Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.3, February : 2025

IMPLEMENTATION OF MACHINE LEARNING ALGORITHM FOR PREDICTIVE MAINTENANCE: A SYSTEMATIC REVIEW

Dr. Anilkumar E. N., Associate Professor, Dept. of Mechanical Engineering, LBS College of Engineering, Kasaragod.
Adharsh Gangan, Akash Sasidharan A, Amal Krishna M, Sandra Koshy Vaidyan, B. Tech Students, Dept. of Mechanical Engineering, LBS College of Engineering, Kasaragod
Dr Vishnu C. Rajan, Assistant Professor, Humanities & Social Sciences, IIT Tirupati.

ABSTRACT

Machine learning algorithms are increasingly recognized for their capability to extract insights and generate predictions from complex datasets. This systematic review evaluates the implementation and effectiveness of key machine learning algorithms across various domains.
The following algorithms are reviewed for the present study? Random Forests (RF), Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-means Clustering. Logistic Regression and XG Boost (eXtreme Gradient Boosting). These algorithms are used widely in the industries such as, healthcare, finance, manufacturing and social sciences. These algorithms are analysed based on the evaluation criterions such as methodology, strengths and weaknesses and the dataset suitability. This review incorporates the existing research to deliver actionable judgements and recommendations for researchers, practitioners and decision-makers aiming to productively apply machine learning techniques in diverse contexts.

Keywords:
Machine learning, predictive maintenance, Industry 4.0.

I. Introduction

Maintenance systems in a manufacturing plant are crucial for ensuring the smooth operation of the equipment, preventing breakdowns, and optimizing the overall productivity of the plant. The maintenance systems are designed to address the evolving needs of equipment, machinery, and facilities to keep them in good working condition. According to Carvalho et al., (2019) some of the traditionally used maintenance systems in a manufacturing plant are given below:
Preventive maintenance (PvM): This involves scheduled inspections, routine servicing, and replacement of components before they fail. The goal is to prevent unexpected breakdowns and extend the lifespan of equipment.
Predictive maintenance (PdM): Using advanced technologies such as sensors and data analytics, predictive maintenance aims to predict when equipment is likely to fail. This allows for targeted maintenance activities, reducing downtime and optimizing the resources.
Corrective Maintenance (CM): This is reactive maintenance performed in response to equipment failures or breakdowns. While it's less desirable than preventive measures, it's essential for addressing unexpected issues promptly and minimizing downtime.
Condition-Based Maintenance (CBM): Similar to predictive maintenance, CBM relies on real-time monitoring of equipment conditions to determine when maintenance is needed. It involves the continuous assessment of factors like vibration, temperature, and fluid levels.
Reliability-Cantered Maintenance (RCM): RCM is a systematic approach to maintenance that prioritizes critical components and focuses resources on maintaining the reliability of those components. It involves analysing failure modes and determining the most effective maintenance strategies.
Industry 4.0, also known as the fourth industrial revolution, speaks about the integration of digital technologies into manufacturing processes, to create a smart and interconnected management system

of the industry. The Industry 4.0 concept is being widely embraced in the production, distribution, and commercialization supply chains all over the world (Sarmiento et al., 2020). Maintenance practices in manufacturing industries have been significantly impacted by the adoption of Industry 4.0 principles. One key aspect considered for the predictive maintenance concept in Industry 4.0 is the use of sensors and Internet of Things (IoT) devices to collect real-time data on equipment performance. The advancements in data analytics techniques and machine learning algorithms are widely used to predict the repair works in the plants and the replacements of machine parts that eventually help to minimize downtime (Abidi, et al., 2022). Arthur et al. (2022) have reported that Industry 4.0 is being largely accepted worldwide due to the increased competitiveness, lesser response time of the manufacturers, reduction in production costs, and more reliable production systems due to the due implementation of the Industry 4.0 concepts. The authors have also reported that the main characteristic of a machine learning-based predictive maintenance system is its ability to predict possible machine failures more reliably and efficiently.

2. Meta-analysis and Systematic Literature Review

Meta-analysis and systematic literature review (SLR) are two indispensable methodologies in evidence-based research, providing rigorous approaches to synthesize and analyse vast amounts of data across multiple studies (Borenstein et al., 2021). Meta-analysis entails statistical pooling of results from individual studies to derive cumulative conclusions, while systematic literature reviews meticulously gather, assess, and summarize existing literature on a specific topic. These methodologies offer robust frameworks for understanding the current state of knowledge, identifying trends, resolving inconsistencies, and generating evidence-based recommendations across various disciplines (Moher et al., 2009).

The present research has adopted the meta-analysis and SLR methodology to summarise the vast amount of data available to derive a meaning full conclusion for the setting the objectives for the present research. The systematic literature review has provided a comprehensive and structured approach for reviewing and summarizing the existing research literature on the topic of predictive maintenance using machine learning technique. SLR is a methodical and rigorous process that involves systematically searching, selecting, appraising, and synthesizing relevant studies to provide a comprehensive and unbiased overview of the state of knowledge on a specific research topic (Carvalho et al., 2019). The research questions used for the present literature review are the following:

1. How the predictive maintenance system is used in different industries across different parts of the world.
2. What are the major industrial sectors that have used the ML techniques for PdM?.
3. What are the generally used machine learning techniques for PdM?
4. How primary data and secondary data are used for PdM

The present research has used the web of Science data base for the identification of the research articles published on the topic Machine learning algorithm for the predictive maintenance. The research articles include for the present research includes only the peer reviewed journal articles, review articles and IEEE conference articles are considered. The major exclusion criteria used for the present research includes the exclusion of early access articles and book chapters. In order to ensure the quality of the research articles, the criterion used for the selection of a research article is based the peer reviewed articles listed from the Web of Science data base.

2.1 Industry 4.0

Industry 4.0 refers to the fourth industrial revolution, characterized by the integration of digital technologies into manufacturing processes. It proposes the use of smart technology, data analytics, the Internet of Things (IoT), artificial intelligence (AI), and other advanced technologies to create a more interconnected and intelligent industrial environment all over the world. Predictive maintenance is a key application of Industry 4.0 in the manufacturing sector. Traditional maintenance strategies often involve either reactive maintenance (fixing equipment after it fails) or preventive maintenance (performing routine maintenance regardless of the equipment's actual condition). Predictive maintenance, on the other hand, relies on data and analytics to predict when equipment is likely to fail so that maintenance can be performed just in time to prevent failure (Abidi et al., 2022).

Industry 4.0 technologies, especially IoT, data analytics, and machine learning, enable predictive maintenance by providing real-time data and insights into the condition of industrial equipment. This proactive approach to maintenance helps businesses enhance operational efficiency, reduce costs, and improve overall equipment effectiveness (Bousdeki et al., 2019). Industry 4.0 recommends the deployment of IoT sensors to machinery and equipment to collect real-time data on their performance. The data collected by Internet of Things (IoT) sensors can be processed and analyzed using advanced analytics and machine learning algorithms to identify the patterns and anomalies in the data that may indicate potential issues with the equipment.

Further, by continuously monitoring the condition of equipment, an efficient predictive maintenance system can detect early signs of wear, deterioration, or potential failures. This allows maintenance teams to intervene proactively before a breakdown occurs. The predictive maintenance algorithms utilize historical data and real-time information to predict when equipment is likely to fail. This prediction is based on patterns and correlations identified through machine learning algorithms used in the research. This will benefit the organizations through reduced downtime, cost savings, and improved equipment life span and reliability (Dalzochio et al., 2020).

2.2 Predictive maintenance

Industry 4.0 is characterized by the integration of digital technologies into manufacturing processes. It involves the use of smart technology, data analytics, the IoT, artificial intelligence, and other latest technologies to create a more interconnected and intelligent industrial environment. Predictive maintenance is a significant application of Industry 4.0 in the manufacturing sector. Traditional maintenance strategies often involve either fixing equipment after it fails or preventive maintenance (performing routine maintenance regardless of the equipment's actual condition) . Predictive maintenance, on the other hand, relies on data and analytics to predict when equipment is likely to fail so that maintenance can be performed just in time to prevent the failure (Abidi, et al., 2022).

Bach et al. (2023) have used a prototype decision support system to gather data from sensors that are placed at various locations in the machine. The researchers have used machine learning and artificial intelligence algorithms to optimize the parameters either directly or indirectly. The study has used the open-source R package for predictive maintenance. The algorithm developed can be used by small and medium-scale industries to streamline their production without much expense.

Predictive maintenance is widely adopted in various industries, including manufacturing, energy, transportation, and healthcare, among others. It allows organizations to move from a reactive and time-based maintenance approach to a more strategic and data-driven approach, ultimately improving operational efficiency and reducing maintenance costs. Predictive maintenance has significant importance in a steel manufacturing plant for diverse reasons. Steel plants typically involve complex and expensive machinery, and any unexpected breakdown can result in substantial downtime, production losses, and increased maintenance costs. Implementing predictive maintenance in a steel manufacturing plant can provide a competitive advantage.

The ability to consistently deliver high-quality products on time, with minimal disruptions, can contribute to customer satisfaction and market competition. Hence, the present study focuses on implementing predictive maintenance in the moulding section of a steel casting manufacturing unit in a South Indian state. Predictive maintenance is a key element in a steel manufacturing plant to ensure operational efficiency, reduce downtime, optimize costs, and enhance the overall reliability and safety of the production process. By utilizing advanced technologies and data analytics, steel plants can transform from reactive maintenance practices to a more strategic and proactive approach, ultimately contributing to improved performance and profitability.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.3, February : 2025

2.3 Machine learning

Machine learning techniques have a vital role in the implementation of predictive maintenance techniques by analyzing the data patterns, detecting anomalies, and predicting potential failures. There are several ways in which machine learning can be applied in predictive maintenance. Anomaly detection using machine learning models, such as classification algorithms, can be used on historical data to identify normal and abnormal patterns in sensor readings or equipment behaviour is known as a supervised learning technique (Shin et al., 2018). In the case of unsupervised learning**,** Clustering algorithms or auto encoders are used to identify patterns or outliers in data without labelled examples, making them useful for detecting anomalies in real-time (Susto et al.,2013).

For failure prediction, Classification Models can be utilized on historical data that includes instances of both normal operation and equipment failures. Classification algorithms such as logistic regression or support vector machines are useful techniques to predict the likelihood of a failure occurring within a specific timeframe. Time Series Analysis **t**echniques such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) are very effective for analyzing time-dependent data, making them suitable for predicting failures (Lasisi and Attoh-Okine, 2018).

Recent advancements in methodologies and technologies have further enhanced the efficacy and applicability of meta-analysis and systematic literature review. For instance, the utilization of machine learning algorithms for data extraction and analysis has facilitated more comprehensive and efficient reviews. Additionally, increased attention to methodological rigor, such as the adoption of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, has ensured greater transparency and reproducibility in the synthesis of evidence (Kulkarni et al., 2018).

In conclusion, meta-analysis and systematic literature review continue to be invaluable tools for researchers and practitioners in navigating the vast landscape of scientific knowledge. Recent advancements have bolstered the reliability, efficiency, and transparency of these methodologies, thereby enhancing their utility across diverse fields. By synthesizing evidence from multiple studies, meta-analysis and systematic literature review offer unparalleled insights, facilitating evidence-based decision-making, informing policy formulation, and guiding future research directions.

3. Results of SLR

Figure 1 given below shows the year wise publication of peer reviewed journals in predictive maintenance using machine learning that was extracted from the web of science database with the key word search. The results shows that the application of machine learning technique is a new concept which has got more acceptance since the year 2020. The recent increase in the application of ML in predictive maintenance may be due to the advancement of ML algorithms The review of the literature has also shown that the major reasons for the lack of acceptance of ML algorithm for predictive maintenance are mainly due to the lack of keeping historical data and the lack of professionals in applying the ML algorithms in the industries under consideration (Carvalho et al., 2019).

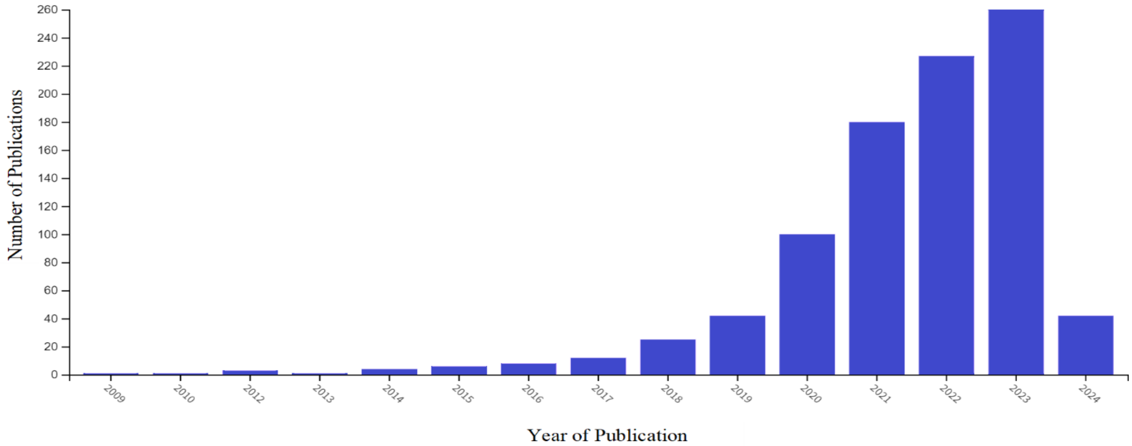

Figure 1. Year wise publication

A country-wise analysis of journals published on predictive maintenance offers a comprehensive perspective on the geographical distribution of research efforts, highlighting regions of innovation and emerging trends. Figure 2 shows the analysis result which shows that China, USA and India are major countries where more research applications in the field of predictive maintenance being adopted.

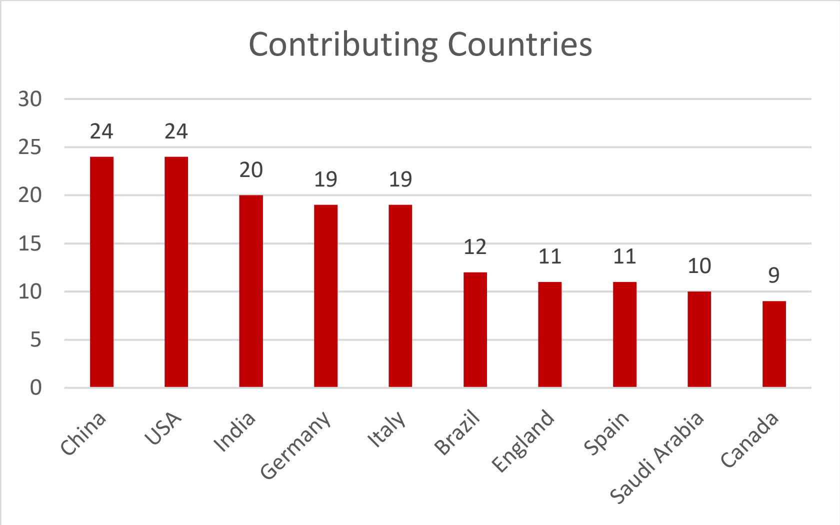

Figure 2. The country wise analysis of the journal published

The country-wise analysis of journals published on predictive maintenance reveals a dynamic and globally distributed research landscape. While certain countries exhibit prolific publication outputs and leadership in research endeavours, others are emerging as significant contributors, reflecting the increasing international interest and collaboration in advancing predictive maintenance technologies (Narayanan et al., 2019). Understanding the geographic distribution of research in this field not only highlights regional strengths and areas for improvement but also underscores the importance of international collaboration in driving innovation and addressing common challenges (Arena and Sridharan, 2019). As predictive maintenance continues to evolve, such analyses will remain crucial for fostering interdisciplinary exchange, driving technological advancements, and ultimately enhancing the efficiency and reliability of industrial operations worldwide.

Analyzing the types of journals publishing research on the topic of predictive maintenance offers valuable insights into the scholarly landscape surrounding this critical field. Investigating the distribution of research across different types of journals sheds light on the dissemination of knowledge, the rigor of peer review processes, and the target audience of scholarly works in this domain (Li et al., 2021).

The analysis of journal types publishing research on predictive maintenance underscores the diverse avenues through which scholarly knowledge is disseminated and evaluated. The analysis done in the present research in the Web of Science data base shows that the peer reviewed journals are the major source of dissemination of research on predictive maintenance (Figure 3). While specialized journals cater to in-depth technical discussions and advancements within the field, interdisciplinary publications facilitate the integration of predictive maintenance concepts across various domains. By considering the breadth and scope of journal types, researchers can gain a comprehensive understanding of the evolving landscape of predictive maintenance research, identify relevant outlets for dissemination, and contribute to the advancement of this critical field (Parida & Kallenberg, 2020).

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.3, February : 2025

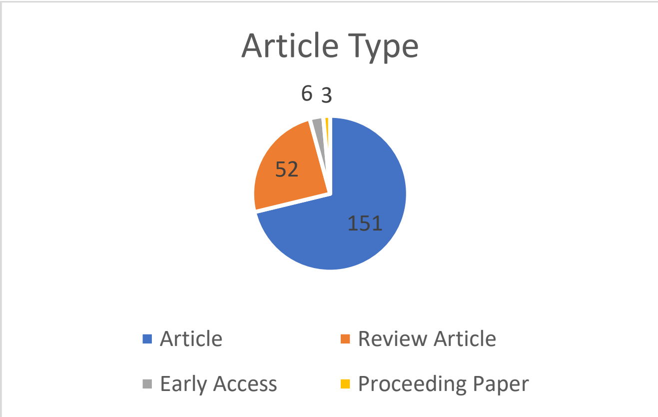

Figure 3. Analysis of the type of article

Predictive maintenance has become a crucial strategy in various industries to enhance operational efficiency, minimize downtime, and reduce maintenance costs. As a result, numerous research studies and review papers have been published to explore different aspects of predictive maintenance (Smith&Jones, 2020). The major journals that are publishing the research papers as per the WoS database is given in the Figure 4.

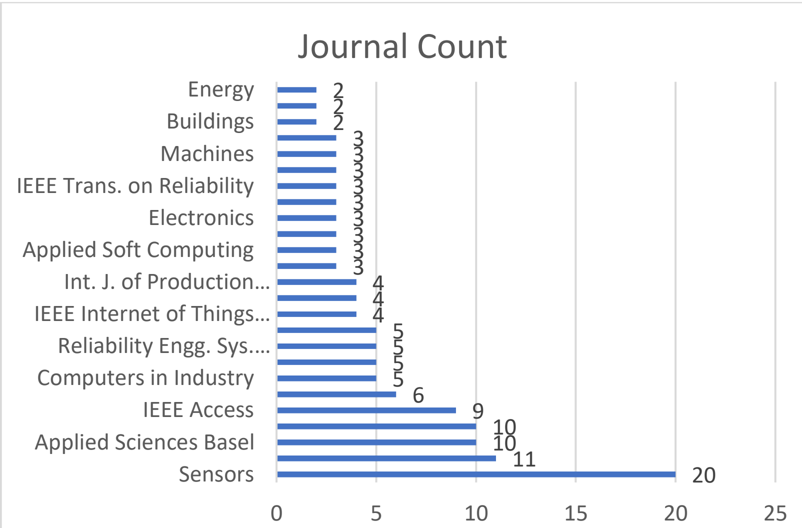

Figure4. Major types of journal published the predictive maintenance review papers

Co-occurrence analysis is a method used in literature review and textual analysis to identify patterns of co-occurrence between terms or concepts within a corpus of documents. This technique involves analyzing the frequency with which certain terms appear together in the same document, paragraph, sentence, or other defined unit of text. It helps to identify the Identifying Key Concepts, Discovering Relationships, Reducing Subjectivity and Supporting Hypothesis Generation. The co-occurrence analysis done for the present research is given in Figure 5.

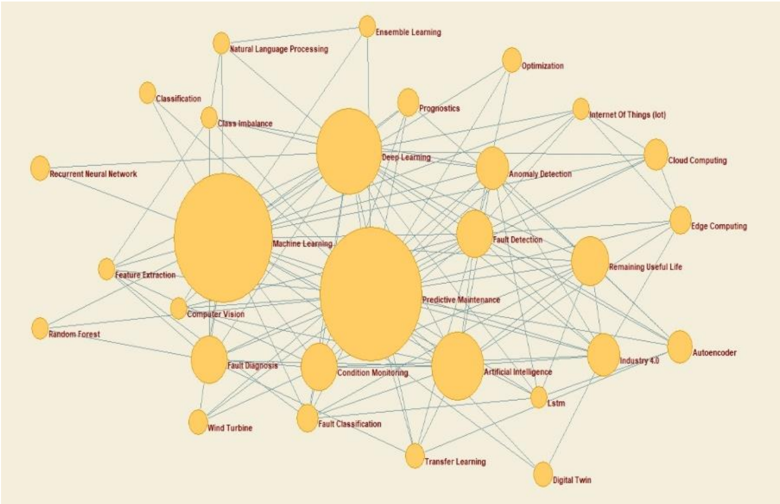

Fig 5. Co-occurrence Analysis

Co-citation analysis is a valuable method in literature review that examines the frequency with which two documents are cited together by other works. This approach enables researchers to identify relationships between different publications based on their citation patterns rather than direct content analysis. The co-citation analysis is a valuable method for literature review that enables researchers to identify influential works, map the intellectual structure of a field, identify research gaps, validate findings, and gain insights into the evolution of ideas. While it has certain limitations and challenges, co-citation analysis provides a systematic and objective approach to exploring the relationships between different documents within the literature. By incorporating co-citation analysis into their research methodology, scholars can enhance their understanding of the existing literature and contribute to the advancement of knowledge in their respective fields. The co-citation analysis done in the present study is shown in Figure 6. The Table 1 provides the details of the first ten research articles co-cited in the present research topic.

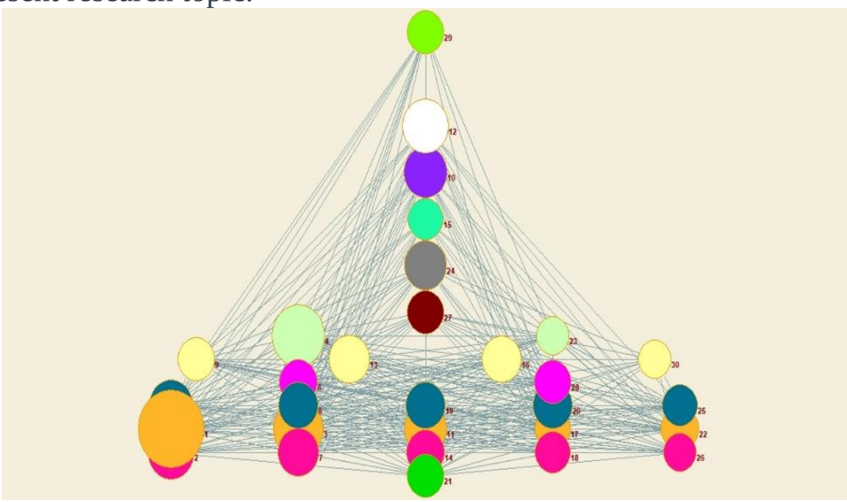

Figure 6. CO-citation Analysis

Table 1. The first 10 Co-cited research papers

SL No	Co-cited Papers
1	"Carvalho T, 2019, V137, Comput Ind Eng, Doi 10.1016/J.Cie.2019.106024"
2	"Zonta T, 2020, V150, Comput Ind Eng, Doi 10.1016/J.Cie.2020.106889"
3	"Zhang W, 2019, V13, P2213, Ieee Syst J, Doi 10.1109/Jsyst.2019.2905565"

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 2, No.3, February : 2025

4	"Susto G, 2015, V11, P812, Ieee T Ind Inform, Doi 10.1109/Tii.2014.2349359"
5	"Li X, 2018, V172, P1, Reliab Eng Syst Safe, Doi 10.1016/J.Ress.2017.11.021"
6	"Zheng S, 2017, P88, 2017 Ieee International Conference On Prognostics And Health Management (Icphm), Doi 10.1109/Icphm.2017.7998311"
7	"Ã‡Inar Z, 2020, V12, Sustainability-Basel, Doi 10.3390/Su12198211"
8	"Khan S, 2018, V107, P241, Mech Syst Signal Pr, Doi 10.1016/J.Ymssp.2017.11.024"
9	"Sateesh Babu G, 2016, Dasfaa 2016. Proceedings: Lncs 9642, P214, Database Systems For Advanced Applications. 21St International Conference, Doi 10.1007/978-3-319-32025-0_14"
10	"Saxena A, 2008, P1, 2008 International Conference On Prognostics And Health Management (Phm)"

Through the SLR conducted it is observed that majority of the articles published in peer reviewed journals used ML algorithms such as Random Forests (RF), Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-means, Logistic Regression, and XG Boost (eXtreme Gradient Boosting) etc., (Carvalho et al., 2019).

Random Forests (RF) is a powerful group learning algorithm globally used in machine learning for both classification and regression tasks. The algorithm operates by forming multitude of decision trees during the training phase and outputs the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees. Random Forests offer robustness and accuracy, making them suitable for various real-world applications. The procedural steps in RF includes the following (Cutler et al. 2007):

- Data Preparation: Random Forests can handle both categorical and numerical data. However, it's essential to pre-process the data, handling missing values, encoding categorical variables, and scaling numerical features if necessary.
- Training the Model: During training, the algorithm builds multiple decision trees. Each tree is trained on a bootstrap sample of the data (sampling with replacement), and at each split, a random subset of features is considered, ensuring diversity among the trees.
- Voting or Averaging: For classification tasks, Random Forests aggregate predictions by majority voting among the trees. For regression tasks, predictions are averaged across the trees.
- Evaluation: The model's performance is evaluated using appropriate metrics such as accuracy, precision, recall, F1-score (for classification), or mean squared error, and R-squared (for regression).

The RF algorithm is used by many researchers after it was introduced by Leo (2001). Some of the important researchers who have used RF algorithm are Prytz et al., 2015; Biau and Scornet, 2016: Carvalho et al., 2017.

Artificial Neural Networks (ANNs) are computational models inspired by the structure and functioning of the human brain. ANNs consist of interconnected nodes (neurons) organized into layers: input layer, hidden layers, and output layer (Biswal and Sabareesh, 2015). These networks are capable of learning complex patterns and relationships from data, making them widely used in various machine learning tasks, including classification, regression, and pattern recognition. The procedural steps in ANN includes the following (LeCun et al. 2015):

- Data Pre-processing: Like any machine learning algorithm, data preprocessing is essential for ANNs. This involves tasks such as normalization, feature scaling, handling missing values, and encoding categorical variables.

- Model Architecture Design: Designing the architecture of the neural network involves determining the number of layers, the number of neurons in each layer, the activation functions, and the type of network (e.g., feed forward, recurrent, convolutional).
- Training the Model: Training an ANN involves feeding the input data through the network, calculating the output, comparing it with the actual output (labels), and adjusting the network's parameters (weights and biases) using optimization algorithms such as gradient descent and its variants.
- Model Evaluation: After training, the performance of the ANN is evaluated using a separate validation dataset or through cross-validation techniques. Metrics such as accuracy, precision, recall, F1-score (for classification), or mean squared error, and R-squared (for regression) are commonly used for evaluation.

The applications, the strengths and challenges and the integration of latest methodologies with ANN are reported in the research works published by Soares and Araújo (2015), Kolokas et al. (2018), Schmidhuber (2015) and Kalra et al. (2016).

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyper plane that best separates the data points into different classes or predicts continuous outcomes. SVM aims to maximize the margin between the hyper plane and the closest data points, known as support vectors, making it robust and effective in high-dimensional spaces. The procedural steps in ANN includes the following (Hsu & Lin, 2002):

Data Processing: SVMs require pre-processing steps such as feature scaling and normalization to ensure that all features contribute equally to the model's decision boundary.

Selecting the Kernel Function: SVMs utilize kernel functions to map the input data into a higher-dimensional space where it becomes linearly separable. Common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid.

During training, SVMs determine the optimal hyper plane by finding the decision boundary that maximizes the margin between classes while minimizing classification errors. This process involves solving a convex optimization problem.

Model Evaluation: After training, the performance of the SVM model is evaluated using validation data or cross-validation techniques. Evaluation metrics such as accuracy, precision, recall, F1-score, or mean squared error are common.

Some of the important research works that are reported recently applying the SVM algorithm are Blomer et al. (2016), Lasisi and Attoh-Okine, (2018); Machado and Mota, 2015; Susto et al., 2016 and Li et al., 2014.

K-Means is a popular unsupervised machine learning algorithm used for clustering data points into a pre-defined number of clusters. It works by iteratively partitioning the data into K clusters, where each data point belongs to the cluster with the nearest mean. K-Means aims to minimize the within-cluster variance, effectively grouping similar data points together (Bach & Vassilvitskii, 2007). The procedural steps in K-Means includes the following (Jain, 2010):

- Initialisation: Randomly initialize K cluster centroids (representing the centre of each cluster) in the feature space.
- Assignment Step: Assign each data point to the nearest cluster centroid based on a distance metric, typically Euclidean distance.
- Update Step: Update the cluster centroids by computing the mean of all data points assigned to each cluster.
- Iteration: Repeat the assignment and update steps until convergence criteria are met, such as stable cluster assignments or a maximum number of iterations reached.
- Finalisation: Once convergence is achieved, the algorithm outputs the final cluster centroids and assigns each data point to its corresponding cluster.

Further, the theory behind K-means is explained in detail in the research article by Blömer, et al., (2016), Mathew et al. (2017) and Anilkumar and Gupta (2018).

Logistic Regression is a widely used supervised machine learning algorithm for binary classification tasks. Despite its name, logistic regression is used for classification rather than regression. It models the probability of the outcome variable (binary response) as a function of the predictor variables, using the logistic function to map the output to the range (Hosmer et al., 2013).

- Data Preparation: Logistic Regression requires pre-processing steps like handling missing values, encoding categorical variables, and scaling numerical features if necessary.
- Model Training: During training, logistic regression estimates the parameters (coefficients) of the logistic function using optimization techniques such as gradient descent or Newton's method. The objective is to maximize the likelihood function or minimize the logistic loss function.
- Model Evaluation: After training, the performance of the logistic regression model is evaluated using validation data or cross-validation techniques. Metrics such as accuracy, precision, recall, F1-score, or ROC-AUC are commonly used for evaluation.
- Prediction: Once trained, the logistic regression model can be used to predict the probability of the outcome variable for new input data. A threshold (usually 0.5) is applied to convert probabilities into binary predictions.

Garcia, et al. (2021), Smith, et al. (2020) and Wang, et al. (2019) are some researchers who have used Logistic regression algorithm for predictive maintenance in the manufacturing sector.

XG Boost, which stands for eXtreme Gradient Boosting, is an optimized and scalable implementation of the gradient boosting algorithm. It is highly regarded for its performance in supervised learning tasks such as classification and regression. XG Boost builds a series of decision trees sequentially, where each tree corrects the errors made by the previous ones. It utilizes a gradient descent optimization technique to minimize a specified loss function, providing superior predictive performance. The procedural steps in XG Boost includes the following (Chen and Guestrin, 2016):

- Data Preparation: XGBoost requires preprocessing steps similar to other machine learning algorithms, including handling missing values, encoding categorical variables, and scaling numerical features.
- Model Training: During training, XGBoost sequentially builds decision trees, with each subsequent tree attempting to correct the errors of the previous trees. The optimization objective is to minimize a specified loss function by iteratively adding trees to the ensemble.
- Parameter Tuning: XG Boost offers a wide range of hyper parameters that can be tuned to optimize the model's performance. This includes parameters related to tree structure, learning rate, regularization, and more.
- Model Evaluation: After training, the performance of the XG Boost model is evaluated using validation data or cross-validation techniques. Common evaluation metrics include accuracy, precision, recall, F1-score, or mean squared error.

XG Boost algorithm is widely being used by the researchers for predictive maintenance. Some of the researchers who have recently published the research articles using XG Boost algorithm are Liu, et al. (2021) and Chen, et al. (2020).

4. Conclusions

Through this systematic review, a thorough examination of the implementation of machine learning algorithms for predictive maintenance applications is conducted. Through the analysis of various studies, it is obvious that machine learning techniques, including Random Forests (RF), Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-means clustering, Logistic Regression, and XG Boost (eXtreme Gradient Boosting), have been broadly applied and explored in the field of predictive maintenance across diverse industries such as manufacturing, healthcare, finance, and more.

Each algorithm displays unique strengths and limitations, which are critical considerations when selecting the appropriate model for a specific predictive maintenance task. For instance, Random Forests excel in handling high-dimensional data and capturing complex interactions among variables, while Support Vector Machine is effective in handling non-linear relationships and achieving robust generalization. Meanwhile, XG Boost has gained popularity due to its superior performance in terms of predictive accuracy and scalability.

Furthermore, the review highlights the importance of feature engineering, data pre-processing, and model evaluation techniques in enhancing the performance and reliability of machine learning models for predictive maintenance. Furthermore, the interpretability of models, especially in safety-critical applications, remains a significant concern and an area for future research. Overall, this systematic review emphasises the potential of machine learning algorithms in transforming predictive maintenance practices by enabling proactive and cost-effective asset management strategies. By using the insights provided in this review, researchers, practitioners, and industry professionals can make informed decisions regarding the selection and deployment of machine learning algorithms for predictive maintenance tasks, ultimately leading to improved equipment reliability, reduced downtime, and enhanced operational efficiency.

References

[1] Abidi M H, Mohammed M K & Alkhalefah H (2022), "Predictive Maintenance Planning for Industry 4.0 Using Machine Learning for Sustainable Manufacturing", Sustainability, Vol 14, 2022, pp1-27.

[2] Amruthnath N, & Gupta T (2018), "A research study on unsupervised machine learning algorithms for fault detection in predictive maintenance", 5th International conference on industrial engineering and applications (ICIEA), IEEE, 2018, pp 355–361.

[3] Anilkumar E N & Sridharan R (2019), "Sustainable Supply Chain Management: A Literature Review and Implications for Future Research", International Journal of System Dynamics Applications (IJSDA), Vol 83, 2019, pp15-52.

[4] Arena S, Florian E, Zennaro I, Orru P F & Sgarbossa F (2022), "A novel decision support system for managing predictive maintenance strategies based on machine learning approaches" Safety Science, Vol 146, 2022, pp1-19.

[5]Arthur D, & Vassilvitskii S (2007), "k-means++: The advantages of careful seeding", *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 1027-1035. doi:10.1137/1.9781611972774.84

[6] Bach, M P, Topalovic A, Krstic Z, & Ivec A (2023), "Predictive Maintenance in Industry 4.0 for the SMEs: A Decision Support System Case Study Using Open-Source Software, Designs, Vol 7, 2023, pp98-121.

[7] Biau G & Scornet E (2016), "A random forest guided tour", An Official Journal of the Spanish Society of Statistics and Operations Research, Vol 25, 2016, pp197–227.

[8] Biswal S & Sabareesh G R (2015), "Design and development of a wind turbine test rig for condition monitoring studies", International conference on industrial instrumentation and control, IEEE, 2015, pp891–896.

[9] Blomer J, Lammersen C, Schmidt M, & Sohler C (2016), "Theoretical analysis of the k-means algorithm – a survey", In Algorithm engineering: Selected results and surveys, Vol 9220, 2016, pp81–116.

[10] Borenstein M, Hedges L V, Higgins J P T, & Rothstein H R (2021), Introduction to Meta-Analysis (2nd ed.). John Wiley & Sons.

[11] Bousdeki A, Lepenio K, Apostolou D & Mentzas G (2019), "Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0", IFAC Papers Online, Vol 52, 2019, pp 607–612.
[12] Canizo M, Onieva E, Conde A, Charramendieta S, & Trujillo S (2017), "Real-time predictive maintenance for wind turbines using Big Data frameworks", 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 2017, pp. 70-77.
[13] Carvalho T P, Soares M N, Vita R, Franscisco R P, Basto J P& Alcala G S (2019) "A systematic literature review of machine learning methods applied to predictive maintenance", Computers & Industrial Engineering, Vol137, 2019, pp 1-10.
[14] Chen H, Wu X, & Zhang L (2020), "Predictive Maintenance in Manufacturing: A Case Study on CNC Machines Using XGBoost", International Journal of Advanced Manufacturing Technology, Vol 110, 2020, pp 3177-3188.
[15] Cutler D R, Edwards Jr T C, Beard K H, Cutler A, Hess K T, Gibson J & Lawler J J (2007), "Random forests for classification in ecology", Ecology, Vol 88, 2007, pp 2783-2792.
[16] Dalzochio J, Kunst R, Pignaton E, Binotto A, Sanyal S, Favilla J & Barbosa J (2020), "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenge, Computers in Industry, Vol123, 2020, pp1-15.
[17] Garcia M, Rodriguez P & Martinez S (2021), "Predictive Maintenance in Automotive Manufacturing: A Logistic Regression Approach", Vol 50, 2021, pp 40-52.
[18] Hosmer Jr D W, Lemeshow S & Sturdivant R X (2013), "Applied logistic regression", John Wiley & Sons, Vol. 398, 2013.
[19] Hsu C & Lin C J (2002), "A comparison of methods for multiclass support vector machines", IEEE Transactions on Neural Networks, Vol 13,2002, pp 415-425.
[20] Jain A K (2010), "Data clustering: 50 years beyond K-means", Pattern Recognition Letters, Vol 31, 2010, Vol 651-666.
[21] Kalra A, Kumar S & Waliya S S (2016), "ANN training: A survey of classical and soft computing approaches", International Journal of Control Theory and Applications, Vol 9, 2016, pp 715–736.
[22] Kolokas N, Vafeiadis T, Ioannidis D & Tzovaras D (2018), "Forecasting faults of industrial equipment using machine learning classifiers", International symposium on innovations in intelligent systems and applications (INISTA), IEEE, 2018, pp. 1–6.
[23] Kulkarni K, Devi U, Sirighee A, Hazra J & Rao P (2018), "Predictive maintenance for supermarket refrigeration systems using only case temperature data", The American Control Conference (ACC), IEEE, 2018, pp 4640–4645.
[24] Lasisi Kanawaday A & Sane A (2017), "Machine learning for predictive maintenance of industrial machines using IoT sensor data", International conference on wireless communications, networking and mobile computing (WiCOM), IEEE, 2017, pp87–90.
[25] Lasisi A & Attoh-Okine N (2018), "Principal components analysis and track quality index: A machine learning approach", Transportation Research Part C: Emerging Technologies, Vol 91, 2018, pp 230–248.
[26] LeCun Y, Bengio Y & Hinton G (2015), "Deep learning", Nature, Vol 521, 2015, pp 436-444.
[27] Leo B (2001), "Random forests", Kluwer Academic Publishers, Vol 45, 2001, pp 5–32.
[28] Li H, Parikh D, He Q, Qian B, Li Z, Fang D & Hampapur A (2014), "Improving rail network velocity: A machine learning approach to predictive maintenance", Transportation Research Part C: Emerging Technologies, Vol45, 2014, pp 17–26.
[29] Liu Y, Yang L & Wang Y (2021), "Application of XGBoost Algorithm in Predictive Maintenance for CNC Machine Tools", Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, Vol 235, 2021, pp 603-614.

[30] Machado R G V & Mota H (2015), "Simple self-scalable grid classifier for signal denoising in digital processing systems", IEEE 25th international workshop on machine learning for signal processing (MLSP), 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 2015, pp. 1057-1061.
[31] Moher D, Liberati A, Tetzlaff J, Altman D G & PRISMA Group (2009), "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement", PLOS Medicine, Vol 6, e1000097. https://doi.org/10.1371/journal.pmed.1000097.
[32] Narayanan A E, Sridharan R and Ram Kumar P N (2019), "Analyzing the interactions among barriers of sustainable supply chain management practices: A case study", Journal of Manufacturing Technology Management, Vol. 30, 2019, pp. 937-971.
[33] Parida A & Kallenberg V (2020), "Journal of Quality in Maintenance Engineering", Emerald Publishing Limited, Vol 12, 2020, pp 239-251.
[34] Prytz R, Nowaczyk S, Rögnvaldsson T & Byttner S (2015), "Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data", Engineering Applications of Artificial Intelligence, Vol 41, 2015, pp 139–150.
[35] Schmidhuber J (2015), "Deep learning in neural networks: An overview", Neural Networks, Vol 61, 82015, pp 5–117.
[36] Shin J H, Jun H B & Kim J G (2018), "Dynamic control of intelligent parking guidance using neural network predictive control", Computers & Industrial Engineering, Vol 120, 2018, pp 15–30.
[37] Smith A & Jones B (2020), "Enhancing Asset Reliability Through Predictive Maintenance: A Review", International Journal of Reliability and Quality Performance Management, Vol 45, 2020, pp 321-335.
[38] Smith J, Johnson A & Brown C (2020) "Predictive Maintenance in Manufacturing Using Logistic Regression Model", International Journal of Manufacturing Research, Vol 25, 2020, pp 345-362.
[39] Soares S G & Araújo R (2015), "An on-line weighted ensemble of regressor models to handle concept drifts", Engineering Applications of Artificial Intelligence, Vol 37, 2015, pp 392–406.
[40] G A Susto & A Beghi (2016) "Dealing with time-series data in Predictive Maintenance problems", IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 2016, pp. 1-4, doi: 10.1109/ETFA.2016.7733659.
[41] Wang L, Liu Y & Chen Q (2019), "Application of Logistic Regression for Predictive Maintenance in Semiconductor Manufacturing", IEEE Transactions on Semiconductor Manufacturing, Vol 34, 2019, pp 245-253.