International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V71I3P108

Volume 71 Issue 3, 50-56, March 2023

Review Article

Machine Learning-Based Predictive Maintenance of Industrial Machines

Chaitali R. Patil[1], Sanika K. Jadhav[2], Asmeeta L. Bardiya[3], Ankita P. Davande[4], Mahee P. Raverkar

[12345]*Department of Computer Engineering, K. K. Wagh Institute of Engineering Education & Research, Nashik(SPPU) Maharashtra, India*

Received: 02 February 2023 Revised: 13 March 2023 Accepted: 23 March 2023 Published: 31 March 2023

Abstract *- IoT (IIoT) enables machines, people, cloud computing and analytics to function together, improving the performance and productivity of industrial processes. Earlier maintenance of industrial machines used methods such as periodic maintenance (PM) or condition-based maintenance (CBM). PM often leads to a waste of personnel and material since, in many cases, maintenance is unnecessary and could be postponed. CBM requires a high level of expert knowledge to define the threshold values. In contrast to these approaches, predictive maintenance (PdM) is a more efficient and effective maintenance approach involving monitoring the state and health of industrial machines to identify potential failures and threats before they create a mishap leading to severe property loss, production, life, etc. Prediction requires machine learning models based on large amounts of data for each system component. The proposed system acquires data about machine health through a standard input mechanism. The system will collect, store and send audio data for processing in the edge devices as well as the cloud. It will use sensor data visualization tool to analyse system health. Machine health reports will be periodically generated, and tools to forecast machine failures will be used. The system generates alarms and sends notifications to the concerned officials of the industry. The proposed methodology will achieve real-time computing prediction of failures of industrial machines. Industry 4.0 focuses on process optimization, reducing costs and increasing efficiency, which imbibes a major motivation.*

Keywords *- Audio Data, Data Visualization, Industrial IoT (IIoT), Machine Learning, Predictive Maintenance (PdM), Sensors.*

1. Introduction

In industries, machinery requires continual maintenance for proper working. Industries follow various maintenance approaches for their machinery, depending on their specific needs and constraints. Here is a summary of some common maintenance approaches:

- Reactive Maintenance: This approach involves waiting for a piece of equipment to fail and then repairing or replacing it. Reactive maintenance is typically the least expensive option in the short term, but it can lead to increased downtime and higher costs in the long run.

- Preventive Maintenance: This approach involves performing maintenance tasks on a fixed schedule, such as replacing parts or conducting inspections, regardless of whether or not the equipment is showing signs of wear or malfunction. Preventive maintenance can help reduce the likelihood of unexpected failures but can also lead to unnecessary maintenance and downtime.

- Predictive Maintenance: This approach involves using data and analytics to predict when a piece of equipment is likely to fail and then taking action to prevent the failure from occurring. Predictive maintenance can help minimize downtime and reduce maintenance costs, but it requires significant data collection and analysis investment.

Nowadays, many companies still follow a periodic (PM) or condition-based maintenance (CBM) approach. While in PM, industrial machines are maintained at regular intervals, CBM involves defining threshold values for particular sensors, which trigger maintenance operations when exceeded. PM often leads to a waste of personnel and material since, in many cases, maintenance is not necessary and could be postponed [7].

Among all the maintenance approaches, predictive maintenance using machine learning is a powerful tool that can help organizations improve equipment reliability and reduce costs associated with unplanned downtime. Predictive maintenance is a technique used to predict equipment failure and prevent unplanned downtime by analyzing data from various sources, including sensors, historical maintenance records, and other operational data. Machine learning is an important tool used in predictive maintenance to automatically identify patterns and anomalies in this data, which can help identify potential issues before they become critical. Machine learning models can be trained using historical data to identify patterns that indicate equipment failure is likely to occur. These models can then predict when a piece of equipment is likely to fail and alert maintenance teams to take action before the failure occurs. Using machine learning in predictive maintenance can lead to significant cost savings by reducing downtime, improving equipment reliability, and extending equipment lifespan. It also enables maintenance teams to focus their efforts on the most critical issues rather than spending time on routine maintenance tasks.

In this research work, we propose using various machine learning algorithms like Random Forest, Support Vector Machine (SVM), Decision Tree, etc., for predicting the type of failure that can arise in the machine and comparative analysis of these algorithms.

The rest of this paper is structured as follows: In Section 2, we briefly describe related works in the domain of PdM and machine failure prediction. Then, we present our proposed PdM methodology for failure prediction in Section 3. Section 4 presents our experimental results and performance analysis in terms of feature selection. Finally, we conclude the paper with a summary in Section 5.

2. Literature Survey

Several studies were published previously on the predictive maintenance of industrial machines by various researchers using various techniques, as summarized below briefly.

Gian Antonio Susto has proposed a methodology that has been demonstrated for a semiconductor manufacturing implant-related maintenance task and shown to guarantee better performance than classical PvM approaches and a single SVM classifier distance-based PdM alternative [1]. This case study has also shown that SVMs offer superior performance to k-NN classifiers when implementing MCPdM and that, in general, MC-PdM-known also consistently outperforms PvM approaches.

The Industrial Internet of Things is used to gain valuable insights from machine information along with the data kind components for predictive modelling [2]. The researchers have explored using autoregressive integrated moving average forecasting trauma plating machines to predict quality defects, downtime and maintenance. Machine learning has been proven to be an important component in the industrial Internet of Things for quality management and quality control. It enhances performance and improves the manufacturing process.

In another study by Marina Paolanti, the methodology was implemented in the experimental environment on the example of a real industrial group, producing accurate estimations [3]. Data has been collected by various sensors, machine PLCs and communication protocols and made available to the Data Analysis Tool. The proposed PdM methodology allows dynamical decision rules to be adopted for maintenance management, which is achieved by training a Random Forest approach on Azure Machine Learning Studio. Preliminary results show a proper behaviour of the approach on predicting different machine states with high accuracy (95%) on a data set of 530731 data readings on 15 different machine features collected in real time from the tested cutting machine.

Similarly, Weiting Zhang and other authors proposed a methodology demonstrated as a universal end-to-end predictive maintenance methodology for the time-to-failure prediction of industrial machines [4]. This includes a universal sensor handling and feature extraction approach based on integral values, feature transformation, feature selection, adjustable target class labelling, training of different machine learning models, and their hyperparameter tuning. The results show that a feed-forward neural network with multi-class labelling managed to achieve the best prediction quality in terms of accuracy, F1-score, and kappa of 97.79% and 94.03%, respectively.

K. Liulys, in his study, shows that the process of regularly checking the equipment is performed so that the machine does not break down. However, preventive maintenance costs many irrelevant and repetitive checking [5]. It leads to greater costs that cannot be afforded frequently. Therefore, a new concept called predictive maintenance can be implemented in such machines to reduce the downtime cost and decrease the number of checking. Through predictive maintenance, the total time required for all the checking and the total cost spent on all the checking is reduced tremendously. Machine learning is the process of predicting output values by implementing programs on huge gathered data of core values.

Farzam Farbiz has introduced a cognitive analytics-based framework for machine condition monitoring and anomaly detection. We applied the proposed framework to an industrial robot use case and validated the proposed approach [6]. The machine model generated by the proposed framework can learn the machine performance adaptive with the new data and detect anomalies of the robot movement in real-time. Although the proposed solution has demonstrated promising results and its unsupervised machine learning model can classify the data despite noise and outliers, it should be noted that it has been applied to one use-case study so far.

In a more recent study, the paper's authors conducted a survey of PdM of industrial equipment [7]. Initially, a brief introduction of the industrial PdM scheme was proposed, demonstrating the challenges faced. The paper's main aim is to provide a systematic overview of the PdM, propose an industrial PdM scheme for automatic washing equipment, and demonstrate the challenges faced when conducting a PdM research study.

Condition monitoring has been prevailing in industrial machinery failures for years. The idea of protecting against failures has been in the industries to reduce the cost of machine maintenance. [8] the paper describes a system that collects data from 30 industrial pumps at a thermochemical plant. This data is gathered and processed through the Random Forest Algorithm to establish relevant information. The paper has articulated the challenges that arise while implementing machine learning algorithms on the data and Systems performance.

The Industrial Internet of Things can be defined as the internet of things technologies used in the manufacturing processes to harness the data gathered from the sensors implemented on the machinery [9]. The Industrial Internet of Things is used to gain valuable insights from machine information along with the data kind components for predictive modelling.

A study by Amruthnath, N. and Gupta shows fault detection has been an important subject prevailing in industries in recent years [10]. Detecting faults is important to reduce the break time cost early and ensure proper running of the manufacturing processes. The research paper has proposed unsupervised learning for predictive maintenance by the system on the exhaust fan. Algorithms such as PCAT2 statistic, Fuzzy C-Means clustering, Hierarchical clustering, K-Means, and model-based clustering are used by the researchers for predictive maintenance. These algorithms are used, and the best one suitable for predictive maintenance has been proposed to ensure robustness.

3. Proposed Approach

3.1. Functional Requirements

3.1.1. Data Storage

The data for predictive maintenance is stored and collected from the Thingspeak IOT cloud in a time-series format. The data is stored in two types of formats:

- Main Storage: A data lake is used to store the data through an online IOT cloud. The stored here is used for pre-processing and analysing the data.
- Backup File: A readily available dataset from kaggle is stored in csv format and is used for developing a machine learning model and analyzing the data.

3.1.2 Data Transfer

It involves data movement from the cloud to the proposed system for analysis. The data transfer occurs through a secure communication channel provided by the IOT cloud.

3.1.3 Data Pre-Processing and Processing

- Data Cleaning: Data cleaning is performed as the raw data may contain irrelevant or undesired elements. It manages the feeding of missing data and noisy values through procedures such as outlier detection and others.
- Data Transformation: Data Normalization is performed to scale the data values into the desired range. In this phase, attributes are selected as per suitability, and new attributes from existing or known attributes are generated.
- Data Processing: In this stage, machine learning algorithms are applied to process the data. Random Forest, SVM, and Decision Tree are used for predicting the data using values such as pressure, volume, temperature and others.

3.2. System Architecture

The system architecture for the predictive maintenance system consists of several components such as industrial machines, Thingspeak IOT cloud for storage and a user interface screen.

The edge device is used to collect the sensing data of the industrial machine. Thingspeak cloud analytic platform stores collected raw data in time-series format. A local dataset from Kaggle is used in case of internet connectivity issues. The edge device has been programmed using libraries such as Flask, Scikit-Learn and Python 3.6 Programming Language. UCI Machine Learning Repository has been used while implementing the system and gathering the data from the industrial machine.

The data collected from the IOT cloud will be given to the model development part, which is processed and trained using machine learning algorithms such as Random Forest, Support Vector Machine, Decision Tree and K-Nearest Neighbors algorithm, and the predicted results will be displayed on the user interface screen.

The proposed methodology will discuss the system’s behaviour and maintenance schedule through predictive analysis. The system will be continuously assessed through the gathered data on pressures, vibrations, temperatures, power consumption others. Random Forest Algorithms and other algorithms will be implemented on the gathered data to predict their behaviour. Through these algorithms, data will be processed to obtain the changes in the values of parameters like vibration, volume, etc., so that the breakdown period of any equipment can be predicted.

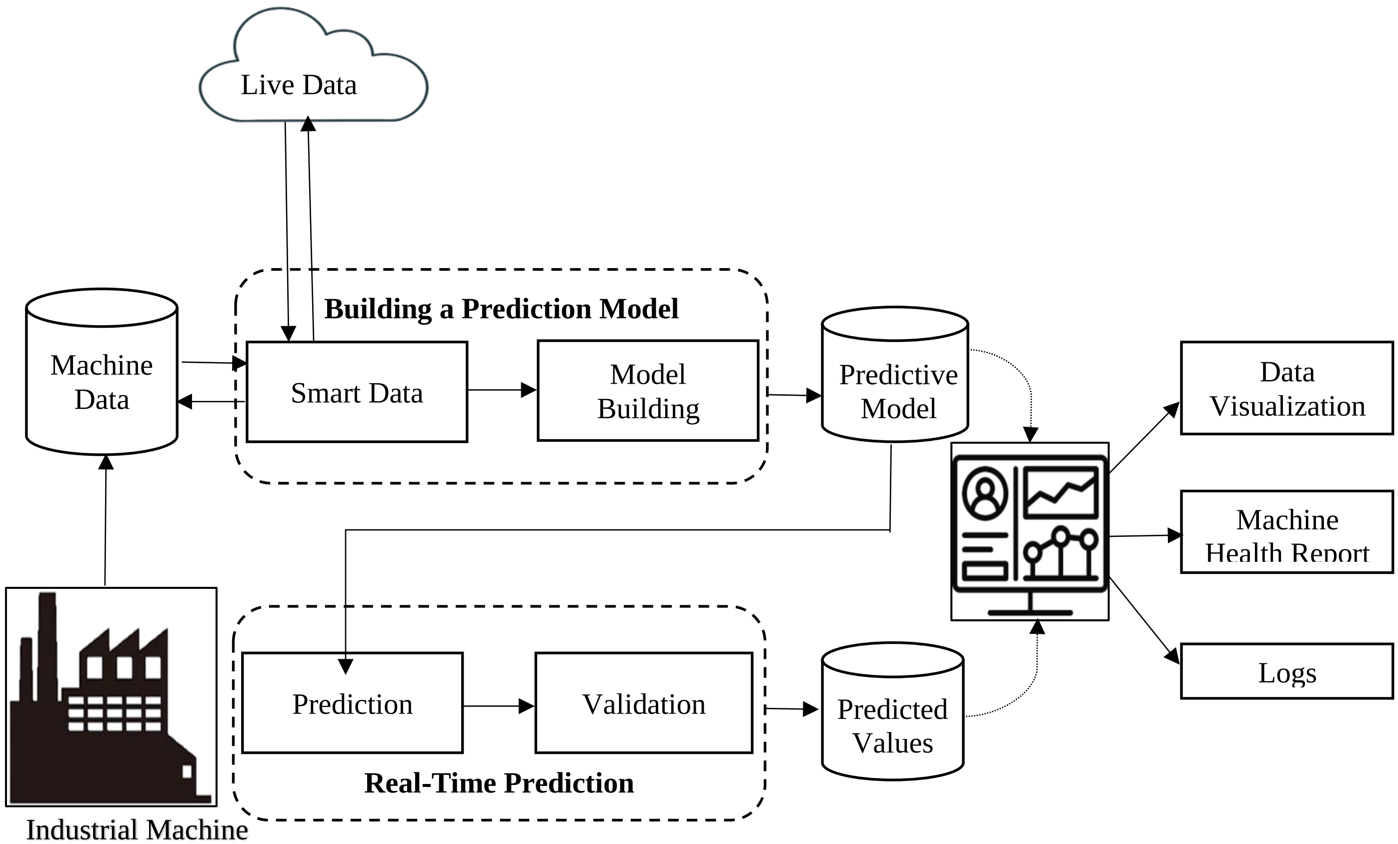

Fig. 1 System Architecture

3.3. Algorithm

3.3.1. Support Vector Machine

The SVM model is typically used to handle binary classification jobs. SVMs have been extensively used in industrial equipment to identify a particular status based on the acquired signal. The SVM model can also be used to complete multiclass jobs due to the variety of fault types and the capability of mapping low-dimension features to hyperplanes. In conclusion, the main goal of SVM is to locate a hyperplane and split data points appropriately on both sides of the hyperplane, and the optimization object is symbolized by

$$\text{argmax}\ (\text{w}, \text{b}) \left\{\frac{1}{||w||} min[y_t(w^T \cdot x_t + b)]\right\}$$

$$s.t\ \ y_t(w^T \cdot x_t + b) \geq 1 \qquad (1)$$

where (x_t, y_t) refers to a sample that contains features and labels.

SVMs work by finding the best hyperplane that separates the data into different classes. The algorithm tries to maximize the margin between the hyperplane and the closest data points from each class. SVMs also use a kernel function that transforms the data into a higher-dimensional space, where it may be easier to separate the data points.

3.3.2. Decision Tree

Decision trees can be used in the predictive maintenance of industrial machines to identify potential faults before they occur. Decision tree models are a type of machine learning algorithm that can be used for classification and regression tasks. They work by partitioning the input data into a set of binary decisions that split the data into smaller subsets until a terminal node is reached, which provides the final prediction.

To use decision trees in predictive maintenance, historical data on machine performance and maintenance records can be collected and used to train a decision tree model. The model can then be used to predict when maintenance is needed based on real-time data collected from sensors on the machine.

The primary problem when implementing a decision tree is choosing the best attribute for the root node and sub-nodes. So, a method known as attribute selection measure, or ASM, can be used to address these issues. Using this measurement, we can easily choose the ideal characteristic for the tree nodes. There are two widely used ASM methods, which are as follows:

1. Information Gain: After segmenting a dataset based on an attribute, information gain measures variations in entropy. It figures out how much knowledge a characteristic gives us about a class. An indicator of the impurity in a particular attribute is entropy. It defines data unpredictability. Entropy is computed as

$$H = -\sum_{t=1}^{n} p(x_t) \log_2 p\,(x_t) \qquad (2)$$

2. Gini Index: When using the CART (Classification and Regression Tree) method to create a decision tree, the Gini index is a purity or impurity indicator. It is preferable to have a trait with a low Gini index than one with a high Gini index.

$$Gini = 1 - \sum_{t=1}^{n} p_t^2 \qquad (3)$$

3.3.3. Random Forest Algorithm

A machine learning method called Random Forest is frequently employed for predictive maintenance. It is an example of an ensemble learning method that combines different decision trees to increase prediction accuracy.

Based on past data, Random Forest can be used in predictive maintenance to forecast machine failures or breakdowns. The algorithm analyses large amounts of data from sensors, devices, and other sources to find patterns and trends that point to possible issues. The algorithm then makes predictions about potential failures or breakdowns using this knowledge.

One of the key advantages of using Random Forest for predictive maintenance is its ability to handle large, complex datasets with many variables. The algorithm can handle numerical and categorical data and automatically select the most important features for making predictions.

To implement Random Forest for predictive maintenance, you will need to:

- Collect and clean historical data on equipment failures and maintenance activities.
- Identify relevant variables and features that may be predictive of future failures.
- Train the Random Forest model using the historical data.
- Test the model on new data to evaluate its accuracy and performance.
- Use the model to generate predictions about future failures or breakdowns.

3.3.4. Logistics Regression

A dataset can be analysed using the statistical technique of logistic regression if one or more independent factors affect the result. It is frequently applied to classification issues in machine learning and data analytics, where the result is a binary (yes or no) or categorical variable.

Based on the independent factors, the logistic regression model calculates the probability of the outcome variable. It is a kind of generalized linear model that converts the linear output of the independent factors into a probability value between 0 and 1 using the logistic function.

The linear regression equation yields the logistic regression equation. The following are the mathematical methods to obtain Logistic Regression equations:

1. We know the equation of the straight line can be written as:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \qquad (4)$$

2. In Logistic Regression, y can be between 0 and 1 only, so for this, let's divide the above equation by (1-y):

$$\frac{y}{1-y}; 0\ for\ y = 0\ and\ infinity\ for\ y = 1 \qquad (5)$$

3. But we need a range between -[infinity] to +[infinity], then take the logarithm of the equation, it will become:

$$\log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \qquad (6)$$

The above Equation 6 is the final equation for Logistic Regression.

4. Results and Discussion

4.1. Results

We have successfully built a predictive system prototype capable of sending data from the Thingspeak cloud on the edge device.

Using a predictive maintainenace dataset from kaggle, the condition of the VMC (Vertical Machining Centre) machine, such as air temperature, rotational speed, torque, tool wear and process temperature, were monitored.

Different machine learning algorithms train the model on a given data set. Given below are the types of failure that may occur in a VMC Machine:

1. tool wear failure (TWF)
   the tool will be replaced or fail at a randomly selected tool wear time between 200 – 240 mins.

2. heat dissipation failure (HDF)
   heat dissipation causes a process failure if the difference between air- and process temperature is below 8.6 K, and the tool's rotational speed is below 1380 rpm.

3. power failure (PWF)
the product of torque and rotational speed (in rad/s) equals the power required for the process. If this power is below 3500 W or above 9000 W, the process fails.

4. overstrain failure (OSF)
if the product of tool wear and torque exceeds 11,000 minNm for the L product variant (12,000 for M, 13,000 for H), the process fails due to overstrain.

The performance of the system is evaluated using the confusion matrix

4.1.1. Confusion Matrix

The performance of the classification models for a specific collection of test data is evaluated using a matrix called the confusion matrix. Important predictive metrics like recall, specificity, accuracy, and precision are visualized using it. Because they provide clear comparisons of values like True Positives, False Positives, True Negatives, and False Negatives, confusion matrices are helpful.

1. Accuracy - Accuracy is the most intuitive performance measure, and it is simply a ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

2. Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. (Precision = TP/TP+FP)

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

3. Recall - Recall is the ratio of correctly predicted positive observations to all observations in the actual class. (Recall = TP/TP+FN)

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

4. F1-score - f1-score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

The bulk of prediction was performed on the product having the L category. Our main task was to find whether the VMC machine would fail or not and what will be the type of failure that may occur in case the machine failed.

Based on the analysis shown in Table 1, Random Forest Algorithm works based on multiclass classification and achieves maximum accuracy. After Random Forest, the Decision tree algorithm gives maximum accurate predictions. It has been observed that Support Vector Machine(SVM) algorithms have less accurate results and cannot perform multiclass classification well.

Table 1. Analysis of Algorithm

Algorithm	Accuracy	Precision	Recall	F1
Random Forest	99.24	96.62	82.7	89.1
Decision Tree	98.06	71.77	77.11	71.9
Logistics Regression	96.2	96.2	96.2	96.2
Support Vector Machine	96.15	96.15	96.15	96.2

5. Conclusion

The review of prior research work conducted on predictive maintenance using machine learning is discussed in this paper. The conclusion drawn from the overall discussion is that implementing a predictive maintenance system can lower the costs associated with breakdowns in the various sectors.

The research has been successfully conducted for the maintainenace of VMC Machine on current data. The proposed system discussed can detect 4 types of faults in the VMC Machine. Thus, it can be concluded that the system can reduce the cost of maintenance and maximise an equipment's life. Thus, it can be concluded that the system can reduce the cost of maintenance and maximise an equipment's life.

References

[1] Gian Antonio Susto et al., "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812-820, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ameeth Kanawaday, and Aditya Sane, "Machine Learning for Predictive Maintenance of Industrial Machines Using IoT Sensor Data," *in 2017 8th IEEE International Conference on Software Engineering and Service Science*, pp. 87-90, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Marina Paolanti et al., "Machine Learning Approach for Predictive Maintenance in Industry 4.0," *IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications*, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Weiting Zhang, Dong Yang, and Hongchao Wang, "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey," *IEEE Publication*, vol. 13, no. 3, pp. 2213-2227, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kaorlis Liulys, "Machine Learning Application in Predictive Maintenance," *2019 Open Conference of Electrical, Electronic and Information Sciences (es- Tream)*, pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Farzam Farbiz, Yuan Miaolong, and Zhou Yu, "A Cognitive Analytics Based Approach for Machine Health Monitoring, Anomaly Detection, and Predictive Maintenance," *15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1104-09, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Marwin Zufle et al., "A Predictive Maintenance Methodology: Predicting the Time-to- Failure of Machines in Industry 4.0," *IEEE 19th International Conference on Industrial Informatics (INDIN)*, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ido Amihai et al., "An Industrial Case Study Using Vibration Data and Machine Learning to Predict Asset Health," *2018 IEEE 20th Conference on Business Informatics*, pp. 178-185, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] P. Strauß et al., "Enabling of Predictive Maintenance in the Brownfield Through Low-Cost Sensors, an IIoT-Architecture and Machine Learning," *IEEE International Conference on Big Data (Big Data)*, pp. 1474-83, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Nagdev Amruthnath, and Tarun Gupta, "A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance," *5th International Conference on Industrial Engineering and Applications*, pp. 355-361, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] B. Balaji et al., "Fault Prediction of Induction Motor Using Machine Learning Algorithm," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 8, no. 11, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] R. Yam et al., "Intelligent Predictive Decision Support System for Condition-Based Maintenance," *The International Journal of Advanced Manufacturing Technology*, vol. 17, no. 5, pp. 383–391, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[13] Archit P. Kane et al., "Predictive Maintenance Using Machine Learning," *ArXiv*, 2022. [CrossRef] [Publisher Link]

[14] Ayushi Chahal, and Preeti Gulia, "Deep Learning: a Predictive IoT Data Analytics Method," *International Journal of Engineering Trends and Technology*, vol. 68, no. 7, pp. 25-33, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Wo Jae Leea et al., "Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data," *CIRP Life Cycle Engineering (LCE) Conference*, vol. 80, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Joel Anto Williams N et al., "Machine Predcitive Mainteneace System for Industrial Applications," *International Journal of Current Research*, vol. 14, no. 05, pp. 21410-21412, 2022. [CrossRef] [Publisher Link]