Enhanced Equipment Reliability Through Predictive Maintenance Modeling: A Machine Learning Approach

Author: Zhe Wee Ng

UC Berkeley, 387 Soda Hall, Berkeley, CA 94720, USA

Email: derrickngzhewee@gmail.com*, ngzhewee@berkeley.edu

Tel: 1-8586251251

Zhe Wee Ng (Derrick) is an undergraduate Computer Science student at UC Berkeley, graduating in May 2025. His research focuses on leveraging Artificial Intelligence and Machine Learning to address real-world challenges through interdisciplinary applications in advanced manufacturing, sustainability, biomedical research, and generative AI. Derrick's projects include predictive maintenance modeling to optimize industrial reliability, NLP techniques for sustainable design insights, single-cell data analysis for biomedical innovation, and diffusion-based 3D video generation for adaptive systems. He is particularly passionate about applying advanced AI techniques to develop scalable, data-driven solutions that enhance decision-making and system efficiency in dynamic environments.

Abstract: The emergence of predictive maintenance (PdM), which moves away from corrective maintenance methods, represents a significant advance in improving operational efficiency and effectively reducing costs. The presented paper deals with an investigation using state-of-the-art machine learning (ML) methods. Specifically, it focuses on a Long Short-Term Memory (LSTM) network for the implementation of PdM and the subsequent prediction of equipment failures to minimize unexpected production downtime. This study conducts an analysis of the existing literature examining different models and their application in different industries, emphasizing the growing importance of PdM in today's context. Their research approach is quite

thorough as they discuss the process of data collection and preparation prior to the implementation of LSTM networks for forecasting purposes. It appears that LSTM outperforms standard models in terms of effectiveness and accuracy in predicting the lifetime of equipment, making it a robust solution to the challenges of predictive maintenance challenges. A performance comparison of LSTM with traditional models shows that it has greater accuracy and independence in predicting equipment failures, making it a viable solution to predictive maintenance problems. The results of this study represent an important contribution to research by presenting a realistic ML-based predictive maintenance framework that demonstrates its potential to reduce costs and improve operational efficiency. In addition, this study provides new areas for future research, such as investigating other ML models and using real-time data to increase predictive accuracy. The results of this study highlight the revolutionary power of advanced predictive approaches to improve maintenance tactics and enhance industrial practices.

Keywords: *LSTM; PdM; RNN; Convolution LSTM; Bi-LSTM; Predictive Maintenance;*

1. Introduction

Maintenance accounts for a large proportion of production costs in industry and is critical to maintaining operational efficiency and equipment life. Effective maintenance techniques can drastically reduce costs, avoid unexpected production downtime and extend the life of machinery. As a result, the attention of industry professionals has shifted to improving maintenance methods, which has led to the development of numerous maintenance techniques. In the past, corrective maintenance (CM), sometimes referred to as “run-to-failure”, was the most common strategy, where repairs or replacements were only made after equipment failure, resulting in costly production downtime [1]. The enormous costs associated with these unforeseen failures prompted the development of proactive maintenance solutions. Preventive maintenance (PM) was developed as a solution. It involves scheduled inspections and replacement of parts at regular intervals, especially after a certain number of operating cycles as prescribed by the manufacturer. However, PM posed problems, such as premature replacement of working components, which increased maintenance costs, or delayed replacement, which required costly remedial action. The introduction of the Internet of Things (IoT) and Industry 5.0 technologies changed maintenance processes and led to the introduction of condition-based maintenance (CBM). This technique uses sensors and monitoring devices to automate inspections and enables maintenance interventions based on real-time data when equipment performance deviates from set values. Building on the CBM concepts, predictive maintenance (PdM) has emerged as a more complex and successful technology. PdM combines cyber-physical systems (CPS), IoT, robotics, mechatronics, and information technology, and data science to forecast equipment failure and predict the remaining useful life (RUL) of resources, facilitating maintenance planning that minimizes disruptions [2].

In addition, prescriptive maintenance goes beyond PdM by not only predicting failures, but also proactively recommending ways to improve RUL and address specific failure modes. For example, while PdM can predict impending equipment failure due to elevated bearing temperatures, prescriptive maintenance can recommend operational changes, such as reducing speed to extend equipment life [3]. Despite the benefits of prescriptive maintenance, its successful implementation depends on the robust integration of PdM and CBM, as accurate predictions rely on well-established CBM frameworks. In this framework, the contributions of this study are twofold:

i. To discover the existing methodologies of predictive maintenance
ii. Propose predictive maintenance solutions to reduce equipment downtime and increase productivity in the industry.

iii. To evaluate the effectiveness of the proposed ML model for predictive maintenance in terms of timely equipment failures.

Further, section 2 discusses the literature studies of recent work, and Section 3 discusses proposed methodology. Section 4 shows the result of the performance parameter. Formerly, Section 5 has a result discussion. Thereafter, the conducted work of this article has been concluded in the Section 6 with future possibilities.

2. Literature Review

In this section, the existing state-of-the-art methods are elaborated to examine the overall structure of proposed domain.

2.1 Overview of Predictive Maintenance Strategies

By using PdM, consumers can receive notifications about when their car needs its next scheduled maintenance, when parts need to be repaired, when warranties or warranties expire, when engine oils need to be changed, when wheels need to be rotated and other important details. Therefore, a driver or user's vehicle will always be in better condition if they follow the predictive maintenance schedule instead of carelessly using and ignoring the vehicle [4]. Based on the data collected by the IoT sensors, analytical maintenance of the vehicle can be performed. Using blockchain technology, this can be linked to incentives so that the user benefits financially by following the predictive maintenance schedule and receiving discounts or prizes [5]. The manufacturer, the insurance provider and the user can all benefit from the blockchain-based solution. The insurance provider can impose penalties or rewards depending on the user's compliance or non-compliance with the maintenance plans. The blockchain's decentralized ledger, decentralized data management and cryptographic primitives and protocols guarantee the immutability, transparency and security of the data in this system. Automated cars with IoT connectivity have features that predict when the car needs to be serviced. This could be thought of as a hardware card that can record all the parameters of the system. The database prompts the user when a parameter exceeds an uncontrollable threshold. The system is also able to predict the general state of the system. It is significant to know that Hyundai Motors' Bluelink technology records every system parameter [6], [7].

Everything that happens in the entire history of the vehicle - both good and bad - is recorded. The database also records all cases in which the driver has exceeded the speed limit or failed to observe safety instructions. Not only is it recorded, but it also indicates what needs to be done next, enabling predictive maintenance. The chip-based predictive maintenance system collects vehicle data and is connected to the cloud. Once processed, the data can be used to predict which part needs to be serviced. Predictive maintenance has become smarter with the advent of artificial intelligence. It is also able to predict a driver's driving performance [8].

2.2 Existing Studies on Predictive Maintenance and Machine Learning

One of the most vital components of Industry 5.0 is the processing of sensor data to enable better decision-making [9]. However, the use of decision-making algorithms is hindered by the ambiguity of predictive analytics, the lengthy procedure and the time limits to reach a conclusion. As PdM has become more popular and effective over the last century, there has been a surge in the development of techniques that can improve maintenance decision making to extend the life of machinery.

Ansari et al. [10] have proposed an integrated strategy that combines Dynamic Bayesian Network (DBN) and a data model to predict future events, identify cause-effect relationships and offer improvements for maintenance planning. A novel architecture for a PHM method was presented by Sarazin et al. [11] in an attempt to improve the extraction of data values. This lambda design consists of 2 layers: a velocity layer and a load layer. The speed layer can apply maintenance policies and insights from machine learning algorithms over the memory layer. In addition, the system must process a large amount of data, i.e. it must handle big data problems while maintaining system compatibility. Cheng et al. [12] used advanced technology to develop a predictive maintenance schedule. To develop a better strategy for equipment and facilities maintenance, an integrated PdM arrangement construction built on Building Information Modeling and the potential of the Internet of Things to facilitate maintenance monitoring was developed. This is because both BIM and IoT can increase the efficiency of PdM. The future state of the MEP components was predicted using ANN and SVM machine learning approaches.

Calabrese et al. [13] have developed a strategy for a large Italian wood processing company that uses machine learning for industrial wood processing machines. The predicted probability of failure is calculated from time series event data using tree-based categorization models. To estimate the RUL of machines used in the wood industry, the researchers employed temporal feature engineering approaches and trained a set of classification algorithms. A self-customizable sample of machines is tested without shutting down the machine to show the effectiveness of the proposed technique. Uhlmann et al. [14] proposed a method based on ML models for exploring and recognising offline data from different sources. In this case, clusters can be found in the data using the sensors. They identified the operation of conventional machine tools and three false scenarios. These results were also applied in the creation of a disease screening model that facilitated the creation of machine tools for PdM solutions.

To reduce the strength of the RNN algorithm, Markiewicz et al [15] developed a unique method in which the processing is performed by sensors. One sensor performed the data analysis, and only one packet was transmitted at that time, perhaps because the machine was operating incorrectly. Since the ultra-low power electronics used in the system can supply the sensors with the collected energy. This configuration substantially increased the processing capacity with low energy consumption. Zenisek et al [16] presented a Random Forest (RF) using deep learning technique to detect potential concept deviations in a continuous stream of data. The concept was to use sensor-equipped devices to provide data about a particular device condition to detect future failures and degradation. These advances showed the potential to reduce material and time costs by preventing failures and improving performance. However, high-quality data must first be screened to enable calculations to be made. Additionally, based on earlier synthetic dataset experiments, a novel technique was devised to determine the implications of data as a potential indication of atypical system performance.

The application of PdM for a milling cutting instrument was explained by Traini et al. [17] and verified with real data sets. Overall, this model provided a basic outline for creating a tool to analyse the wear level, an elementary engineering tool, and error prevention to increase productivity in conjunction with machine-human collaboration. Chen et al. [18] developed the Cox Proportional Hazard Deep Learning (CoxPHDL) technique to solve the problem of constrained and sanitised data in maintenance. In order to achieve a useful result, the aids of improved dependability and machine learning must be considered in the first place. An autoencoder was first used to generate a valid representation of the nominal data. The CoxPHM was then used to compute the time between failures (TBF) for the stored data. A BIM and IoT-based FMM model were presented by Cheng et al. [19]. This paradigm included application and info layers to achieve the best maintenance performance. SVM and ANN machine learning approaches were used to assess the probable condition of the mechanical, electrical and piping components.

To perform PdM on nuclear infrastructures, an ML technique was established by Abidi et al. [20]. The unusual events that can occur in a nuclear structure were compared and navigated using logistic regression (LR) and SVM. The SVM provided the most accurate evaluation results. In addition, the parameters of the LR and SVM algorithms were optimized. A novel model with a significantly lower probability density was developed to correlate data from nuclear infrastructure, although previous research used a large amount of data. Abidi et al. [20] presented a five-phase intelligent PdM planning approach that includes statistics correction, data normalization, feature collection, accurate decision making, and evaluation. Predictive maintenance increases the sustainability of production by reducing breakdowns, downtime and material waste. Both time and material waste can be reduced through the use of efficient PdM. For predictive maintenance, Mishra and Manjhi [21] proposed a failure prediction model. The proposed machine learning approach utilizes many inputs to predict the future downtime of ATMs at the module level.

The PdM model with RNN prediction was investigated by Rahhal and Dia [22]. With the help of IoT sensors, a large database was created to collect the data from devices. Moreover, the prediction made with the collected data has given encouraging results. Vanilla-RNN and LSTM-RNN, two different types of NN, have been used and these two have provided better predictions for bulb lighting. In addition, M. Paolanti et al. [23] have worked on PdM and developed a Random Forest ML technique. Using a variety of sensors and communication protocols, real-time industrial data is collected and analyzed in the Azure cloud architecture to then use the ML model. Finally, a comparison of the results was also carried out using the simulation tool analysis. Three categories of historical data were used. Zhang et al. [24] proposed a two-layer method for predicting production failures. Methods such as PCA and grouping were applied. In addition, Random Forest is used to apply the final model which performs better. To overcome the shortcomings of previous research, such as the long training times, high processor complexity, and time-data dependency when using statistical analysis methods, a new model that outperforms previous attempts must be developed. When industrial plants fail, there is a considerable loss of production and quality as well as unplanned downtime until the plant can function as intended again.

It is therefore to be expected that PdM will reduce unplanned downtimes of production facilities. PdM also helps to reduce the time and cost of unnecessary maintenance and repair work when a machine breaks down. PdM therefore guarantees the best possible use of resources. PdM can be performed through the use of mechanical analysis, lubricant analysis, electromagnetic thermography, acoustic monitoring and vibration analysis. The machine for manufacturing car parts is one of the various machines used in the manufacturing industry that will be further investigated to achieve the goals of the research. The development of PdM systems saves money and time and prevents the entire production line from breaking down. In the assessment by Li et al. [14], potential system failures are categorized into four groups: Atmospheric influences, elemental waste, human error and incorrect treatment techniques.

2.3 Industry Trends and Applications

Industry 4.0 is making a number of changes to the paradigm of industrial automation. The concept of smart manufacturing can be realised through the use of IoT and CPS technologies that introduce cognitive automation and ultimately lead to the creation of smart products and services [25]. Companies will have to deal with the problems of a much more dynamic environment resulting from this new technology. Most of these industries are not spontaneously able to deal with this evolving reality in which a multitude of events do not always lead to improved performance [26]. One of the basic ideas of Industry 5.0, which is transforming conventional production into smart, sensor-connected factories with ubiquitous technologies, is the source of vast amounts of data. A practical realisation of this idea is the use of data analytics in the development of decision support systems which can enable more efficient decision making and speed up the recovery from errors [27], [28]. The Design Support System in Industry

5.0, that evolved from the Decision Support Systems, offers strategy supports by utilizing machine learning algorithms to suggest, for instance, new product versions based on a product's attributes [29].

PdM is another application that relies heavily on data; in this case, CPS can provide a self-sustaining mind. With this method, the industry is able to predict product performance degradation and autonomously accomplish and improve product service requirements [30], [31]. Predictive maintenance has numerous advantages for production, but also disadvantages that need to be overcome. PdM has several advantages, such as higher productivity, fewer system failures [32–35], less unplanned downtimes [36–37], better management of financial and human resources [38], and better planning of maintenance interventions. In addition to diagnosing problems, ML can also be used to predict and anticipate failures. For example, ML can be used to estimate the lifetime of a machine by processing the huge amount of data to train an algorithm.

3. Methodology

The methodology used for performance analyses of the dateset is illustrated in Figure 1. It generally comprises four steps, namely data collection and preprocessing, training of the ML model, implementation of the LSTM and performance evaluation. Further, it is described below as per the methodology used for PdM:

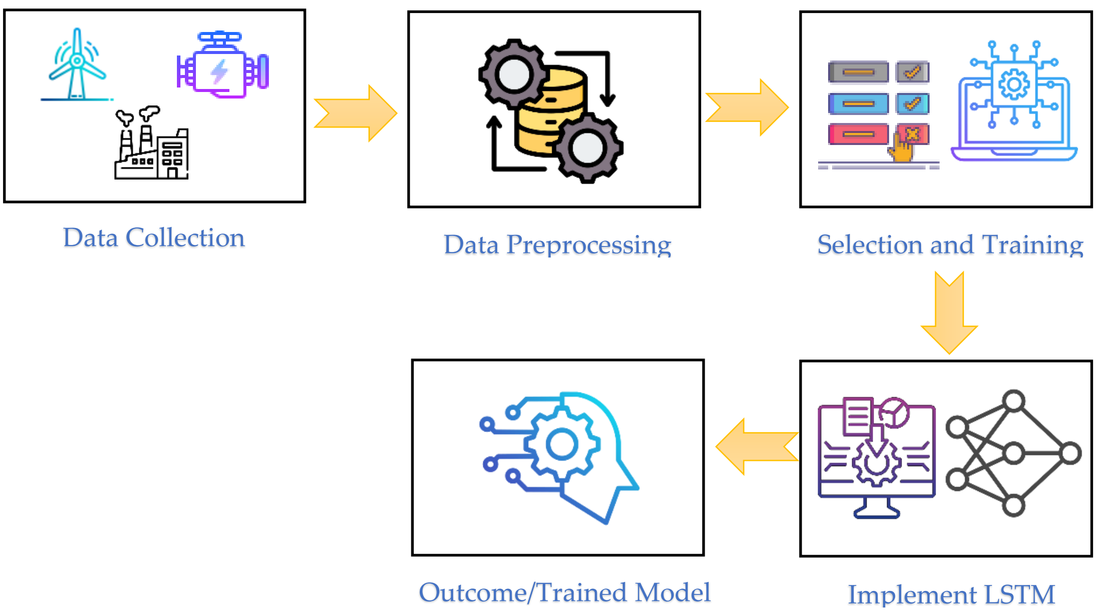


Fig. 1 Proposed Methodology

i. Data Collection and Preprocessing

Here, two datasets from Kaggle is considered to test the effectiveness of the proposed algorithm. The two datasets are 3W and PMSG wind turbine datasets. The first dataset deals with adverse events in oil wells and the second with short circuit faults in wind turbines. Before processing, first of all normalized the data between 0 and 1 and removed noise and missing values from the data points.

ii. Selection and Training of Machine Learning Models

As you have seen above, there are many ML models in the literature that can deal with this type of problem. So, the challenge is to find the perfect model for this problem from these models. Since LSTM is a popular machine learning model and it can handle time series prediction data very well, so used this model in this scenario. Another question is how many data points can be used for training, how many data points for validation and how many data points for testing. Here, in two data sets, used 80% of the information for training, 10% is used for validation and the rest of the data points for testing.

iii. Implementation of LSTM

After validating the data and selecting the training, an LSTM model for predicting downtime is implemented. In this model, the first layer is an LSTM layer with 100 units, followed by another LSTM layer with 50 units. After each LSTM layer, a dropout is performed to stop overfitting. The last layer is a deep output layer through 9 units for the 3W dataset and three units for the PMSG dataset of the wind turbine. Figure 2 shows the architectural diagram of the LSTM model and how it works. The core components of an LSTM cell include the cell state and three gates: the input gate, the forget gate and the output gate. The cell state acts as a memory that carries important information through the time steps, while the gates regulate the flow of information. The aforementioned gates use sigmoid and tanh activation functions to control the flow of information and ensure that the network can maintain relevant data over long sequences. This unique architecture allows LSTMs to overcome the vanishing gradient issues that traditional RNNs face. This makes them extremely effective for tasks such as time series prediction, natural language processing, and other applications that require modeling long-range dependencies in data. The forgetting gate decides which information should be discarded, the input gate handle the addition of new data to the cell state, and the output gate decides which part of the cell state would be used to generate the hidden state for the next time step.

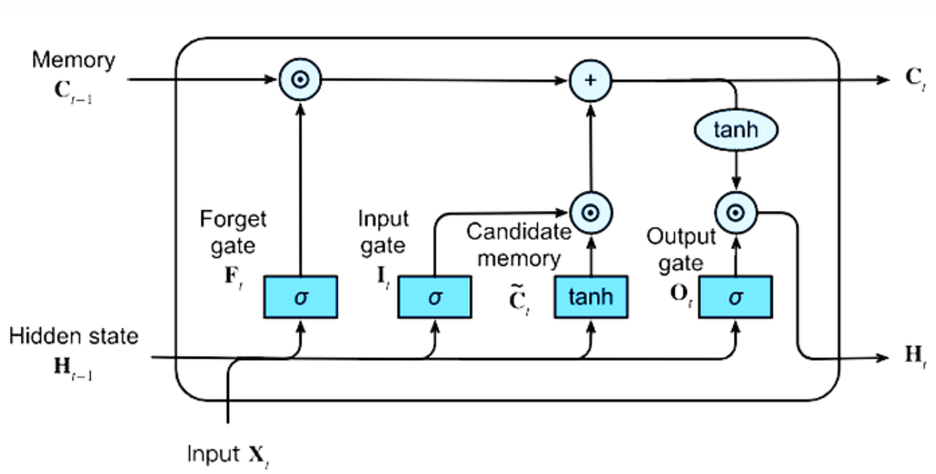


Fig. 2. LSTM Model Architecture

iii. Outcome/Trained Model

The result after implementing the proposed model provides a significantly improved prediction of machine faults, which in turn can be used for predictive maintenance of industrial equipment to increase production. It also helps to reduce the cost of maintenance, which is usually done manually by the mechanics, taking time to find the faults. The increased maintenance time affects production.

3.1 Description of the Dataset and Features Used

In this study, two data set from Kaggle are used to assess the efficiency of the proposed method. They are the 3W and wind turbine PMSG data sets [39], [40]. The first dataset refers to undesirable events in oil wells, the second to short circuit faults in wind turbines. All features are used in the 3W and wind turbine PMSG datasets. The wind turbine PMSG short circuit fault dataset and the 3W dataset contain detailed time series data required for predictive maintenance modeling in the industry. The Wind Turbine PMSG dataset was developed to detect and predict short circuit faults in wind turbine generators and includes key electrical and mechanical characteristics such as voltage, current, power generation, vibration and temperature. The 3W dataset, on the other hand, is intended for the oil and gas sector. It collects critical operating data from oil wells such as pressure, temperature, flow rate, vibration level and pump speed. Both datasets contain a binary target variable (fault indicator for the wind turbine PMSG dataset and equipment status for the 3W dataset) that indicates whether a fault has occurred, making them ideal for training machine learning models to predict equipment failures.

Table 1 Sample data of both datasets

Feature Name	Description	Data Type	Range/Values	Dataset
Timestamp	Time of data recording	Datetime	N/A	Both
Voltage	Voltage level of the wind turbine generator	Float	Varies	Wind Turbine PMSG

Current	Current flow in the generator	Float	Varies	Wind Turbine PMSG
Power	Power output of the generator	Float	Varies	Wind Turbine PMSG
Frequency	Electrical frequency output	Float	Varies	Wind Turbine PMSG
Vibration	Vibration levels detected	Float	Varies	Both
Temperature	Temperature within the system	Float	Varies	Both
Fault Indicator	Indicates fault occurrence (1 = Fault, 0 = No Fault)	Integer	0, 1	Wind Turbine PMSG
Pressure	Pressure level within the oil well	Float	Varies	3W
Flow Rate	Flow rate of oil through the well	Float	Varies	3W
Pump Speed	Speed of the oil well pump	Float	Varies	3W
Equipment Status	Operational status (1 = Normal, 0 = Fault)	Integer	0, 1	3W

These data sets are characterized by their rich and diverse continuous features. A selection of data can be found in Table 1. The ML model datasets are divided into training, validation and testing datasets, with 80% of the data used for training, 10% for validation and 10% for testing. This division guarantees that the models can learn efficiently from a variety of situations and at the same time are able to generalize to new, previously unknown data. The 50 epochs used for training represent a compromise between underfitting and overfitting, allowing the models to converge well while avoiding over-specialization on the training data. Both datasets are crucial for improving predictive maintenance tactics, as they allow models to reliably predict problems and optimize maintenance schedules. ML models use these data sets to improve operational efficiency, reduce downtime and prevent costly breakdowns in both the renewable energy and oil and gas industries.

The core objective of predictive maintenance is to forecast equipment breakdowns based on past time series data. Consider a $dataset\ (D = \{(x_t, y_t)\}_{t=1}^{T}), where (x_t \in R^n)\ is\ the feature\ vector\ at\ time\ (t), and (y_t \in \{0,1\})$ is the binary label indicating whether a fault occurred (1 for fault, 0 for normal operation). The intent is to train a predictive function $(f: R^n \to [0,1])$ that estimates the probability of a fault occurring at time (t). The LSTM network (Long Short-Term Memory) is a form of recurrent neural network (RNN), which is characterized by the recording of temporal relationships in time series and is therefore ideally suited for predictive maintenance applications. The LSTM cell is characterized by the following equations:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad \text{(Input Gate)} \tag{1}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad \text{(Forget Gate)} \tag{2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad \text{(Output Gate)} \tag{3}$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \qquad \text{(Candidate Cell State)} \qquad (4)$$

$$C_t = f_t \odot C_t - 1 + i_t \odot \tilde{c}_t \qquad \text{(Cell State Update)} \qquad (5)$$

$$h_t = o_t \odot \tanh(\tilde{c}_t) \qquad \text{(Hidden State)} \qquad (6)$$

Where:

- i_t is the input gate, $f_t$is the forget, o_t is the output gates, and $\tilde{c_t}$ is the candidate cell state.
- W_i, W_f, W_o, W_c, U_i, U_f, U_o, U_c, and b_i, b_f, b_o, b_c are the learnable parameters of the LSTM.
- σ is the sigmoid activation function, whereas $\tanh$ is the hyperbolic tangent activation function.
- $\odot$ denotes element-wise multiplication.

3.2 Loss Function

To train the LSTM model, minimized a loss function that quantifies the difference between the predicted probability $(\widehat{y_t}) = f(x_t)$ and the actual label y_t. For binary classification, the binary cross-entropy loss function is typically used:

$$L(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\left[y_t \log\left((\widehat{y_t})\right) + (1 - y_t)\log\left(1 - (\widehat{y_t})\right)\right] \qquad (7)$$

Where θ indicates the computational parameter (weights and biases of the LSTM layers).

3.3 Model Training and Optimization

The model variables (θ) are changed using optimization approaches such as gradient descent. The update rule for the parameters is given by:

$$\theta^{(k+1)} = \theta^{(k)} - \eta \nabla_\theta L(\theta) \quad (8)$$

Where:

- η is the rate of learnings.
- $\nabla_\theta L(\theta)$ is the value of the loss function relative to the parameter (θ).

3.4 Evaluation Metrics

The model's performance is assessed using metrics like mean absolute error (MAE) and coefficient of predictability R^2. It is calculated as follows:

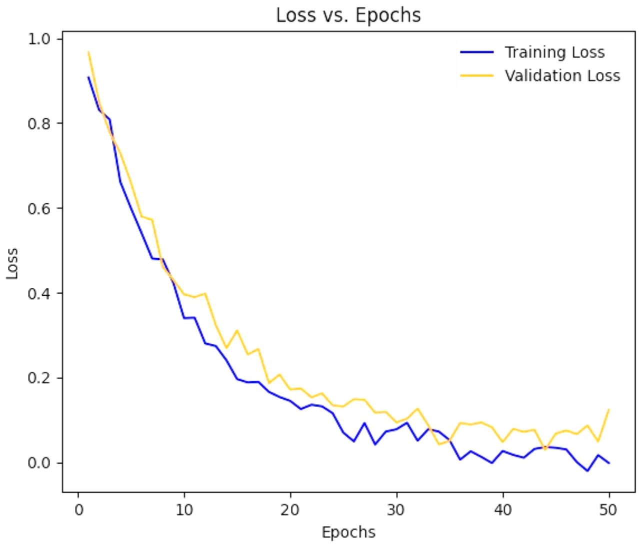
$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|y_t - \widehat{(y_t)}\right| \quad (9)$$

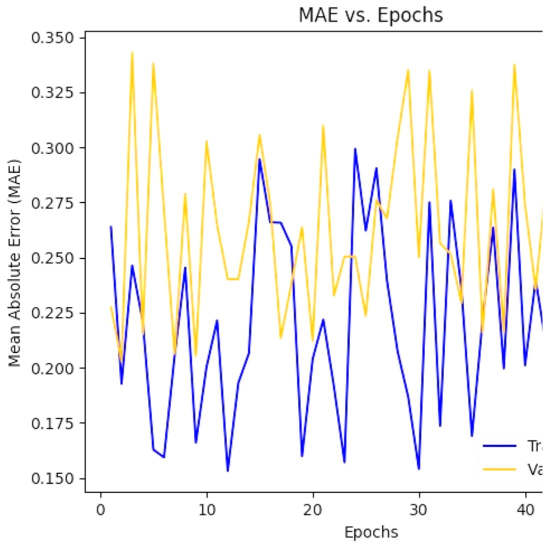
$$R^2 = 1 - \frac{\sum_{t=1}^{T}\left(y_t - \widehat{(y_t)}\right)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \quad (10)$$

Where, $\bar{y}$ is the mean of the actual labels, R^2 represents the fraction of variation in the dependent variable that can be projected from the independent variables, with values closer to one suggesting a better fit.

4. Results

The metrics used to assess model performance are MAE and R2Score for the goodness of fit of deep learning models. The models were trained for a total of 50 epochs, a strategic decision made after careful consideration of the training process. Training a model over multiple epochs offers distinct advantages that contribute to its performance and resilience. The graphs shown in Figure 3(a-c) provide important insights into the performance of the LSTM model throughout the training process. Each graph represents a key evaluation metric - loss, MAE and R2Score - as a function of the number of training epochs. These metrics supported the evaluation of model learning on unseen data.





(a) (b)

Fig. 3(a-c) Performance evaluation of LSTM model

i. Loss vs. Epochs

The first figure 3(a) shows the training and validation loss over 50 epochs. Loss functions measure how well the model's predictions match the actual values, with lower values indicating better performance.

- **Training Loss:** The training loss curve shows a steady decrease over the epochs, representing that the model is gradually learning after the training data. A consistent decrease in loss indicates that the model is fitting the training data more accurately as training progresses.
- **Validation Loss:** The validation loss initially decreases, indicating a better generalization to new data. However, if the validation loss plateaus or even slightly increases within a certain number of epochs, this may indicate the onset of overfitting, where the model tends to memorize the training data rather than learn generalizable patterns.

This pattern is common in models that use machine learning and emphasizes the significance of tracking both training and validation loss to find the best stopping point for training.

ii. MAE vs. Epochs

The second Figure 3(b) shows the MAE for both the training and validation sets across epochs. MAE is a simple statistic that evaluates the average number of errors in the predictions without considering their direction. A lower MAE value indicates higher accuracy.

- **Training MAE:** The training MAE reduces with time, showing that the model's predictions improve on the training set with learning.
- **Validation MAE:** Similar to the training MAE, the validation MAE initially decreases, indicating that the model makes better predictions with unseen data. If the validation MAE remains close to the training MAE, this indicates that the model is not overfitting and can be effectively generalized to newly acquired data.

If the MAE value for validation differs from the MAE value for training after a few epochs, this may indicate that the model is beginning to overfit the training set and lose its ability to standardize.

iii. R2Score vs. Epochs

The third Figure 3(c) presents the R2Score for both the training and validations set. The percentage of the dependent variable's variation that can be predicted from the independent variables is shown by the R2Score, sometimes referred to as the coefficient of determination. A model with a R2Score nearer 1 is thought to account for a larger percentage of the variation and, hence, has higher predicted accuracy.

- **Training R2Score:** The training R2Score increases over the epochs, suggesting that the model is capturing more of the variance in the training data as it learns.
- **Validation R2Score:** It often follows the same pattern as the R2Score for training. A steady improvement in the validation R2score means that the model makes better and better predictions from unseen data. If this score plateaus or starts to decrease, this may indicate that the model is starting to overfit.

The collective effect of these three diagrams gives a complete picture of the training process of the model. The model's ability to learn and generalize is demonstrated by a steady decrease in losses and MAE, as well as an improvement in R2Score. However, careful monitoring of the validation metrics is crucial, as deviations from the training metrics could indicate overfitting. By monitoring these metrics, one can determine the optimal number of training epochs and adjust the model to improve performance. The structure of the LSTM network in combination with the appropriate loss functions and op-timing methods enables effective prediction of device failures in time series data. The decision to use 80% of the data for training, 10% for validation and 10% for testing ensures that the model is also well suited for new data. The selection of 50 epochs for training ensures that there is a balance between achieving model convergence and avoiding overfitting, which makes the model robust for real-world predictive maintenance applications. In subsections 4.1 and 4.2, compared the LSTM model with RNN and different versions of LSTM, such as convolutional LSTM, Bi convolution LSTM. The results are broken down by dataset to analyze.

4.1 Results on 3W dataset:

Figure 4 and Table 2 show the performance of RNN using R2Score and MAE. After getting the results, it has been analyzed that RNN gives the best result at 50 units, and as the units increases the performance get down by increasing MAE.

Table 2 Results for RNN model on 3W dataset

Dynamic Parameters	R_2Score	Mean absolute error
RNN units = 10	0.8273	0.0504
RNN units = 50	0.8427	0.0478
RNN units = 100	0.7182	0.0657

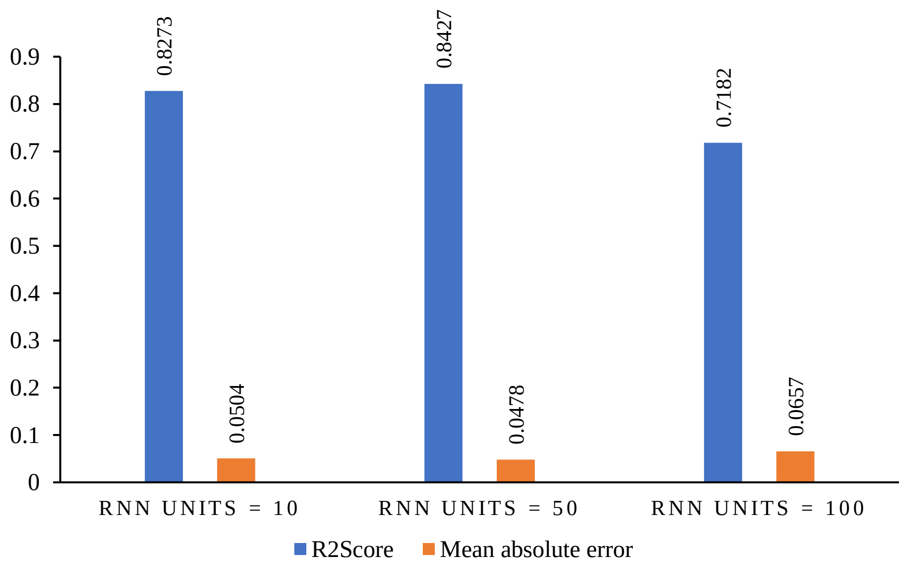


Fig. 4. Performance of RNN model

Figure 5 and Table 3 show the evaluation of the Convolutional LSTM model on the 3W dataset using R2Score and MAE. The results show that the Convolutional LSTM at 100 units yields an R2Score of 0.9362 and an MAE of 0.0213, while the R2Score at 50 units is lower at 0.9312.

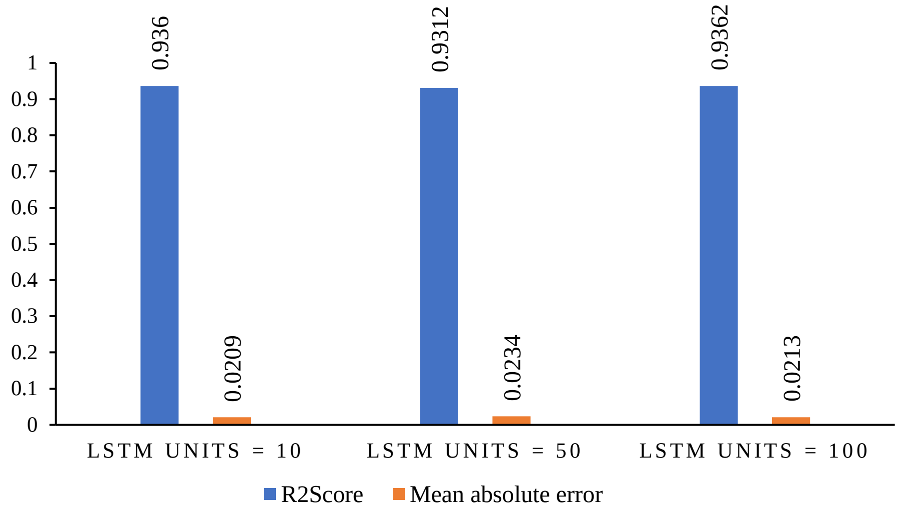


Fig. 5. Performance of Convolutional LSTM

Table 3 Results for Convolutional LSTM model on 3Wdataset

Dynamic Param-eters	R_2Score	Mean absolute error
LSTM units = 10	0.936	0.0209
LSTM units = 50	0.9312	0.0234
LSTM units = 100	0.9362	0.0213

Figure 6 and Table 4 show the performance of the Conv. Bi-LSTM model on 3Wdataset using R2Score and MAE. After getting the results, it has been analyzed that Bi-LSTM gives the 0.9847 R2Score at 100 units, and MAE is 0.0227 whereas the R2Score at 10 units is less i.e. 0.9557 and MAE is 0.0475. So, it indicates the MAE decreases as the units increases.

Table 4 Results for Conv. Bi-LSTM model on 3Wdataset

Dynamic Parameters	R_2Score	Mean absolute error
units = 10	0.9557	0.0475
units = 50	0.985	0.0203
units = 100	0.9847	0.0227

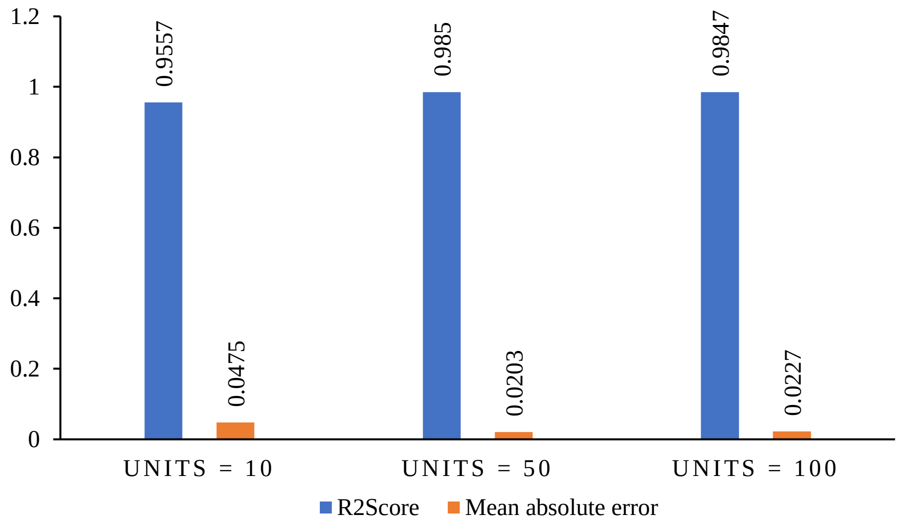


Fig. 6. Performance of Conv. Bi-LSTM

4.2 Results on Wind turbine PMSG dataset:

The results for RNN on the PMSG dataset of a wind turbine are shown graphically in Figure 7 for better understanding. Figure 7 and Table 5 show the performance of the RNN model for the wind turbine PMSG dataset using R2Score and MAE. After analyzing the results, it was found that the RNN model at 10 units gives an R2Score of 0.6173 and MAE of 0.0688, while the R2Score at 100 units is lower at 0.5587 and the MAE is lower at 0.0988, which is higher than the other units compared. This means that the MAE increases as the number of units increases.

Table 5 Results for RNN model on Wind turbine PMSG dataset

Dynamic Parameters	R_2Score	Mean absolute error
RNN units = 10	0.6173	0.0688
RNN units = 50	0.6458	0.0897
RNN units = 100	0.5587	0.0988

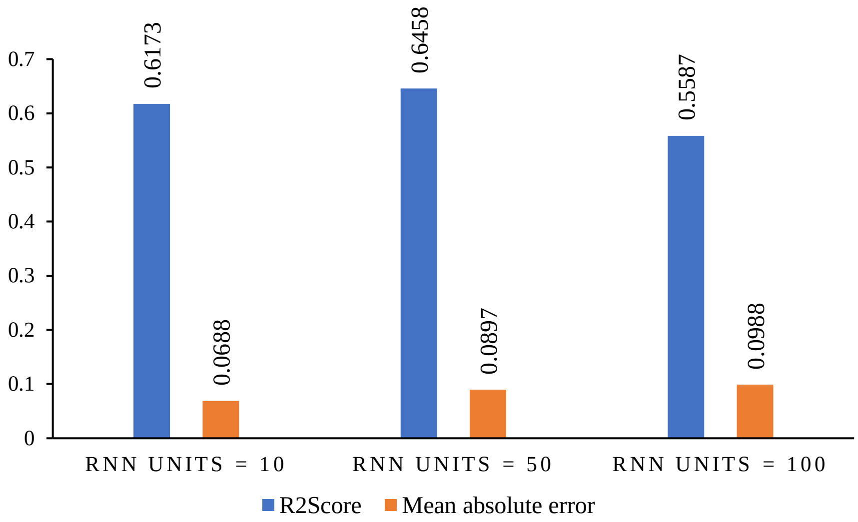


Fig. 7. Performance of RNN model

Figure 8 and Table 6 show the performance of Convolutional LSTM on the wind turbine PMSG dataset on R2Score and MAE. The analysis of the results shows that the Convolutional LSTM at 100 units gives an R2Score of 0.8051 and MAE of 0.0329, while the R2Score at 10 units is lower, i.e. 0.7961 and the MAE is 0.0305. This therefore means that the MAE value changes slightly as the number of units increases, while the R2Score increases as the number of units increases.

Table 6 Results for Conv-LSTM model on Wind turbine PMSG dataset

Dynamic Parameters	R_2Score	Mean absolute error
LSTM units = 10	0.7961	0.0305
LSTM units = 50	0.8013	0.0388
LSTM units = 100	0.8051	0.0329

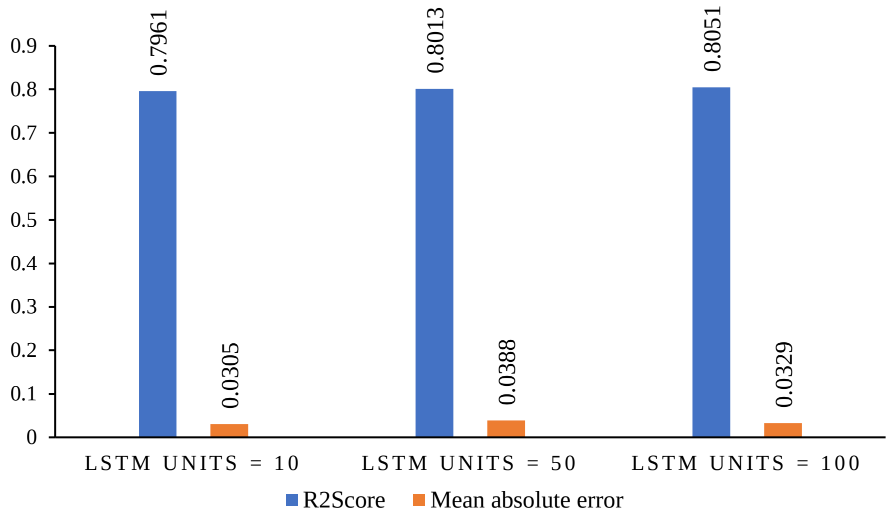


Fig. 8. Performance of Convolutional LSTM

Figure 9 and Table 7 show the performance of the Convolutional Bi-LSTM on the wind turbine PMSG dataset using R2Score and MAE. The analysis of the results shows that the Convolutional Bi-LSTM at 100 units provides an R2Score of 0.8712 and MAE of 0.0225, while the R2Score at 10 units is lower, i.e. 0.8448 and the MAE is 0.0411. This means that the MAE decreases as the number of units increases, while the R2Score increases as the number of units increases.

Table 7 Results for Conv. Bi-LSTM model on Wind turbine PMSG dataset

Dynamic Parameters	R_2Score	Mean absolute error
units = 10	0.8448	0.0411
units = 50	0.8558	0.0236
units = 100	0.8712	0.0225

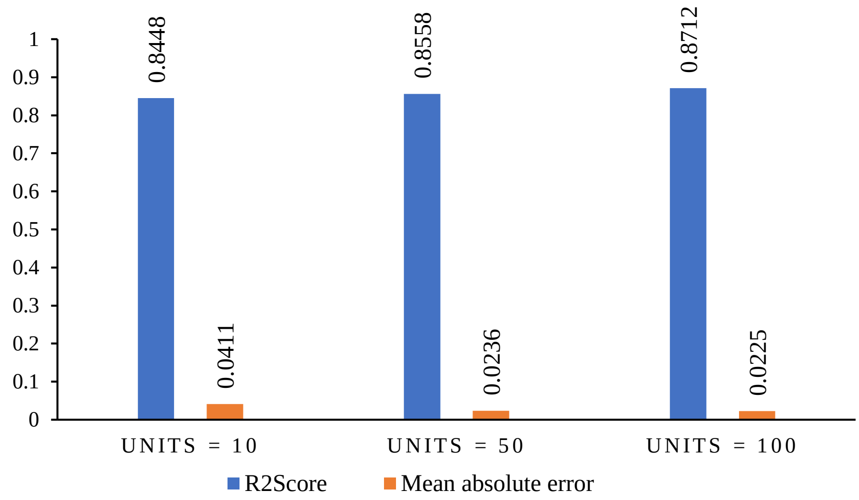


Fig. 9. Performance of Conv. Bi-LSTM

After analyzing the results, it was found that the LSTM outperforms the RNN. As with the LSTM, the R2score increases while the MAE decreases with the number of units performed. It helps in early prediction of faults for maintenance. Early prediction of faults helps to improve production in the industry while reducing maintenance time. The reduced maintenance time directly reduces the cost of repairs and increases production.

5. Discussion

This section discusses the interpretation of the results, the accuracy of the model and the implications of the research findings for the future. Figure 4 to Figure 6 illustrate the R2score and the mean absolute error of the 3W dataset. Then Figure 7 to Figure 9 show the R2Score and the mean absolute error of the wind turbine dataset. In this research, RNN and two versions of LSTM are used to compare the proposed method with LSTM. If the R2Score is high, then the performance of the approach is good and if the mean absolute error is less, then the performance of a method is good. From Table 2 to Table 9, it can be concluded that in both datasets, RNN method performs worst and the best method is Convolutional Bi-LSTM. The performance of Convolutional LSTM is worse than Convolutional Bi-LSTM and the performance of LSTM is worse than Convolutional LSTM. For RNN, Convolutional LSTM and Convolutional Bi-LSTM, a variable set of units is used to check the performance of each method. The variable units are 10, 50, and 100 respectively. In most cases using 100 units the performance is the best among all.

Implications of the Research Findings: In the study, the proposed methodology shows the effectiveness of the LSTM and its different versions in two data sets. In LSTM, a fixed set of LSTM units is used. For RNN, Convolutional LSTM, Convolutional Bi-LSTM, the number of units varies. The results show that Convolutional Bi-LSTM performs best among all comparative methods.

Practical Applications and Future Directions: In this article, two data sets are used to test the effectiveness of the LSTM and its different versions. Of all the LSTM versions, the Convolutional Bi-LSTM performs the best. Then another version of the LSTM can be used, the Attention model, to check the performance. This type of model can be used when time series data is available or when a system produces data at certain time intervals.

6. Conclusion

This study presents a comprehensive evaluation of predictive maintenance models, focusing on the performance of RNN and LSTM variants, including Convolutional LSTM and Convolutional Bi-LSTM. The analysis based on two critical data sets - 3W and wind turbine - shows that the Convolutional Bi-LSTM consistently outperforms the other models, with the highest R2 scores and the lowest MAE across different turbine configurations. The plots show a steady decline in training and validation losses across epochs, with validation losses beginning to diverge slightly, indicating the onset of overfitting. The lower MAE and higher R2score show the increasing accuracy of the model and its ability to capture temporal relationships in the data. Despite the encouraging results, the model's performance is fundamentally limited by the quality and quantity of accessible data, which may limit its applicability to various industrial situations. Even though the Convolutional Bi-LSTM is very promising, further progress could be made by exploring advanced architectures such as attention mechanisms and the incorporation of real-time data streams to improve flexibility and robustness in dynamic contexts. Finally, this study demonstrates the effectiveness of LSTM-based models, namely Convolutional Bi-LSTM, in predictive maintenance applications, laying the foundation for further research. The presented approaches are promising for use in industry, especially in optimizing maintenance schedules and reducing unnecessary downtime. Future research should focus on overcoming the highlighted limitations and investigating the recommended additions to advance the field of predictive maintenance.

Author Contributions: Conceptualization, methodology, writing—final manuscript, funding.

Funding: No funding is received for this work.

Conflicts of Interest: The author declares no conflict of interest financial or otherwise.

Acknowledgement: Author is thankful to the Reviewers and Journal team for giving this opportunity to get this work accepted.

References

[1] Hassan, M. A., Zardari, S., Farooq, M. U., Alansari, M. M., & Nagro, S. A. (2024). Systematic Analysis of Risks in Industry 5.0 Architecture. *Applied Sciences*, *14*(4), 1466.

[2] Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, 12(19), 8211.

[3] Hao, Q., Xue, Y., Shen, W., Jones, B., & Zhu, J. (2010, May). A decision support system for integrating corrective maintenance, preventive maintenance, and condition-based maintenance. In *Construction Research Congress 2010: Innovation for Reshaping Construction Practice* (pp. 470-479).

[4] Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia cirp*, *16*, 3-8.

[5] Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia cirp*, *16*, 3-8.

[6] Sreedharan, R., & Unnikrishnan, V. A. (2017). Moving towards Industry 4.0: A systematic review. *International Journal of Pure and Applied Mathematics*, *117*(20), 929–936.

[7] Fox, H., Pillai, A. C., Friedrich, D., Collu, M., Dawood, T., & Johanning, L. (2022). A review of predictive and prescriptive offshore wind farm operation and maintenance. *Energies*, *15*(2), 504.

[8] Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, *50*, 57-70.

[9] Amruthnath, N., & Gupta, T. (2018, April). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In *2018 5th international conference on industrial engineering and applications (ICIEA)* (pp. 355-361). IEEE.

[10] Ansari, F., Glawar, R., & Sihn, W. (2020). Prescriptive maintenance of CPPS by integrating multimodal data with dynamic Bayesian networks. In *Machine Learning for Cyber Physical Systems: Selected papers from the International Conference ML4CPS 2017* (pp. 1-8). Springer Berlin Heidelberg.

[11] Sarazin, A., Truptil, S., Montarnal, A., & Lamothe, J. (2019). Toward information system architecture to support predictive maintenance approach. In *Enterprise Interoperability VIII: Smart Services and Business Impact of Enterprise Interoperability* (pp. 297-306). Springer International Publishing.

[12] Cheng, J. C., Chen, W., Chen, K., & Wang, Q. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation in Construction*, *112*, 103087.

[13] 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.

[14] Uhlmann, E., Pontes, R. P., Geisert, C., & Hohwieler, E. (2018). Cluster identification of sensor data for predictive maintenance in a Selective Laser Melting machine tool. *Procedia manufacturing*, *24*, 60-65.

[15] Markiewicz, M., Wielgosz, M., Bocheński, M., Tabaczyński, W., Konieczny, T., & Kowalczyk, L. (2019). Predictive maintenance of induction motors using ultra-low power wireless sensors and compressed recurrent neural networks. *IEEE Access*, *7*, 178891-178902.

[16] Zenisek, J., Holzinger, F., & Affenzeller, M. (2019). Machine learning based concept drift detection for predictive maintenance. *Computers & Industrial Engineering*, *137*, 106031.

[17] Traini, E., Bruno, G., D'antonio, G., & Lombardi, F. (2019). Machine learning framework for predictive maintenance in milling. *IFAC-PapersOnLine*, *52*(13), 177-182.

[18] Chen, C., Liu, Y., Wang, S., Sun, X., Di Cairano-Gilfedder, C., Titmus, S., & Syntetos, A. A. (2020). Predictive maintenance using cox proportional hazard deep learning. *Advanced Engineering Informatics*, *44*, 101054.

[19] Gohel, H. A., Upadhyay, H., Lagos, L., Cooper, K., & Sanzetenea, A. (2020). Predictive maintenance architecture development for nuclear infrastructure using machine learning. *Nuclear Engineering and Technology*, *52*(7), 1436-1442.

[20] Abidi, M. H., Mohammed, M. K., & Alkhalefah, H. (2022). Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing. *Sustainability*, *14*(6), 3387.

[21] Mishra, K., & Manjhi, S. K. (2019). Failure prediction model for predictive maintenance. In *Proceedings of the 7th IEEE International Conference on Cloud Computing in Emerging Markets (CCEM 2018)* (pp. 72–75).

[22] Rahhal, J. S., & Abualnadi, D. (2020). IoT based predictive maintenance using LSTM RNN estimator. In *Proceedings of the 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE 2020)*.

[23] Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine learning approach for predictive maintenance in Industry 4.0. In *Proceedings of the 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA 2018)*.

[24] Zhang, D., Xu, B., & Wood, J. (2016). Predict failures in production lines: A two-stage approach with clustering and supervised learning. In *Proceedings of the 2016 IEEE International Conference on Big Data*.

[25] Kunst, R., Avila, L., Binotto, A., Pignaton, E., Bampi, S., & Rochol, J. (2019). Improving devices communication in Industry 4.0 wireless networks. *Engineering Applications of Artificial Intelligence*, *83*, 1-12.

[26] Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia cirp*, *16*, 3-8.

[27] Romeo, L., Loncarski, J., Paolanti, M., Bocchini, G., Mancini, A., & Frontoni, E. (2020). Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. *Expert Systems with Applications*, *140*, 112869.

[28] O'Donovan, P., Gallagher, C., Leahy, K., & O'Sullivan, D. T. (2019). A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning engineering applications. *Computers in industry*, *110*, 12-35.

[29] Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. *Computers in industry*, *101*, 1-12.

[30] Wang, K., & Wang, Y. (2018). How AI affects the future predictive maintenance: a primer of deep learning. In *Advanced Manufacturing and Automation VII 7* (pp. 1-9). Springer Singapore.

[31] Schmidt, B., & Wang, L. (2018). Predictive maintenance of machine tool linear axes: A case from manufacturing industry. *Procedia manufacturing*, *17*, 118-125.

[32] Liu, Z., Jin, C., Jin, W., Lee, J., Zhang, Z., Peng, C., & Xu, G. (2018, June). Industrial AI enabled prognostics for high-speed railway systems. In *2018 IEEE international conference on prognostics and health management (ICPHM)* (pp. 1-8). IEEE.

[33] Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O. P., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering–a tertiary study. *Information and software technology*, *52*(8), 792-805.

[34] Zenisek, J., Wolfartsberger, J., Sievi, C., & Affenzeller, M. (2018). Streaming synthetic time series for simulated condition monitoring. *IFAC-PapersOnLine*, *51*(11), 643-648.

[35] Bumblauskas, D. P. (2010). *Smart Maintenance Decision Support Systems (SMDSS)*. Iowa State University.

[36] Nuñez, D. L., & Borsato, M. (2018). OntoProg: An ontology-based model for implementing Prognostics Health Management in mechanical machines. *Advanced Engineering Informatics*, *38*, 746-759.

[37] Cheng, J. C., Chen, W., Chen, K., & Wang, Q. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation in Construction*, *112*, 103087.

[38] Daniyan, I., Mpofu, K., Oyesola, M., Ramatsetse, B., & Adeodu, A. (2020). Artificial intelligence for predictive maintenance in the railcar learning factories. *Procedia Manufacturing*, *45*, 13-18.

[39] Adônis, B. (2020). *Wind turbine PMSG - Short-Circuit Fault*. Kaggle.com. https://www.kaggle.com/datasets/brunoadonis/wind-turbine-pmsg-short-circuit-fault-mcsa

[40] *3W Dataset* - *Undesirable events in oil wells*. (n.d.). Kaggle.com. https://www.kaggle.com/datasets/afrniomelo/3w-dataset