Machine Learning for Predictive Maintenance: A Literature Review

Moses Laksono Singgih
Department of Industrial and Systems Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
moseslsinggih@ie.its.ac.id

Fauziyyah Firdausi Zakiyyah
Department of Industrial and Systems Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
firdazakiyyah21@gmail.com

Andrew
Department of Industrial and Systems Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
andrewonlyy@gmail.com

2024 Seventh International Conference on Vocational Education and Electrical Engineering (ICVEE) | 979-8-3315-0510-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICVEE63912.2024.10823713

***Abstract* – Purpose:** *The advancement of Industry 4.0 has driven the adoption of predictive maintenance as a strategic step in optimizing equipment performance. Today, machine learning plays a key role in predicting potential equipment failures to minimize equipment downtime. However, assessing the accuracy of the predictive maintenance model presents a challenge for researchers in creating a reliable predictive maintenance model. Therefore, this review specifically focuses on the evaluation of the outputs of various related literatures.*

Design/methodology/approach: *The approach used in making this review focuses on exploring the method of developing a predictive maintenance model with machine learning algorithms. This search is also carried out with the help of special software to search the literature database and screen the suitability of the research results, which is the main purpose of this literature review.*

Findings: *Predictive maintenance models with machine learning algorithms such as support vector machines and random forests, including their integration with time-series classification and AdaBoost, demonstrate the ability of real-time abnormality prediction and equipment failure classification, as well as improve operational efficiency and resource optimization with high stability on complex data.*

Originality value: *The results of this literature review are expected to be a guide for researchers and practitioners in exploring and validating ML algorithms to develop predictive maintenance models of equipment that can be accessed in real-time.*

***Index Terms* – AdaBoost Classifier, Machine Learning, Predictive Maintenance, Random Forest, Supervised Vector Machine, Time-series Classification**

I. INTRODUCTION

In the fast-growing world of industrial maintenance, machine learning techniques are increasingly being used to optimize asset performance and reduce downtime. This article examines the most recent research, trends, and machine learning applications in predictive maintenance. Predictive maintenance is a proactive method that seeks to discover possible problems before they cause damage to equipment, which is becoming increasingly crucial in today's economy. 4.0 [1]. Machine learning algorithms can evaluate vast volumes of data and uncover relevant trends, making them extremely useful in predictive maintenance applications [2]. This review provides an overview of the latest solutions in this field, highlighting various applications of predictive maintenance in different environments that utilize a variety of analytical methods and algorithms with the same goal, which is to identify potential failures early to enable efficient maintenance planning.

In most industry contexts, the ability to accurately predict equipment failures is critical [3]. Many manufacturing industries face the challenge of equipment breakdowns that can lead to costly downtime. Technological developments in the industrial era 4.0 have encouraged the adoption of innovations such as predictive maintenance [4]. Predictive maintenance has become a key strategy implemented by many industries to optimize equipment performance and minimize unplanned downtime [5]. In this context, machine learning is critical because of its capacity to evaluate historical and real-time data and forecast equipment faults before they occur [6].

However, the main challenge in developing predictive maintenance models is achieving high accuracy. Although several machine learning methods have been applied, such as decision trees, support vector machines, and artificial neural networks, there is always potential for development, particularly in terms of model stability and adaptation to complicated and dynamic data [7]. Along with the development of Industry 4.0, the need for more efficient and reliable predictive maintenance solutions is increasingly urgent. Therefore, it is necessary to perform an in-depth examination of the current literature to understand the trends, applications, and obstacles in the development of machine learning-based predictive maintenance models [8]. This review aims to present a comprehensive review of the research that has been conducted as well as provide guidance for researchers and practitioners in exploring and validating machine learning algorithms for the development of models that can be accessed in real-time.

II. RESEARCH STRATEGY

The research strategy for related literature begins with identifying the relevance of the retrieved information to the search keyword. The literature research was conducted using research platform, such as ScienceDirect, MDPI, and IEEE. The time frame set for the selection of main journal articles only covers literature published between 2019 and 2024. To thoroughly complete the review of main literatures, additional supporting journals research are carried out to support the validity of this literature review. The process of determining the topic begins with the definition of machine learning and predictive maintenance. Conducting journal selection using keyword machine learning and predictive maintenance, after shorting journals based on several provisions, a journal review and documenting the results of the review are carried out. The results of the documenting are analyzed to determine key findings and issues in the journal. Finally, conclusions were made based on the analysis of journal reviews.

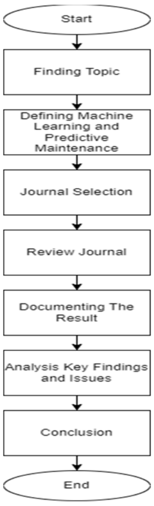

Figure 1. Research Selection Flow Diagram

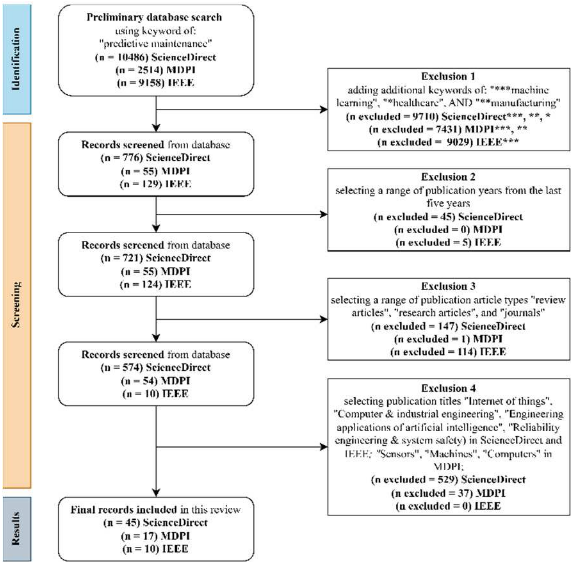

Figure 2 PRISMA flow diagram

The PRISMA flow diagram (Figure 2) illustrates the process of literature selection on the topic of “predictive maintenance” from the databases ScienceDirect, MDPI, and IEEE. The process begins with an initial search, yielding 22,158 articles. This was followed by a filtering step using additional keywords such as “machine learning”, “healthcare”, and “manufacturing”, narrowing the results down to 960 articles. Subsequently, a screening was conducted based on a publication date range within the last five years, reducing the number to 900 articles. The next step involved selecting specific types of articles, including “review articles”, “research articles”, and “journals”, resulting 638 articles. The final exclusion step focused on eliminating specific publication titles such as “Internet of Things” and “Computer & Artificial Engineering”, ultimately leaving 72 relevant articles for this review, comprising 45 from ScienceDirect, 17 from MDPI, and 10 from IEEE.

Subsequently, the selected literatures were further eliminated based on similarities in methodology and machine learning algorithm performance results. This was followed by screening several articles based on the completeness of related information, data processing outcomes, uniqueness, and novelty of the methods and cases. The result yielded four relevant articles, supported by 68 supplementary articles to strengthen the review of the four main articles.

III. Generation of Key Findings

A. *Time Series Classification: Integration of Histogram-based Information Gain Binning (HIGB) and Coefficient improved Bag of Pattern (CoBoP)*

Medical imaging equipment such as Computed Tomography (CT) provides significant benefits in helping patients understand the diagnosis of their illnesses and aiding the treatment process [9]. However, abnormalities in such equipment often occur during its operational life, such as unstable currents [10]. The presence of abnormalities can lead to a decline in the reliability of equipment functions in providing accurate medical imaging information for disease diagnosis. In addition, equipment failure can disrupt hospital operations [11]. Ironically, the failure of medical equipment failure contributes to an increase in the death rate every year [12].

The problems outlined in this case involve the presence of abnormalities that trigger the failure of medical equipment, such as CT scan, leading to inaccurate information in medical imaging information and disrupting hospital operations. This issue is predominantly due to the implementation of preventive maintenance strategies in most hospitals in Canada for medical equipment maintenance, based on the recommendations of medical equipment manufacturers [13]. Previous research on medical equipment maintenance strategies has highlighted the limitations of preventive maintenance in handling frequent and random frequencies [14]. In addition, the high cost incurred for maintenance activities, which can be triggered by the high cost of replacing the entire defective component or the high frequency of maintenance [15]. In 2019, Sabah et al. found that approximately 37% of hospitals had implemented predictive maintenance strategies for their medical equipment, a 5% rise from the year before [16]. This indicates the growing adoption of predictive maintenance strategies in medical equipment. Therefore, the development of a predictive maintenance model is necessary for planning the maintenance strategy for medical imaging equipment [15] .

A new data-based model, SAX-HCBOP, has been developed to improve prediction analysis of medical equipment failures. SAX-HCBOP is a Time Series Classification (TSC) model that employs the Histogram-based Information Gain Binning (HIGB) and Coefficient improved Bag of Pattern (CoBoP) methods [15]. Developed using Python and Spyder, the model’s processing follows three main stages: data collection, predictive model development, and model evaluation, as follows:

1. Data Collection
   Data collection involved one year of time-series data from CT equipment, including real-time monitoring of oil temperature and tube arc data, which was used as the classification label [15]. Data collection was conducted out for one year obtained from Sichuan University Hospital, China.
2. Data Pre-processing
   Data was pre-processed to clean inaccuracies, handle missing values, and correct abnormalities encountered during data collection [15]. Piecewise Aggregate

Approximation (PAA) was applied to compress data and align sampling frequencies into fixed time intervals, while linear interpolation addressed any missing values that arose due to changes in sampling frequency after compression.

3. Data-based Predictive Model Development
   The process begins with the extraction of the raw data obtained from the sensors based on the extraction features, including derivative – changes in the signal rate, offset – the difference between the current value and the moving average value based on historical data, and safety distance – the deviation between the current value and the given safety threshold [15]. The HIGB and CoBOP methods were then applied to transform time series data into a bag-of-words paradigm or a method of converting time series data into short word representations, where the frequency of these words' occurrences is then counted. Efficient algorithms like LightGBM handled the computational demands of large-scale, real-time predictive maintenance, ensuring scalability and performance without degradation in real-world applications.
4. Predictive Model Evaluation
   The predictive model was evaluated by comparing the SAX-HCBOP model with existing predictive maintenance approaches, such as TDF, TDE [17], DrCIF [18], and DTW-KNN [19] focusing on accuracy, recall, precision, and F1-score metrics [15].

There are several factors that affect the performance of machine learning algorithms, including:

Data Quality: Low-quality data, such as noisy or incomplete datasets, reduces model accuracy and increases overfitting risks. Preprocessing techniques like PAA improve data uniformity and mitigate missing values but increase computational demands, especially in real-time applications [20].

Real-time Processing: ML algorithms in real-time systems must process data continuously and deliver fast predictions. Models like SAX-HCBOP and lightweight algorithms like LightGBM enhance computational efficiency and accuracy, balancing complexity and performance [15].

Complexity of Maintenance Systems: Dynamic environments require frequent retraining and updates of ML systems. Deep learning models capture complex patterns but are costly and less interpretable. Classical models like SVM and DT are easier to maintain but struggle with complex data [21]. Information Gain Binning (IGB) offers a balance between computational efficiency and model flexibility [22].

Scalability is crucial for ML algorithms in real-world applications with large datasets and complex features. Algorithms like LightGBM, using Gradient Boosting and feature bundling, handle large-scale data efficiently [15]. Time-series models like SAX and SFA use symbolic transformations to reduce computational load, enhancing scalability. However, real-world scenarios, especially in IoMT, often involve imbalanced datasets and frequent model updates, challenging scalability and performance [23].

The SAX-HCBOP predictive model improves maintenance by accurately predicting CT equipment abnormalities, achieving 90.4% accuracy, 74.7% recall, 41.7% precision, and 53.5% F1-score, outperforming other methods like TDF, TDE, DrCIF, and DTW-KNN. Its discrete data representation handles rapidly changing conditions effectively, providing reliable predictions even with incomplete data or frequent system changes [15]. This enhances predictive maintenance, optimizing medical resource use, reducing equipment malfunction costs, and ensuring continuous high-quality medical imaging.

B. AdaBoost Classifier for Predictive Maintenance

The reliability of machine operation is very important to maintain the quality of the products produced. The traditional reactive maintenance approach, which repairs equipment after a breakdown, is beginning to be abandoned [24]. The use of machine learning and data analysis in predictive maintenance is considered to be an alternative maintenance system that is more proactive and can reduce maintenance costs [5]. On circular knitting machines, predictive maintenance is implemented by building an Internet of Things (IoT) system by recording machine speed data and operational stops on the machine [25]. The data that has been collected is trained with a machine learning model to predict when the machine will stop or break down [26]. The suggested approach is projected to provide various advantages over standard maintenance techniques, including less engine downtime, greater machine availability, and enhanced output [24]. The system's ability to identify the causes of machine failures can result in maintenance interventions being more focused and making it easier for the system to carry out appropriate actions on machines, thereby reducing unplanned downtime [27].

The problem discussed in this journal is a reactive approach that is still often used in the maintenance of textile industry equipment, where repairs are carried out only after the damage has occurred. This method results in significant machine downtime and high maintenance costs. This publication highlights the need to transition from reactive to proactive and cost-effective predictive maintenance. Predictive maintenance employs data analysis and machine learning to forecast breakdowns before they occur, decreasing downtime and costs associated with equipment failures [26]. In recent years, research on the development of predictive maintenance systems for circular knitting machines has been increasing [26]. According to Haarman et al, the impact of maintenance reaches 15 to 60% of the total operational costs of all manufacturers [28]. However, the company does not measure the exact amount of expenses associated with maintenance. Therefore, further studies are needed on the use of modern technology to be able to change these problems [1]. In this review, the researcher focuses on predictive maintenance and wants to show the significant benefits that will be obtained in the implementation of industry 4.0.

The repair method involves using a predictive maintenance system based on machine learning and the Internet of Things [29]. Speed and pause data from circular knitting machines are collected in real-time using IoT and stored in a structured database that records machine type, thread configuration, and shift activity. A new dataset, obtained from factories operating under normal conditions, serves as a repository for historical machine records and facilitates the adoption of predictive maintenance procedures by identifying machine shutdowns [26]. The classification model used is the AdaBoost Classifier, an ensemble method that combines several weak classifiers into a strong classifier. The Boosting algorithm starts by training weak classifiers with initial weights, then determines the classifier's weight based on its error rate, updating the sample weight for the next classifier accordingly until the set number of iterations is reached [30]. Customized machine learning models are used to classify machine shutdowns. To achieve high accuracy in machine failure classification, a cross-validation method with grid search is applied to optimize the AdaBoost classifier [31]. The training data comprises 60%, while the test and validation data each comprise 20% [26]. Irrelevant features are removed using the permutation importance technique during feature selection. Machine learning algorithms are then applied and optimized using grid search cross-validation to achieve the best classification results. [32]. Several factors affect the performance of machine learning algorithms:

Data Quality: Clean, relevant data is crucial for accurate predictions. Poor data quality can lead to incorrect results, while a large volume of data improves learning but poses processing and storage challenges [33].

Real-Time Processing: High processing speed and low latency are essential for real-time applications. While methods like random forest and gradient boosting are accurate, they are computationally intensive. Simpler methods like logistic regression or K-nearest neighbors may be more suitable for real-time needs [34]. Data streaming and batch processing can help manage workloads.

Complexity of Maintenance Systems: Maintaining ML models involves regular updates, bug fixes, and performance monitoring [35]. Complex algorithms like deep learning require more resources and time for training and maintenance [36].

Different ML methods have varying computing needs. Different ML methods have varying computational needs. SVMs and neural networks are more intensive than decision trees or naive bayes [37]. Scalability is crucial, with parallelizable algorithms like random forest or gradient boosting being more scalable [38]. The usage of technologies like Hadoop, Spark, and Kubernetes can aid with large-scale data processing and modeling [39]. Data quality, real-time processing, and system complexity impact ML performance. Proper management of these factors is essential for optimal results.

The application of predictive maintenance methods based on machine learning and IoT shows significant advantages over traditional maintenance methods [26]. For instance, the AdaBoost Classifier model used for engine shutdown classification achieved a high classification accuracy of 92% based on 1260 samples [26]. The average value, considering the number of samples in each category, has a precision value of 92%, recall of 92%, and F1 score of 91% [26]. By implementing the AdaBoost classifier algorithm, the system can predict when the machine will stop or break down, thereby reducing heavy equipment downtime and increasing production efficiency.

C. Supervised Machine Learning Algorithms in Predictive Maintenance

One of the most important aspects of healthcare services is the medical equipment, even any minor damage to it can jeopardize the patients' lives [40]. Moreover, during the COVID-19 pandemic outbreak, when the use of medical equipment – particularly respiratory equipment was peaked [41]. Therefore, keeping an eye on the medical equipment usage and maintenance is essential to provide best healthcare services to patients. The World Health Organization (WHO) reported around 83% of the reasons of medical equipment failure can be prevented and are brought on by inadequate maintenance strategies [42]. This is triggered by the lack of use diagnosis officers and medical device repair technicians [40]. In addition, environmental factors can contribute to deterioration of medical equipment, necessitating special treatment to ensure longevity and reliability.

An inappropriate of maintenance strategy can disrupt hospital operations and patient care. Furthermore, miscalculations in the timing of medical equipment failures and maintenance strategies often threaten the lives of patients who need medical treatment [40]. Unlike other industries, medical equipment maintenance must be optimized to meet unpredictable patient needs. Previous research on medical equipment maintenance has involved condition-based maintenance (predictive maintenance), where maintenance activities are conducted upon detecting signs of failure [43]. Shamayleh et al. (2020) demonstrated that the application of predictive maintenance in medical equipment can result in more efficient and cost-effective maintenance scheduling [44]. However, prior research has only addressed a small portion of medical equipment [40]. Therefore, further investigation is required to develop machine learning algorithms that able to predict the maintenance schedules for a wider range of medical equipment.

A predictive failure analysis model for maintaining medical equipment was developed using MATLAB [40]. Three basic processes were engaged in the data processing: data preparation, data classification, and predictive model development:

1. Data Preparation
   A dataset comprising 13,350 units of medical equipment from various public health clinics in Malaysia was categorized into 19 different types of medical equipment [40]. There were thirteen features and criteria for medical equipment failure were identified. Implicitly, this stage also involved the data reductions and adjustments of incomplete and irrelevant data to align with the 13 selected features and criteria, such as asset ID, asset location, and maintenance personnel [40]. This stage ends with the transformation of data into numerical form involving seven features, such as equipment category, function, PM status, maintenance complexity, maintenance scope, problem category, and operation [40]. Additionally, data transformation also involved data normalization with machine learning with z-score approaches with 0 as a mean and 1 as a standard deviation for consistent data [45].
2. Data Classification
   The data classification stage involves categorizing 13 features and criteria for medical device failure into three aspects that will be predicted in this model, consisting of first failure event (FFE), failure-to-year ratio (FYR), and failure rectification group (FRG) [40].
3. Predictive Model Development

A 70:30 split was applied to the dataset in order to develop the predictive model, with 30% of the sample referring to the test dataset and 70% to the dataset [46]. This split technique is commonly used in ML applications to avoid overfitting and generate high model accuracy. Furthermore, the data will later be tested to develop a predictive model for medical device failure analysis using seven supervised machine learning classification algorithms, such as Decision Tree [47], Discriminant Analysis [48], Naïve Bayes [47], Support Vector Machine [47], K-Nearest Neighbour [47], Random Forest [47], and Artificial Neural Network [47]. The classification results were evaluated using metrics like Accuracy, Precision, Recalls, Specificity, F-measure, Misclassification, and Area Under Curve. Hyperparameter testing was then carried out to optimize the model's performance, determining the predictive model with the best accuracy to predict FFE, FYR, and FRG.

Scalability is a critical factor in real-world ML applications. While SVMs and ANNs demonstrate strong performance, it needs distributed computing or cloud infrastructure to handle large-scale tasks [49]. These models require optimization to meet real-time demands. In contrast, simpler model like DTs is more resource-efficient and scale effectively with limited computational resources. However, it may struggle with complex datasets unless enhanced with ensemble methods, such as RFs [50].

One of the major challenges in handling noisy and incomplete data is its potential to degrade prediction accuracy. Applying z-score normalization helps address large-scale differences within the dataset, contributing to noise and outlier management [40]. Moreover, SVMs and DTs have demonstrated reliability in rapidly changing environments, such as medical facilities, due to their ability to efficiently process large volumes of data and provide fast and accurate predictions.

This research results for improving maintenance strategies in the three aspects of medical device failure prediction were determined based on the best-performing classification parameters for each predictive model. The SVM algorithm was the best algorithm to predict FFE with 96.9% accuracy, the DT algorithm was the best algorithm to predict FYR with 83.9% accuracy, and the ANN algorithm was the best algorithm for FRG with 76.7% accuracy [40].

D. Random Forest Classifier

Industrial production is driven by changing market demand and increasingly fierce global competition. To meet these challenges, rapid advances in today's manufacturing technology are needed. Automation is becoming a major trend among emerging technologies. Predictive maintenance is a way of continuously monitoring the machine to detect the condition of the machine before it breaks down [51]. The device is continually monitored by capturing a variety of data, including temperature and vibration. Any variation from the regular pattern may indicate a fault in the device [52]. By making these predictions, downtime on maintenance can be minimized. Asset tracking is another revolutionary method to improve efficiency in the industrial sector [53]. This approach leverages modern technologies such as machine learning, data visualization, cloud computing, and the Internet of Things (IoT) to implement these systems in today's industry, with the aim of improving efficiency and reducing operational costs.

One of the main challenges discussed in this journal is how to ensure effective predictive maintenance, which is continuous monitoring of the condition of industrial machinery to detect anomalies before they develop into serious damage. This is especially important because unplanned machine downtime or downtime can lead to decreased productivity and increased operating costs. The difficulty of finding the location of the necessary assets is another challenge in asset tracking in the industrial environment [54]. The uncertainty of the location of these assets not only leads to a waste of production time but also has the potential to lead to unnecessary order repetition, ultimately resulting in wasted resources and costs [52].

The method carried out to improve maintenance strategies using cloud-based asset tracking and predictive maintenance has several stages, including:

1. Data Collection
   This stage involves capturing data from various sensors installed on industrial assets and machinery. These sensors include RFID tags for asset tracking and temperature and vibration sensors for machine condition monitoring. The initial data used in Random Forest was collected from various sensors inside industrial buildings. These sensors measure parameters such as temperature, relative humidity, and ambient light intensity. This data is an important input to build a classification model.
2. Data Delivery and Storage in the Cloud
   Once the data is collected by RFID sensors and readers, it is transmitted to a cloud system using the MQTT protocol. The use of the cloud allows data storage in databases such as Firebase, which provides quick and secure access for further analysis. Storage in the cloud also ensures that data remains available and accessible at any time for further monitoring and processing [55].
3. Data Analysis and Processing
   The data that has been processed is then divided into two sets of training data (training data) and testing data (testing data). Training data is used to build the Random Forest model, while test data is used to evaluate the performance of the model that has been built. The Random Forest algorithm builds several decision trees (in this case, 20 trees) from the training data [52]. Each decision tree is built from a different subset of data by randomly selecting features. These trees work independently, and their results are combined through majority voting for the final prediction.

In the journal, it is explained that data quality, real-time processing, and the complexity of maintenance systems greatly affect the performance of machine learning algorithms [52]. High data quality is essential to ensure the accuracy and reliability of machine learning models [56]. Clean, complete, and relevant data allows the model to learn correct patterns and make accurate predictions, while poor or incomplete data can lead to model overfitting or underfitting, which reduces prediction performance [57]. Therefore, data preprocessing such as normalization, outlier deletion, and handling of lost data is an important step to ensure high data quality.

Real-time processing allows the system to analyse data in real time and provide a quick response to changing conditions. The algorithms used must be efficient and optimized for fast processing, such as the use of the MQTT protocol for lightweight and fast data communication [58]. Cloud computing is often used to handle the processing of large amounts of data in real-time, allowing for greater scalability and flexibility [52]. The computational requirements for algorithms like Random Forest Classifier are significant because they build many decision trees, but they are highly scalable and can be implemented in a cloud computing environment to handle large amounts of data.

The results of the improvements obtained in this journal are in the form of prediction model accuracy, where the system implemented using the Random Forest Classifier algorithm with 20 Decision trees managed to achieve an accuracy of 98% in assigning status labels to new data that has never been seen before [52]. This journal states that by adding the RUL estimation feature to the existing monitoring system, maintenance can be more proactive and responsive to potential problems, so that it can improve efficiency and smooth operations without having to experience major disruptions such as broken engines.

Findings:	*The development of a predictive maintenance model using time-series classification with HIGB and CoBoP; AdaBoost classification; Supervised ML Algorithms (SVM, DT, ANN, and RF) can predict abnormalities, classify failures, and proactively optimize the equipment maintenance with high efficiency and consistency.*
Issues:	*The development of the predictive maintenance model faces challenges related to poor data quality, lack of adaptability, resource-intensive training, overfitting risks, and insufficient integration with CMMS for real-time data and validation.*

IV. Conclusions

Research related to the implementation of machine learning algorithms in the development of predictive maintenance models for equipment has expanded to various industrial sectors, including healthcare, textiles, and manufacturing. The literature review focused on identifying and exploring the application of machine learning algorithms in the development of predictive maintenance models for equipment. The results of literature identification and exploration indicate a diversity of machine learning classification algorithms for dataset analysis in predictive maintenance model development, including Decision Tree. Discriminant Analysis, Naïve Bayes, Support Vector Machine, K-Nearest Neighbour, Random Forest, Artificial Neural Network, AdaBoost Classifier. This review also includes a comparison of the SAX-HCBOP time-series-based predictive maintenance model with existing predictive maintenance models, such as TDF, TDE, DrCIF, and DTW-KNN. These algorithms resulted in predictive maintenance models with accuracy performance exceeding 90%. These results align with the contribution of the application of machine learning algorithms in developing accurate predictive maintenance models and contributing to the realization of equipment maintenance strategies and scheduling independent of medical equipment manufacturers' recommendations; enhancing system reliability, performance, and availability; and reducing maintenance system downtime.

The literature review on the development of predictive maintenance models with machine learning algorithms is summarized into two main points for each of the selected main literature, including findings and issues. The findings highlight the discoveries or advantages of the output produced by ML algorithms that contribute to the development of optimal predictive maintenance models of equipment. The issues identified in the literature review focus on the problems faced based on the results of the data processing output in the relevant literature. Some of the issues include poor data quality, lack of adaptability, resource-intensive training, overfitting risks, and insufficient integration with CMMS for real-time data and validation. With this literature review, it is expected that the results related to the development of equipment predictive maintenance models will provide guidance for researchers or practitioners in exploring and validating relevant ML algorithms to realize the development of real-time accessible predictive maintenance models.

References

[1] T. Zonta, C. A. da Costa, R. d. R. Righi, M. J. de LIma, E. S. da Trindade and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Scient Direct,* 2019.

[2] T. N. Tran, T. H. Truong, T. V. Tran and N. H. Hai, "An overview of the application of machine learning in predictive maintenance," *Research Gate,* 2021.

[3] P. Nunes, P. Santos and E. Rocha, "Challenges in predictive maintenance – A review," *Science Direct,* vol. 40, pp. 53-67, 2023.

[4] S. Jayashree, M. N. Reza, C. A. N. Malarvizhi, A. Gunasekaran and M. A. Rauf, "Testing an adoption model for Industry 4.0 and sustainability: A Malaysian scenario," *Science Direct,* vol. 31, pp. 313-330, 2022.

[5] A. L. Paul and J. Oluwaseyi, "Predictive Maintenance: Leveraging Machine Learning for Equipment Health Monitoring," *Research Gate,* 2024.

[6] F. Raza, "AI for Predictive Maintenance in Industrial Systems," *Research Gate,* vol. 2, 2023.

[7] W. wu, A.-D. Li, X.-H. He, R. Ma, H.-B. Liu and J.-K. Lv, "A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China," *Science Direct,* vol. 144, pp. 86-93, 2018.

[8] Y. Wen, M. F. Rahman, H. Xu and T.-L. B. Tseng, "Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective," *Science Direct,* vol. 187, 2022.

[9] D. C. C. M.A, J. W.Eberhard and G. A. Mohr, "Computed tomography part I: Introduction and industrial applications," *JOM,* no. 46, pp. 14-26, 1994.

[10] S. Tabakov, "X-Ray Tube Arcing: Manifestation and Detection during Quality Controol," *Medical Physics International Journal,* vol. 6, no. 1, pp. 157-161, 2018.

[11] J. Rajwade, L. Miller and D. Simon, "Partial-data interpolation method for arc handling in a computed tomography scanner," *Computerized Medical Imaging and Graphics,* vol. 36, no. 5, pp. 387-395, 2012.

[12] C. Heneghan, M. Thompsoon, M. Billingsley and D. Cohen, "Medical-device recalls in the UK and the device-regulation process: retrospective review of safety notices and alerts," *BMJ Open,* vol. 1, pp. 1-6, 2011.

[13] A. Jamshidi, S. A. Rahimi, D. Ait-kadi and A. R. Bartolome, "Medical devices Inspection and Maintenance; A Literature Review," in *Proceedings of the 2014 Industrial and Systems Engineering Research Conference* , Canada, 2014.

[14] A. Khalaf, Y. Hamam, Y. Alayli and K. Djouani, "The effect of maintenance on the survival of medical equipment," *Journal of*

Engineering, Design, and Technology, vol. 11, no. 2, pp. 142-157, 2013.

[15] H. Zhou, Q. Liu, H. Liu, Z. Chen, Z. Li, Y. Zhuo, K. Li, C. Wang and J. Huang, "Healthcare facilities management: A novel data-driven model for predictive," *Artificial Intelligence In Medicine ,* vol. 149, pp. 1-12, 2024.

[16] S. Sabah, M. Moussa and A. Shamayleh, "Predictive Maintenance Application in Healthcare," in *2022 Annual Reliability and Maintainability Symposium (RAMS)*, Tucson, 2022.

[17] M. Middlehurst, J. Large, G. Cawley and A. Bagnall, "The Temporal Dictionary Ensemble (TDE) Classifier for Time Series Classification," in *Machine Learning and Knowledge Discovery in Databases*, Belgium, 2021.

[18] M. Middlehurst, J. Large, M. Flynn, J. Lines, A. Bostrom and A. Bagnall, "HIVE-COTE 2.0: a new meta ensemble for time series classification," *Machine Learning,* vol. 110, pp. 3211-3243, 2021.

[19] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang and E. Keogh, "Generalizing DTW to the multi-dimensional case requires an adaptive approach," *Data Mining and Knowledge Discovery,* vol. 31, pp. 1-31, 2016.

[20] Y. Zhou, F. Tu, K. Sha, J. Ding and H. Chen, "A Survey on Data Quality Dimensions and Tools for Machine Learning," in *The 6th IEEE International Conference on Artificial Intelligence Testing*, Shanghai, 2024.

[21] A. Ucar, M. Karakose and N. Kırımça, "Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends," *Applied Sciences,* vol. 14, no. 2, pp. 1-40, 2024.

[22] C. Emmanouilidis, "Topical collection "applications of machine learning in maintenance engineering and management"," *Neural Computing & Applications,* vol. 35, pp. 2945-2946, 2022.

[23] A. Vijayalakshmi, E. Abishek.B, J. Rubi, J. Dhivya, K. K and A. R. A.S., "Machine Learning-Based Prediction and Analysis of Air and Noise Pollution in Urban Environments," in *2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, India, 2024.

[24] W. J. Lee, H. Wu, H. Yun, H. Kim, M. B. Jun and J. W. Sutherland, "Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data," *ScienceDirect,* vol. 80, pp. 506-511, 2019.

[25] S. N. Elkateb, A. Metwalli and A. Shendy, "An Innovative Online Monitoring System in Knitting Industry," *Research Gate,* 2023.

[26] S. Elkateb, A. Metwalli, A. Shendy and E. A. Abu-Elanien, "Machine learning and IoT – Based predictive maintenance approach for indutrial application," *Alexandria Engineering Journal,* vol. 88, pp. 298-309, 2024.

[27] C. E. Mgbemena and F. N. Okeagu, "Development of an IoT-based real-time remote monitoring device for the maintenance of injection moulding machines in plastic industries," *UNIZIK Journal of Engineering and Applied Science,* vol. 2, 2023.

[28] R. K. Mobley, An Introduction to Predictive Maintenance, USA, 2002.

[29] M. Soori, F. K. G. Jough, R. Dastres and B. Arezoo, "Internet of Things and Data Analytics for Predictive Maintenance in Industry 4.0, A Review," *Research Gate,* 2024.

[30] P. Wang, H. Jiang, T. Gao, W. Qu and Y. Dong, "A Fusion Classifier Algorithm Based on AdaBoost," *IEEE,* 2021.

[31] M. Adnan, A. A. S. Alarood, M. I. Uddin and I. u. Rehman, "Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models," *PMC PubMed Central,* vol. 8, p. 803, 2022.

[32] J. Wang and Y. Liu, "Prediction of Sensitive Consumer Behavior based on Randon Forest with Grid Search Cross-Validation," *IEEE,* 2023.

[33] A. Aldoseri, K. N. Al-Khalifa and A. M. Hamouda, "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges," *MPDI,* 2023.

[34] A. D. Goenawan and S. Hartati, "The Comparison of K-Nearest Neighbors and Random Forest Algorithm to Recognize Indonesian Sign Language in a Real-Time," *Science Journal of Informatics,* vol. 11, 2024.

[35] P. Chen, L. Wu and L. Wang, "AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications," *MPDI,* 2023.

[36] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *MPDI,* 2023.

[37] X. Daniela, C. J. Hinde and R. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *ResearchGate,* 2009.

[38] B. Mesut, A. Baskor and N. B. Aksu, "Chapter 3 - Role of artificial intelligence in quality profiling and optimization of drug products," in *A Handbook of Artificial Intelligence in Drug Delivery*, 2023, pp. 35-54.

[39] E. Shaikh, I. A. Mohiuddin, Y. Alufaisan and I. Nahvi, "Apache Spark: A Big Data Processing Engine," in *2019 2nd IEEE Middle East and North Africa COMMunications Conference (MENACOMM)*, 2019.

[40] A. H. Zamzam, K. Hasikin and A. K. A. Wahab, " Integrated failure analysis using machine learning predictive system for smart management of medical equipment maintenance," *Engineering Applications of Artificial Intelligence,* vol. 125, pp. 1-14, 2023.

[41] M. P. Odena, A. Valls, J. Grifols, R. Farre, F. J. C. Lasosa and B. K. Rubin, "COVID-19 and Respiratory Support Devices," *Paediatric Respiratory Reviews,* no. 35, pp. 61-63, 2020.

[42] WHO, "10 facts on patient safety," WHO, 26 August 2019. [Online]. Available: https://www.who.int/news-room/photo-story/photo-story-detail/10-facts-on-patient-safety.

[43] C. Corciovă, D. Andrițoi and C. Luca, "A Modern Approach for Maintenance Prioritization of Medical Equipment," *Operations Management - Emerging Trend in the Digital Era,* 2020.

[44] A. Shamayleh, M. Awad and J. Farhat, "IoT Based Predictive Maintenance Management of Medical Equipment," *Journal of Medical Systems,* vol. 44, no. 72, pp. 1-12, 2020.

[45] D. Sree and S. Bindu, "Data Analytics: Why Data Normalization," *International Journal of Engineering & Technology ,* vol. 7, pp. 209-213, 2018.

[46] Q. H. Nguyen, H.-B. Ly, L. S. Ho, N. Al-Ansari, H. V. Le, V. Q. Tran, I. Prakash and B. T. Pham, "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil," *Mathematical Problems in Engineering,* vol. 2021, no. 1, pp. 1-15, 2021.

[47] B. Behera and G. Kumaravelan, "Performance evaluation of Machine learning algorithms in Biomedical Document Classification," *International Journal of Advanced Science and Technology ,* vol. 29, no. 5, pp. 5704-5716, 2020.

[48] H. Shiferaw, W. Bewket and S. Eckert, "Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem," *Ecology and Evolution,* vol. 9, no. 5, pp. 2562-2574, 2019.

[49] S.Balamurugan, "Special issue on artificial intelligence and machine learning for real-time image processing," *Journal of Real-Time Image Processing,* vol. 18, p. 1341, 2021.

[50] M. Frank, D. Drikakis and V. Charissis, "Machine-Learning Methods for Computational Science and Engineering," *Computation,* vol. 8, no. 1, pp. 1-35, 2020.

[51] O. Surucu, S. A. Gadsden and J. Yawney, "Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances," *Scient Direct,* vol. 221, 2023.

[52] D. Daji, K. Ghule, S. Gagdani, A. Butala, P. Telele and H. Kamat, "Cloud-Based Asset Monitoring and Predictive Maintenance in an Industrial IoT System," *IEEE,* 2020.

[53] C. Hegedus, A. Franko and P. Varga, "Asset and Production Tracking through Value Chains for Industry 4.0 using the Arrowhead Framework," *Research Gate,* 2019.

[54] R. Gandhewar, A. Gaurav, K. Kokate, H. Khetan and H. Kamat, "CLOUD BASED FRAMEWORK FOR IIOT APPLICATION WITH ASSET MANAGEMENT," *Proceedings of the Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019],* 12-14 June 2019.

[55] I. K. G. Sudiartha, I. N. E. Indrayana, I. W. Suasnawa and S. A. Asri, "Data Structure Comparison Between MySql Relational Database and Firebase Database NoSql on Mobile Based Tourist Tracking Application," *Journal of Physics Conference Series,* July 2020.

[56] Y. Gong, G. Liu, Y. Xue, R. Li and L. Meng, "A survey on dataset quality in machine learning," *Elseiver,* vol. 162, 2023.

[57] M. L. OA, C. J and C. , "Chapter 4Overfitting, Model Tuning, and Evaluation of Prediction Performance," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction [Internet].*, 2022.

[58] M. S. Rahman, T. Ghosh, N. F. Aurna, M. S. Kaiser, M. Anannya and A. S. Hosen, "Machine learning and internet of things in industry 4.0: A review," *Science Direct,* vol. 28, 2023.