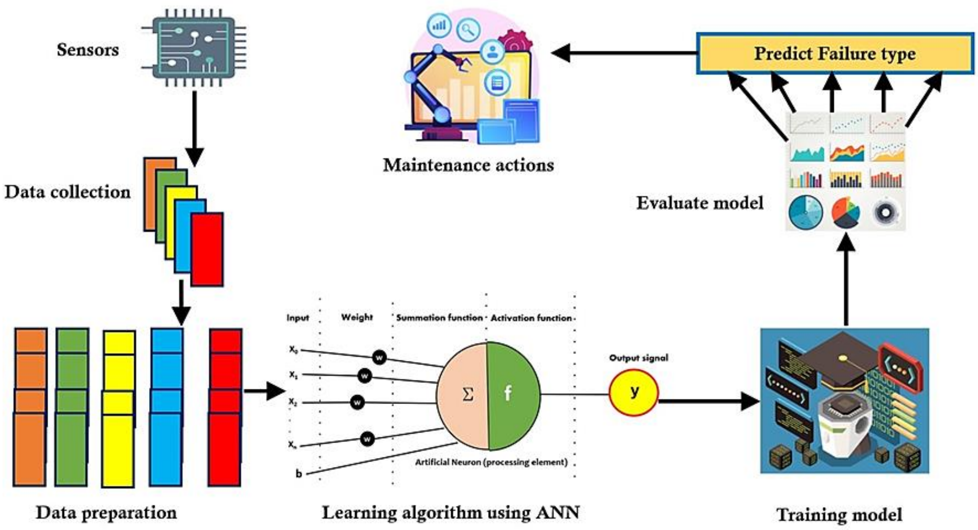

Figure 1 General structure of the proposed predictive maintenance model

In this manuscript, a labelled PdM dataset pertaining to the machine failure modes of an industrial milling machine is employed to develop the predictive model using ML algorithms. The dataset consists of 10,000 data points. There are 5 classes of machine failure modes that are classified (no failure, power failure, tool wear failure, overstrain failure, random failure, and heat dissipation failure). The failure class is dependent on the combination of input parameters such as, Air temperature [K], Process temperature [K], Rotational speed (rpm), Torque [Nm], and Tool wear [min]. The methodology of ML approach implemented for failure classification is shown in Figure 1.

Despite these advancements, several key challenges persist in the field of predictive maintenance, particularly for industrial machining systems:

- Many existing studies focus on model implementation but overlook systematic hyperparameter optimization, which is essential for maximizing model accuracy and reliability.
- Complex deep learning approaches, while powerful, often have high computational costs and limited interpretability, making them less suitable for real-time industrial deployment.
- The issue of class imbalance, where failure events are rare compared to normal operation, often leads to misleadingly high accuracy while minority failure classes remain poorly predicted.
- There is also a lack of clear comparative studies that systematically benchmark classical ML techniques under realistic industrial data scenarios.

While recent studies have explored the application of deep learning architectures such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer-based models for predictive maintenance, these approaches often require significantly higher computational resources and may lack the interpretability required for deployment in resource-constrained or safety-critical industrial environments. In this context, the present study focuses on classical, interpretable machine learning models as a practical benchmarking foundation. These models are widely used in industry due to their ease of deployment, lower training requirements, and transparent decision-making processes.

To address these limitations, the present study makes the following targeted contributions:

1. A comprehensive benchmarking of five widely used and interpretable ML classifiers — Artificial Neural Network (ANN), Discriminant Analysis (DA), Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) — applied to a publicly available PdM dataset for an industrial milling machine.
2. Systematic hyperparameter tuning for each model to optimize their classification performance.
3. Evaluation of model performance across multiple metrics (accuracy, precision, recall, F1-score) to provide a holistic assessment beyond overall accuracy.
4. Feature importance analysis using ANOVA, providing insights into the most critical operational parameters influencing machine failures.
5. Practical assessment of training and inference times, highlighting the feasibility of these models for real-time deployment in industrial environments.

A summary of key related works, their contributions, limitations, and how the present study addresses identified gaps is provided in Table 1.

Table 1 Research background and literature review summary

Author & Year	Approach	Findings	Limitations	How Present Study Addresses Gap
Heymann & Schmitt (2023) [16]	ML pipeline for PdM in polymer 3D printing	Effective prediction of remaining useful life using sensor integration	Focus on regression; limited to specific polymer application	Broad PdM application using classification of failure types
De Luca et al. (2023) [17]	Deep learning (DL) for intelligent PdM strategy	DL models enhance real-time PdM using sensor and IoT data	High computational complexity for real-time industrial deployment	Emphasis on interpretable, low-cost classical ML models
Karabacak (2024) [18]	ML-based tool wear prediction for milling machine	Identified ANN as best model for tool wear prediction	Limited to single failure type (tool wear)	Multi-class failure type classification for milling machines
Gong & Chen (2024) [19]	CNN-based PdM for wind turbines	CNN and IoT integration improve condition monitoring	Application-specific; deep models with limited interpretability	Focus on generalizable, interpretable ML approaches
Efeoğlu & Tuna (2022) [23]	SVM for PdM on AI4I 2020 dataset	Achieved high accuracy (~99%)	Poor performance on random failures, limited model comparison	Systematic comparison of 5 ML models with robust evaluation
Sengupta et al. (2023) [24]	Ensemble model for PdM of armored vehicles	Improved fault prediction using ensemble approach	Application-specific, ensemble complexity	Focus on milling machine PdM with interpretable ML models
Ghasemkhani et al. (2023) [25]	Balanced K-Star for PdM with IoT-enabled data	Improved explainability and performance for PdM	Specific to IoT-enabled systems; limited hyperparameter tuning	Comprehensive hyperparameter optimization of ML models
Present study	ML pipeline for PdM in polymer 3D printing	Effective prediction of remaining useful life using sensor integration	Focus on regression; limited to specific polymer application	Broad PdM application using classification of failure types

By addressing these gaps, this work provides a practical and transparent framework for developing effective predictive maintenance solutions aligned with the goals of Industry 4.0.

2. Materials and methods

This section outlines the dataset used in the present study followed by a description of the ML techniques adopted and the metrics of measuring their predictive performance (available at: https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset).

2.1 Description of the dataset

The chosen dataset from the University of California, Irvine – Machine Learning Repository, “AI4I 2020” has 10,000 data points, which is based on working of an actual industrial-grade milling machine. This is a synthetic dataset designed to reflect real predictive maintenance data encountered in industry. It encompasses a synthetic milling process for classification and explainable artificial intelligence (XAI). The dataset consists of the following operational parameters:

1. Air temperature [K] around 300 K with a standard deviation of 2 K.
2. Process temperature [K] around 310 K with a standard deviation of 1 K.
3. Rotational speed [rpm] computed for power output of 2860 W.
4. Torque [Nm] distributed normally around 40 Nm, with a standard deviation of 10 Nm.
5. Tool wear [min] indicating the minutes of process-related tool wear to the employed tool.

For each datapoint with a particular set of input parameters listed above, the machine condition or type of failure is labelled as:

1. No Failure: Indicates normal operation without any faults.
2. Power Failure: Occurs when the power output deviates significantly from the expected operational range, either due to system overload or power disruption.
3. Tool Wear Failure: Triggered when tool wear exceeds its permissible limit, leading to poor machining quality and potential breakdowns.
4. Overstrain Failure: Results from excessive torque combined with prolonged tool wear, typically when processing hard materials under high loads.
5. Random Failures: Represents unanticipated or stochastic breakdowns without clear precursor patterns, possibly due to material inconsistencies or electronic malfunctions.
6. Heat Dissipation Failure: Happens when thermal limits are breached due to insufficient cooling or excessive heat accumulation in the tool or workpiece.

The five input parameters are used to predict the type of failure or whether the machine is working properly or not using ML methods discussed subsequently. The size distribution of different failure types is presented in Figure 2.

To aid better understanding of the data structure, a few representative entries from the dataset are presented in Table 2. Each row represents a snapshot of the machine’s operational condition defined by five input parameters and a corresponding machine failure label.

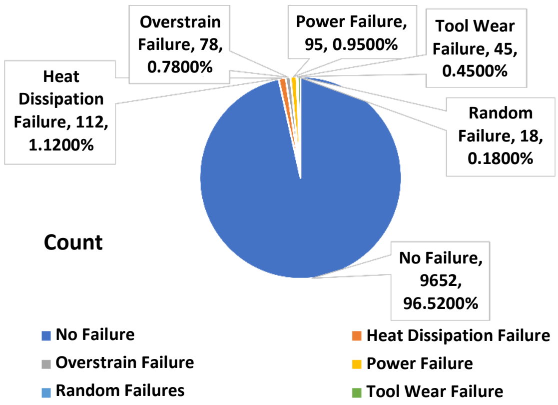

Figure 2 Size distribution of different failure categories

Table 2 Representative examples from the predictive maintenance dataset

Air Temp (K)	Process Temp (K)	Rotational Speed (rpm)	Tool wear (min)	Torque (Nm)	Failure Type
303.5	312.2	1511	27	37.1	No Failure
301.5	311.1	1731	101	27.9	No Failure
300.8	310.1	1405	189	61.2	Power Failure
298	308.7	1268	189	69.4	Power Failure
302.6	311.6	1227	187	68.2	Overstrain Failure
298.4	308.2	1282	216	60.7	Overstrain Failure
297	308.3	1399	132	46.4	Random Failures
298.6	309.8	1505	144	45.7	Random Failures
297.1	308.5	1323	207	44.4	Tool Wear Failure
300.8	310.6	1577	227	37.9	Tool Wear Failure
301.7	309.9	1317	187	49	Heat Dissipation Failure
302.5	310.2	1307	86	54	Heat Dissipation Failure

2.2 ML Classifiers

Five well-established ML classification models, viz., Artificial Neural Networks (ANN), Discriminant Analysis (DA), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Trees (DT) are employed in the present study to investigate their effectiveness in predicting the classification problem for PdM. The dataset was split into 90% for training and 10% for testing. The models were implemented using the Classification Learner Toolbox of MATLAB 2022a®. The process flowchart adopted in the present work to implement the different ML classifiers, the list of hyperparameters optimized and their assessment is described in Figure 3.

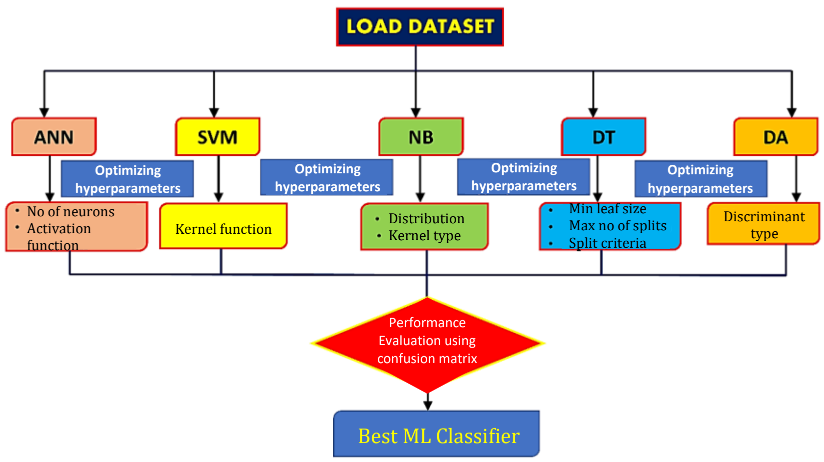

Figure 3 Algorithm of implementation of ML classifiers in the present work

2.2.1 Artificial neural network

ANN involves a computational process of feedforward-backpropagation algorithm to optimize the hyperparameters or unknown coefficients of the mathematical model to arrive at accurate predictions of the required output responses [26-29]. Neural networks consist of certain centers of computations called *neurons* and stages of connections between series of neurons called *layers*. The first layer is the list of inputs, and the last layer is the output or the model prediction. In case of classification problems, output layer computes a classification score to predict the category of the output response for a particular set of input parameters. In this study, a single hidden-layer ANN is optimized with respect to two hyperparameters, viz.: number of neurons between 1 to 100 and the activation function (among ReLU, Sigmoid and Tanh), for the best possible accuracy [30].

2.2.2 Discriminant analysis

Discriminant Analysis (DA) is a supervised ML technique commonly used for classification and dimensionality reduction. It is a simple and reliable algorithm for classification problems employing Bayesian rule to classify the data space into disjoint subdomains using the probability densities [31]. The DA method intelligently optimizes the linear plane to be employed to classify the data space by means of intra-class separability (variance within a group) and inter-class separability (variance between groups) [31-33]. Here, the only hyperparameter optimized is the type of discriminant used in DA, viz.: linear (all classes have same covariant matrix), quadratic (each class can have unique covariant matrix), diaglinear (all classes have same diagonal covariant matrix) and diagquadratic (each class can have a unique covariant matrix, which are diagonal) [33].

2.2.3 Naïve bayes

One of the most straightforward yet efficient supervised classification methods that makes use of the Bayes theorem is the NB method [33]. The algorithm used in the Naïve Bayes classification model assumes that each feature's occurrence stands independent from the others. A conditional probabilistic technique is used by the classifier to build models from a given amount of data in order to learn certain features that belong to a class and make predictions. The model hyperparameters optimized here are the distribution (Kernel or Gaussian) and the type of Kernel (Gaussian, Box, Epanechnikov and Triangle) [34].

2.2.4 Support vector machine

SVM was created in the 1960s, and as it started to gain popularity in the 1990s, it underwent substantial improvements [35]. With superior accuracy and little computational power requirements, it is currently regarded as one of the most effective ML algorithms. SVM is frequently employed for classification goals, while it can also be used for regression. By looking for the best hyperplane or decision boundary (in an N-dimensional space where N is the number of characteristics or input factors), SVM attempts to categorize the dataset into classes or categories [36]. The decision hyperplane's dimension depends on the number of control factors. For instance, if there exist two input features, the hyperplane turns into a line, and its dimension increases with more features. Here, the hyperpameter considered for optimization is the type of Kernel function among Gaussian, Linear, Quadratic and Cubic.

2.2.5 Decision tree

By building a tree-like relationship based on the qualities of the data, decision trees (DT) accomplish categorization of the data. In order to build decision-making rules and identify patterns in data, hierarchical structures are used [37]. They are also used to estimate the relationship between independent and dependent variables. The minimum leaf size, maximum number of splits, and split criteria (which includes Gini's diversity index, twoing rule and largest deviation reduction) are the hyperparameters for optimization in this study.

2.3 Performance evaluation

The confusion matrix is a crucial pictorial representation for the performance assessment of classification models. The confusion matrix provides information on the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. When evaluating an algorithm's accuracy, the True Positive (TP) and True Negative (TN) values reveal how frequently the algorithm labels positive data as positive and negative samples as negative. As a gauge of the algorithm's accuracy, the False Positive (FP) value shows how frequently the algorithm declares a negative sample to be positive. As a gauge of the algorithm's Recall capability, the False Negative (FN) value shows how frequently the algorithm declares a positive sample to be false. Using the TP, TN, FP, FN results, the classification performance of the models are examined using standard metrics of error evaluation. The formulas for the performance metrics: Accuracy, Precision, Recall and F-measure (also called F1-score) are provided in Eqs (1-4), respectively [31-36].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - Measure = \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Accuracy is nothing but the number of correct predictions (TP+TN) as a ratio of the total number of samples. Precision is the probability that the samples (TP+FP) that are predicted to be positive are actually positive. Recall measures the likelihood of the samples (TP+FN) to be predicted correctly. Recall is therefore focusing on whether positive predictions are successfully classified or not. F-measure combines the values of Precision and Recall to by performing harmonic average to produce a single quantity. It is essentially a gauge of a classifier's effectiveness and is frequently used to compare different algorithms. When the Recall rate is desired to be enhanced in most machine learning methods, the Precision value drops as the number of FPs increase. Alternatively, if a higher Precision value is needed, a lower Recall rate will result as the FN number rises. When looking for an algorithm that has good Recall as well as Precision, the F-measure metric is used. Both values must be high for the F1-score to be high, because when the harmonic mean of two numbers, one tiny and the other huge is calculated, the result will be closer to the smaller one.

3. Results and discussion

This section presents the comparison of the results obtained using the various ML techniques to arrive at the model with the best performance.

3.1 Accuracy of the optimized models

The confusion matrices obtained for ANN, DA, NB, SVM and DT classifier models are shown in Figures 4(a-e), respectively. It is found that, ANN model with 10 neurons in the single hidden layer and ReLU (rectified linear unit) activation function shows an overall accuracy of 98.8% (refer Figure 4(a)). It did not predict tool wear failures and random failures, seen by 0% and NaN (i.e., not a number) values in the diagonal. DA showed a maximum of 97.8% accuracy with the quadratic type of discriminant. But it did not correctly predict random failures, which can be seen as 0% in the diagonal entries (refer Figure 4(b)). NB classifier with Gaussian type of Kernel and performed with a maximum accuracy of 96.69%. But it was unable to predict both tool wear failure and random failures, which is evident from the respective diagonal entries (refer Figure 4(c)).

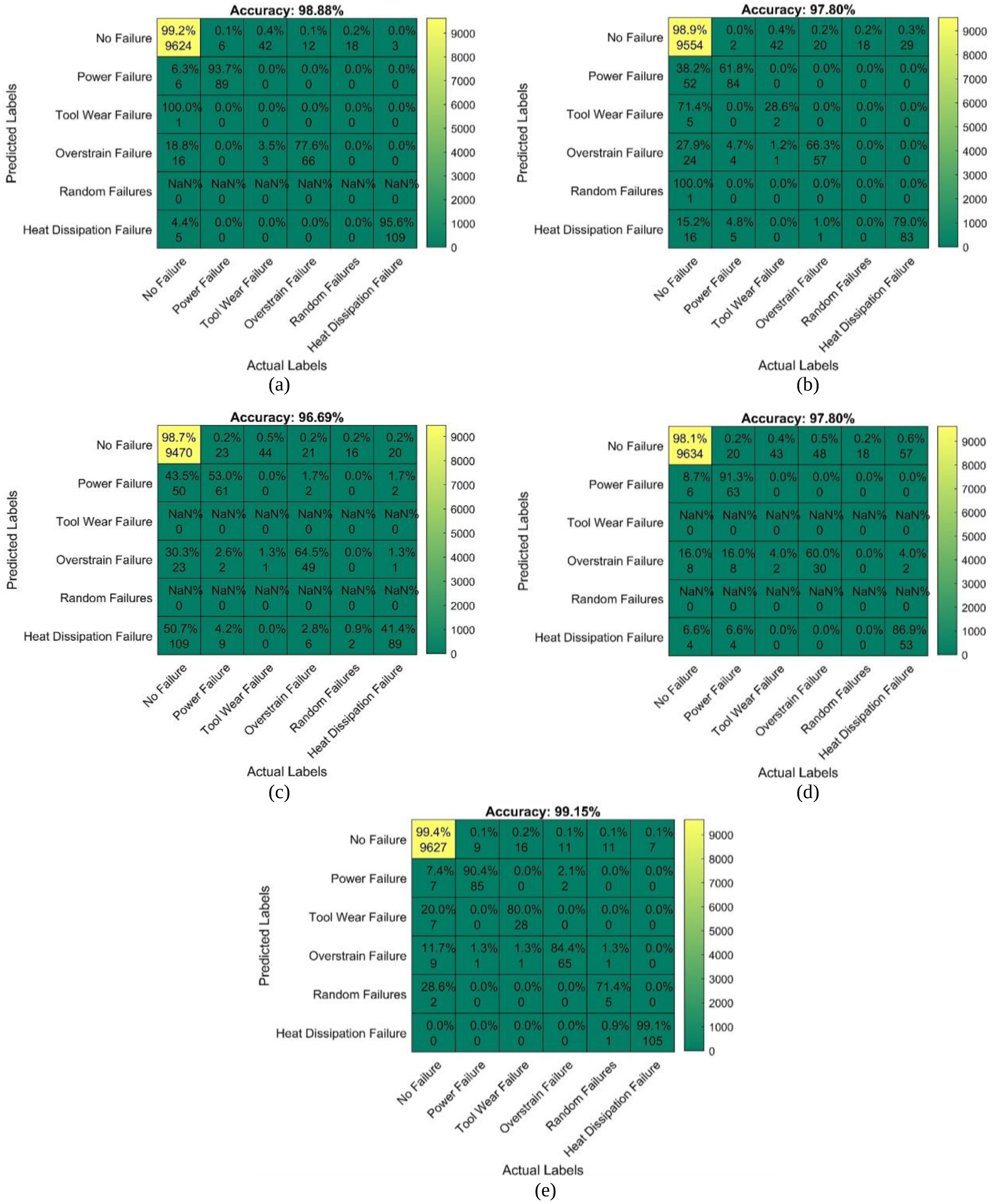

Figure 4 Confusion matrices for (a) ANN, (b) DA, (c) NB, (d) SVM, and (e) DT classifiers

The SVM classifier showed an overall accuracy of 97.8% with a linear kernel function. But similar to NB, it was unable to predict tool wear failure and random failure (refer Figure 4(d)). The DT classifier performed with the best accuracy with a maximum of 99.15% with 391 leaf splits and split criterion of maximum deviance reduction. It is observed from the confusion matrix (refer Figure 4(e)) that all types of failures were predicted effectively by the decision tree classifier demonstrating its superior applicability to be employed in predictive maintenance scenarios and for judicious condition monitoring of industrial milling machines.

Models like Naïve Bayes, ANN and SVM, which rely on strong probabilistic assumptions and density estimation, struggled to learn meaningful patterns for these underrepresented failure types, leading to their misclassification or complete non-detection. Consequently, in some cases, the denominator in Precision or Recall calculations became zero, resulting in NaN values. While other models exhibited NaN values due to their inability to classify the minority failure types, the Decision Tree model achieved classification accuracies ranging from 71.4% to 99.4% for all six failure types, demonstrating its robustness despite class imbalance (Figure 5).

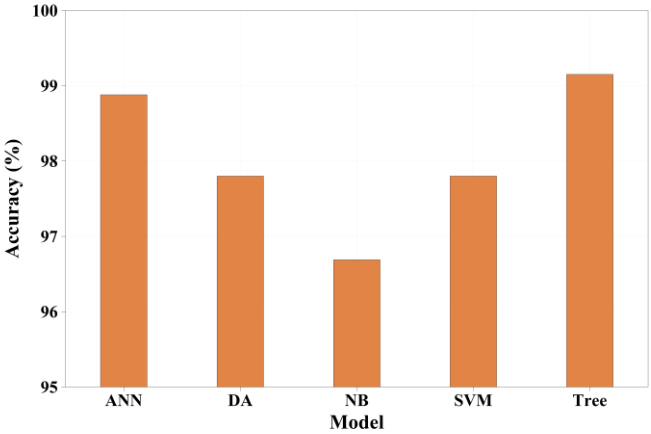

Figure 5 Comparison of accuracy of the optimized models

It is important to note that the predictive maintenance dataset used in this study exhibits a significant class imbalance, with the majority of data points corresponding to the "No Failure" condition and relatively fewer instances representing specific failure types. Such imbalance is common in real-world industrial datasets, where actual machine failures are rare compared to normal operation. While high overall accuracy values are observed, this can sometimes mask poor performance on minority classes, as reflected in the inability of certain models (e.g., NB, SVM) to accurately predict rare failure categories.

To mitigate this challenge and provide a more reliable evaluation, future work will explore the use of data balancing techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE), cost-sensitive learning, or resampling strategies to create balanced training and validation sets. Additionally, confusion matrix evaluations can be performed on balanced validation sets to provide a more comprehensive assessment of model robustness, particularly for minority failure types.

Despite these limitations, the Decision Tree model, as demonstrated in this study, achieved strong classification performance even under imbalanced data conditions, correctly predicting all failure types with reasonable accuracy. This highlights its suitability for PdM applications where imbalanced data is inevitable.

3.2 Performance metrics

The performance metrics which are defined in Eqs. (1-4) give a better understanding of the predictions in comparison to the classification performance of the ML algorithms. From the confusion matrices, the performance parameters, Accuracy, Precision, Recall (also called Sensitivity), and F-Measure (also called F1-score) have been calculated and plotted in Figure 6. The comparative plot clearly demonstrates that the decision tree model (DT) performs the accurately classifies the type of failure based on the condition of the machine and would be able to perform the PdM tasks for condition-based maintenance of the industrial milling machines.

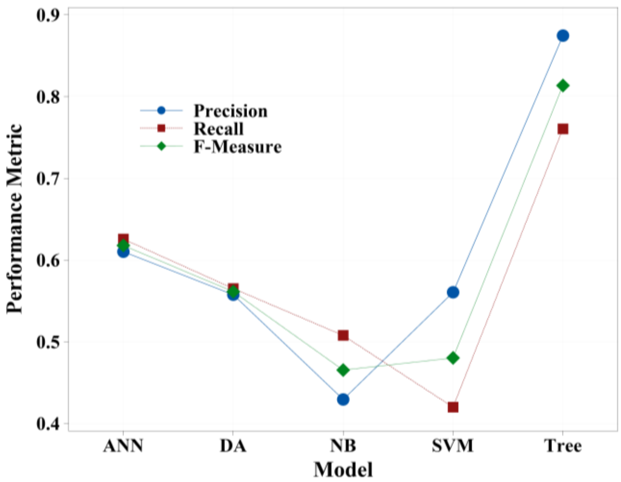

Figure 6 Comparison of classification prediction performance of the different classifiers

Table 3 provides a concise comparative summary of the five machine learning models investigated in this study, combining both their quantitative performance (accuracy) and qualitative characteristics (strengths and weaknesses). The table allows for quick reference to each model's suitability for predictive maintenance applications based on multiple dimensions:

- Accuracy (%) reflects each model's overall classification performance on the test set.
- The strengths column highlights features such as computational efficiency, interpretability, and robustness to class imbalance.
- The weaknesses column identifies limitations like poor performance on minority classes, high training time, or difficulty in model interpretability.

Table 3 Average training and inference time for each ML model

Model	Accuracy (%)	Strengths	Weaknesses
Decision Tree	99.15	Highest accuracy, robust to class imbalance, interpretable	Slightly higher training time
Artificial Neural Network	98.80	High accuracy, flexible	Poor on rare classes, less interpretable
Support Vector Machine	97.80	Good accuracy, robust margin	Fails to detect rare classes
Discriminant Analysis	97.80	Fast, interpretable	Lower precision for failures
Naïve Bayes	96.70	Fastest training/inference	Poor classification of failures

From the table, it is evident that the Decision Tree (DT) model stands out as the most balanced and practically deployable choice — offering the highest accuracy (99.15%), successful classification across all failure types, and ease of interpretation. The Artificial Neural Network (ANN) also performs well but shows reduced reliability for rare failure types and is less transparent. In contrast, models like Naïve Bayes (NB) and Discriminant Analysis (DA) are computationally efficient and interpretable but show limitations in precision and recall, especially under class imbalance.

Overall, this comparative summary reinforces the narrative that while multiple classical ML models are viable for predictive maintenance tasks, their selection should be informed by specific deployment needs, such as accuracy requirements, interpretability, and response time in real-time factory environments.

3.3 Computational efficiency: Training and inference times

In practical predictive maintenance applications, especially those operating under real-time constraints, the computational efficiency of machine learning models is of critical importance. To evaluate the feasibility of real-world deployment, the average training and inference times of each model were recorded using MATLAB's *tic* and *toc* functions on a standard computing platform (Intel® Core^{TM} i7 processor, 2.90 GHz, 16 GB RAM). Table 4 presents the average time taken by each model for training and prediction on the test set (10% of the dataset).

Table 4 Average training and inference time for each ML model

Model	Training time (s)	Inference time (s)
Decision Tree	0.67	0.15
Artificial Neural Network	1.82	0.24
Support Vector Machine	1.25	0.21
Discriminant Analysis	0.31	0.06
Naïve Bayes	0.27	0.05

Naïve Bayes and Discriminant Analysis were found to be extremely fast to train and evaluate, completing both phases in under half a second. ANN and SVM, while achieving strong predictive performance, required more time due to their optimization procedures and complexity. The Decision Tree model, although slightly more computationally intensive than NB and DA, still completed both training and inference under a second, making it a suitable choice for deployment in typical factory settings. Therefore, while high-performing models such as DT and ANN offer excellent accuracy, simpler models like NB and DA may be considered when ultra-low-latency deployment is needed with slightly reduced accuracy.

3.4 Ranking the influence of input parameters

To assess the relative importance of the different governing factors considering in the present analysis on the condition of the milling machine, their comparative influence on the failure type is evaluated against the analysis of variance (ANOVA) algorithm. ANOVA is a statistics-based correlational, filtering and selection technique. By performing univariate statistical tests, it determines which control factors are having higher influence on the output responses. The association between each control factor and the target variable is determined by a one-to-one comparison. Since other features are disregarded while analyzing the relationship between a single control factor (or feature) and the target variable (in the present case, the failure types), this technique is known as univariate. To determine the test score for features, ANOVA utilizes the F-statistic when the input features are numerical [38, 39]. ANOVA determines the test scores for each characteristic and compares them to determine which features are the best.

Figure 7 shows the ranking of the five process parameters considered in the present study on the condition of the milling machine. It is observed that torque, rotational speed, and tool wear have a significant influence on the failure of the machine, while ambient temperature and process temperature have marginal effects. This is especially true in the case of large-scale industrial milling machines wherein they work in well-controlled ambient environmental conditions, in which fluctuations in temperature are minimal. The effect of temperature would become important in situations where the fluctuation in temperature comes into picture [40].

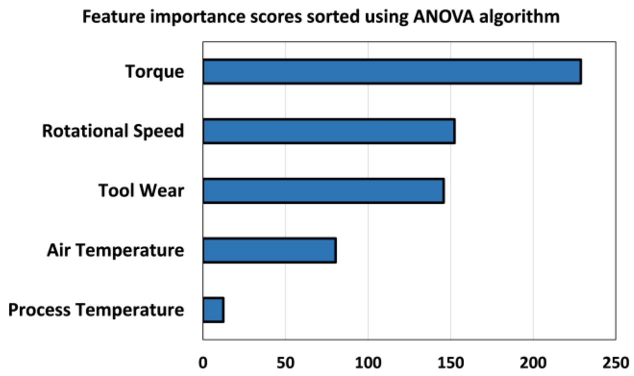

Figure 7 Feature importance ranking using ANOVA

3.5 Comparison with previous studies

To bring forth the contribution of the present study, a literature review was conducted to compare the maximum accuracy of classification achieved in other studies in the literature on the same dataset against that obtained in the present study. Table 5 presents the comparison of the overall accuracy obtained by other studies along with the best accuracy obtained in this paper. It is clearly observed that the Decision Tree model investigated herein achieves the best accuracy with an accuracy of 99.15% (shown in boldface). Although the study by Efeoğlu and Tuna [23] (refer Table 4) showed a close accuracy, their model was unable to predict the random failures, whereas the DT model predicted all types of failures with reasonable accuracy (see Figure 4(e)).

Table 5 Accuracy of classification obtained by previous studies on the same dataset

Author (Year)	Technique	Accuracy
Mota et al. (2023) [41]	Gradient boosting	94.55%
Efeoğlu and Tuna (2022) [23]	SVM (linear kernel)	99.00%
Sengupta et al. (2023) [24]	Ensemble model	98.93%
Ghasemkhani et al. (2023) [25]	Balanced K-Star	98.75%
Present study	**Decision Tree**	**99.15%**

In manufacturing, analyzing welding defects is an important process. Cruz et al. [42] developed a computer vision system based on structured light for welding inspection of liquified petroleum gas (LPG) pressure vessels by using combined digital image processing and deep learning techniques. A convolutional neural network (CNN) based method of inspection prior to the welding and laser triangulation method for post welding inspection is proposed. The proposed process increased the quality index from 95.0% to 99.5%, showing its robustness. Cruz et al. [43] discussed a novel two-step ML approach for dynamic model selection when feedback is not available. The results obtained by applying the proposed approach for predicting surface roughness in micro-milling processes outperform all the individual models evaluated. Specifically, the selected version of the system, which does not include force signals, increased R2 from 0.892 to 0.915 and decreased error from 19.79% to 14.63% when compared to the best individual models' metrics. Beruvides et al. [44] developed a hybrid incremental modeling (HIM) plus simulated annealing (SA) technique applied for predicting the surface roughness in milling processes. Two comparative studies to assess the accuracy and overall quality of the proposed strategy were carried out. The first comparative demonstrated that the proposed strategy is more accurate than conventional methods for predicting surface roughness. The second study also corroborated that hybrid incremental model plus simulated annealing is better than Bayesian network and multilayer perceptron methods for correctly predicting the surface roughness.

Multilayer perceptron (MLP) is one of the most widely applied neural networks at industrial level. However, the main drawback of this approach is related to the setting and tuning of network hyperparameters such as number of hidden layers, number of nodes in the hidden layers, and form of activation functions. HIM + SA has some interesting features such as simple structure and easy training with a few tuning parameters enabled by an optimal setting procedure while outperforming MLP technique [44]. The future scope for improving the present work is the application of such hybrid approaches (HIM + SA) and the coupling of SA and/or Genetic Algorithm (GA) based optimization techniques with deep neural networks to optimize the network parameters [19, 44]. A limitation of the present study is that key techniques such fuzzy systems such as Fuzzy-KNN and neuro-fuzzy systems have not been considered herein. These techniques will be taken into account in further investigations.

One of the key observations in this study is the class imbalance in the dataset, where the majority class ("No Failure") comprises 9,652 instances, while the failure categories have significantly fewer samples (ranging from 18 to 112 instances). This imbalance may contribute to the higher classification accuracy observed for the "No Failure" category compared to the failure types. However, the objective of this study is to evaluate and optimize machine learning models under real-world conditions, where such imbalances are common in industrial predictive maintenance datasets. Moreover, our performance evaluation does not solely rely on accuracy but also considers Precision, Recall, and F1-score, which provide a more comprehensive assessment of model effectiveness across different failure categories. Nevertheless, future work can explore data balancing techniques such as Synthetic Minority Over-sampling Technique (SMOTE), cost-sensitive learning, or advanced ensemble methods to further enhance classification performance for the minority failure classes [45, 46]. Despite these limitations, the current study provides a practical benchmark for machine learning-based predictive maintenance in industrial settings.

Many existing PdM studies focus primarily on overall accuracy without considering the impact of dataset imbalance on minority class prediction. This study highlights both the strengths and limitations of classical ML models under imbalanced conditions and proposes future research directions to incorporate data balancing techniques for more equitable performance evaluation.

The importance of the present work is the attempt in optimizing the hyperparameters of and comparing and assessing five of the most popularly used conventional ML models for the effective predictive maintenance (PdM) of an industrial milling machine based on its process parameters. Also, a statistical evaluation of the process parameters is carried out to identify their relative importance on the machine performance. This study hence can play an important role in the thrust towards smart manufacturing facilities with effective integration between AI, sensors, and data-driven industrial machine systems for making intelligent and timely decisions for predictive maintenance.

While the present study focuses on classical machine learning techniques, it is important to acknowledge the increasing role of deep learning and hybrid intelligent optimization approaches in predictive maintenance research. Methods such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer-based models have demonstrated excellent capabilities in fault diagnosis [47-49], remaining useful life prediction, and system health monitoring due to their ability to automatically extract complex data patterns. However, these methods often require significant computational resources and large training datasets, limiting their suitability for real-time industrial environments where interpretability, low latency, and deployment simplicity are critical.

Furthermore, hybrid approaches, which combine classical ML or deep learning models with intelligent optimization algorithms — such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), or Simulated Annealing (SA) — have shown promise in enhancing model accuracy and robustness while optimizing hyperparameters. Notable examples include the use of Hybrid Incremental Modeling (HIM) with SA or coupling CNNs with evolutionary algorithms for improved PdM performance [50].

Although such advanced methods are not implemented in the present study, the proposed work provides a practical benchmarking foundation that can be directly extended by integrating these emerging technologies in future research. This layered approach allows industries to balance model complexity with operational requirements based on specific application scenarios.

4. Conclusions

The present study investigated the efficiency of multiple popular classification machine learning techniques in a predictive maintenance scenario for an industrial milling machine. The results of the investigation point towards the Decision Tree technique as the most suitable algorithm which can be implemented to identify and predict machine failure and aid in preventive actions during machine condition maintenance. The results obtained in the present study are also compared with those obtained by other researchers to highlight the superior classification accuracy obtained herein. Thus, the results obtained here can serve in building such smart predictive maintenance systems utilizing the available maintenance datasets in industries for improving productivity, agility, and realizing the goals of Industry 4.0.

Although this study emphasizes classical machine learning models for their interpretability and computational efficiency, future work can extend this framework by incorporating deep learning techniques, such as CNNs, LSTMs, or Transformer-based architectures. These methods are capable of automatically learning complex patterns from raw data, potentially improving failure prediction accuracy and enabling more advanced prognostic capabilities. However, appropriate strategies to mitigate their higher computational demands and ensure explainability must be considered to enable their successful adoption in real-world PdM scenarios.

Additionally, exploring hybrid models that integrate classical ML or deep learning techniques with intelligent optimization methods like GA or SA could further improve model performance, robustness, and adaptability. These directions will enhance the applicability of predictive maintenance solutions under diverse industrial scenarios, advancing the goals of smart manufacturing and Industry 4.0.

5. Acknowledgements

We gratefully thank Manipal Academy of Higher Education, Manipal for the support and encouragement provided to conduct this work. The authors have carefully revised the manuscript to improve its language, clarity, and presentation based on the reviewers' feedback.

6. References

[1] Sharma A, Yadava GS, Deshmukh SG. A literature review and future perspectives on maintenance optimization. J Qual Maint Eng. 2011;17(1):5-25.
[2] Faccio M, Persona A, Sgarbossa F, Zanin G. Industrial maintenance policy development: a quantitative framework. Int J Prod Econ. 2014;147:85-93.
[3] Mota B, Gomes L, Faria P, Ramos C, Vale Z, Correia R. Production line optimization to minimize energy cost and participate in demand response events. Energies. 2021;14(2):462.
[4] Ramos C, Barreto R, Mota B, Gomes L, Faria P, Vale Z. Scheduling of a textile production line integrating PV generation using a genetic algorithm. Energy Rep. 2020;6(S8):148-54.
[5] Garg A, Deshmukh SG. Maintenance management: literature review and directions. J Qual Maint Eng. 2006;12(3):205-38.
[6] Shin JH, Jun HB. On condition based maintenance policy. J Comput Des Eng. 2015;2(2):119-27.
[7] Carvalho TP, Soares FAAMN, Vita R, Francisco RP, Basto JP, Alcalá SGS. A systematic literature review of machine learning methods applied to predictive maintenance. Comput Ind Eng. 2019;137:106024.
[8] Udo W, Muhammad Y. Data-driven predictive maintenance of wind turbine based on SCADA data. IEEE Access. 2021;9: 162370-88.
[9] Calabrese M, Cimmino M, Fiume F, Manfrin M, Romeo L, Ceccacci S, et al. SOPHIA: an event-based iot and machine learning architecture for predictive maintenance in industry 4.0. Information. 2020;11(4):202.
[10] Zonta T, da Costa CA, Righi RDR, de Lima MJ, da Trindade ES, Li GP. Predictive maintenance in the Industry 4.0: a systematic literature review. Comput Ind Eng. 2020;150:106889.
[11] Chaudhuri A. Predictive maintenance for industrial IoT of vehicle fleets using hierarchical modified fuzzy support vector machine [Internet]. arXiv [Preprint]. 2023 [cited 2025 Mar 3]. Available from: https://arxiv.org/abs/1806.09612.

[12] Wang L, Zhang C, Zhu J, Xu F. Fault diagnosis of motor vibration signals by fusion of spatiotemporal features. Machines. 2022; 10(4):246.

[13] Abidi MH, Mohammed MK, Alkhalefah H. Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing. Sustainability. 2022;14(6):3387.

[14] Çınar ZM, Nuhu AA, Zeeshan Q, Korhan O, Asmael M, Safaei B. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability. 2020;12(19):8211.

[15] Schiller E, Aidoo A, Fuhrer J, Stahl J, Ziörjen M, Stiller B. Landscape of IoT security. Comput Sci Rev. 2022;44:100467.

[16] Heymann H, Schmitt RH. Machine learning pipeline for predictive maintenance in polymer 3D printing. Procedia CIRP. 2023;117:341-6.

[17] De Luca R, Ferraro A, Galli A, Gallo M, Moscato V, Sperli G. A deep attention based approach for predictive maintenance applications in IoT scenarios. J Manuf Technol Manag. 2023;34(4):535-56.

[18] Karabacak YE. Intelligent milling tool wear estimation based on machine learning algorithms. J Mech Sci Technol. 2024;38:835-50.

[19] Gong L, Chen Y. Machine learning-enhanced loT and wireless sensor networks for predictive analysis and maintenance in wind turbine systems. Int J Intell Netw. 2024;5:133-44.

[20] Dong E, Zhan X, Yan H, Tan S, Bai Y, Wang R, et al. A data-driven intelligent predictive maintenance decision framework for mechanical systems integrating transformer and kernel density estimation. Comput Ind Eng. 2025;201:110868.

[21] Chen C, Shi J, Shen M, Feng L, Tao G. A predictive maintenance strategy using deep learning quantile regression and kernel density estimation for failure prediction. IEEE Trans Instrum Meas. 2023;72:1-12.

[22] Wang L, Cao H, Ye Z, Xu H. Bayesian large-kernel attention network for bearing remaining useful life prediction and uncertainty quantification. Reliab Eng Syst Saf. 2023;238:109421.

[23] Efeoğlu E, Tuna G. Machine learning for predictive maintenance: support vector machines and different kernel functions. J Mach Manuf Reliab. 2022;51:447-56.

[24] Sengupta P, Mehta A, Rana PS. Predictive maintenance of armoured vehicles using machine learning approaches [Interne]. arXiv [Preprint]. 2023 [cited 2025 Mar 3]. Available from: https://arxiv.org/abs/2307.14453.

[25] Ghasemkhani B, Aktas O, Birant D. Balanced K-Star: an explainable machine learning method for internet-of-things-enabled predictive maintenance in manufacturing. Machines. 2023;11(3):322.

[26] Bhat SK, Deepak GD. Predictive modelling and optimization of double ring electrode based cold plasma using artificial neural network. Int J Eng. 2024;37(1):83-93.

[27] Bhat SK, Ranjan A, Upadhyaya YS, Managuli V. Fatigue strength prediction of Cobalt alloys using material composition and monotonic properties: ML-based approach. Mater Res Express. 2025;12(4):046505.

[28] Deepak GD, Bhat SK. Predictive modeling and optimization of pin electrode based cold plasma using machine learning approach. Multiscale Multidiscip Model Exp Des. 2024;7(3):2045-64.

[29] Doreswamy D, Bhat SK, Raghunandana K, Hiremath P, Shreyas DS, Bongale A. ANN-based predictive modelling of the effect of abrasive water-jet parameters on the surface roughness of AZ31 Mg alloy. Manufacturing Rev. 2024;11:21.

[30] Deepak GD, Bhat SK. Optimization of deep neural networks for multiclassification of dental X-rays using transfer learning. Comput Methods Biomech Biomed Eng: Imaging Vis. 2024;12(1):2272976.

[31] Mathew SK, Zhang Y. Acoustic-based engine fault diagnosis using WPT, PCA and Bayesian optimization. Appl Sci. 2020;10(19):6890.

[32] Soofi AA, Awan A. Classification techniques in machine learning: applications and issues. J Basic Appl Sci. 2017;13:459-65.

[33] Owusu KB, Skinner W, Asamoah RK. Acoustic sensing and supervised machine learning for in situ classification of semi-autogenous (SAG) mill feed size fractions using different feature extraction techniques. Powders. 2023;2(2):299-322.

[34] Patil VK, Pawar VP, Kulkarni SP, Mehta TA, Khare NP. Real time emotion recognition with AD8232 ECG sensor for classwise performance evaluation of machine learning methods. Int J Eng. 2023;36(6):1040-7.

[35] Zhu G, Blumberg DG. Classification using ASTER data and SVM algorithms: the case study of Beer Sheva, Israel. Remote Sens Environ. 2002;80(2):233-40.

[36] Kotsiantis SB. Supervised machine learning: a review of classification techniques. Informatica. 2007;31:249-68.

[37] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Boca Raton: CRC Press; 1984.

[38] Gupta SC, Goel N. Performance enhancement of diabetes prediction by finding optimum K for KNN classifier with feature selection method. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT); 2020 Aug 20-22; Tirunelveli, India. USA: IEEE; 2020. p. 980-6.

[39] Moorthy U, Gandhi UD. A novel optimal feature selection technique for medical data classification using ANOVA based whale optimization. J Ambient Intell Human Comput. 2021;12:3527-38.

[40] Luo FJ, Song DL. Study on the influence of ambient temperature on spindle box of the wheel groove milling machine based on ANSYS. Appl Mech Mater. 2013;475-476:1513-6.

[41] Mota B, Faria P, Ramos C. Predictive maintenance for maintenance-effective manufacturing using machine learning approaches. The 17th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2022); 2022 Sep 5-7; Salamanca, Spain. Cham: Springer; 2023. p. 13-22.

[42] Cruz YJ, Rivas M, Quiza R, Beruvides G, Haber RE. Computer vision system for welding inspection of liquefied petroleum gas pressure vessels based on combined digital image processing and deep learning techniques. Sensors. 2020;20(16):4505.

[43] Cruz YJ, Rivas M, Quiza R, Haber RE, Castaño F, Villalonga A. A two-step machine learning approach for dynamic model selection: a case study on a micro milling process. Comput Ind. 2022;143:103764.

[44] Beruvides G, Castaño F, Haber RE, Quiza R, Villalonga A. Coping with complexity when predicting surface roughness in milling processes: hybrid incremental model with optimal parametrization. Complexity. 2017;2017:7317254.

[45] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16(1):321-57.

[46] Ling CX, Sheng VS. Cost-sensitive learning and the class imbalance problem. In: Sammut C, editor. Encyclopedia of Machine Learning. Boston: Springer; 2008. p. 1-8.

[47] Wang R, Dong E, Cheng Z, Liu Z, Jia X. Transformer-based intelligent fault diagnosis methods of mechanical equipment: a survey. Open Phys. 2024;22(1):20240015.
[48] Saeed A, Akram MU, Khattak M, Khan MB. An interpretable hybrid framework combining convolution latent vectors with transformer based attention mechanism for rolling element fault detection and classification. Heliyon. 2024;10(21):e38993.
[49] Fan Z, Li W, Chang KC. A bidirectional long short-term memory autoencoder transformer for remaining useful life estimation. Mathematics. 2023;11(24):4972.
[50] Kobbacy KAH, Jeon J. The development of a hybrid intelligent maintenance optimisation system (HIMOS). J Oper Res Soc. 2001;52(7):762-78.