Journal of Trends and Challenges in Artificial Intelligence

Vol. 03, No. 03 (2026) 151-164, DOI: 10.61552/JAI.2026.03.004 - http://jai.aspur.rs

# A COMPARATIVE EVALUATION OF SUPERVISED MACHINE LEARNING METHODS TO PREDICT MACHINERY CONDITION AND RELIABILITY USING LUBRICATION PARAMETERS

Qadeer Ahmed [1]
Hasanur J. Molla
Ziad M. Aloqifi
Dahham M. Anazi
Saad H. Dossary

Received 02.07.2025.
Revised 09.09.2025.
Accepted 15.10.2025.

Keywords:

*Supervised Machine Learning, Lubrication, Artificial intelligence, Machine learning; Comparative Model Assessment; Machinery Health Prediction, Reliability.*

**Original research**

A B S T R A C T

*Effective machinery lubrication is essential for ensuring the reliability and availability of systems, as it plays a vital role in the smooth operation of machines, much like blood in the human body. Inadequate lubrication and poor management practices can have severe consequences, including equipment failures, compromised safety, and increased maintenance costs, particularly in asset-intensive industries such as oil, gas, and petrochemical facilities. To address this challenge, this study utilizes artificial intelligence, machine learning, and data mining techniques to develop and compare various predictive models for monitoring machine health. The ultimate goal of this research is to identify the most effective approach for predicting machine health using lubrication condition data. To achieve this, relevant lubrication features were carefully selected and incorporated into model development, with their significance in predicting machine health subsequently evaluated. An analysis of real-world lubrication data from diverse machines revealed that the random forest algorithm outperformed other algorithms in meeting our requirements. The findings of this study demonstrate the potential of lubrication data to forecast machine health and improve equipment reliability, providing valuable insights that enable researchers to enhance lubrication management strategies. Overall, this research contributes to the development of enhanced lubrication management strategies, which can help reduce downtime, promote more efficient and safe operations, and ultimately benefit asset-intensive industries.*

© 2026 Journal of Trends and Challenges in Artificial Intelligence

## 1. INTRODUCTION

Lubrication in rotating equipment serves as the lifeblood of machines, playing a vital role in the smooth operation of all types of rotating machinery. Different machines require various types of lubrication, such as grease, oil, and other specialized lubricants (Laval et al., 2013). The choice of lubricant depends on the specific application, with some requiring mineral-based oils and others necessitating synthetic-based oils with tailored additive

[1] Corresponding author: Qadeer Ahmed
Email: qahmed50@gmail.com

packages. A lubricant is typically a mixture of base oil, which constitutes 70-90% of the lubricant and is responsible for its primary characteristics (Pichler et al., 2023). The remaining portion consists of additives, which are complex chemicals designed to enhance the lubricant's performance for specific applications. Historically, lubricants were limited to mineral oils, vegetable oils, fats, and greases (Uppar et al., 2023). However, in recent years, the concept of lubrication has expanded to include solid lubricants, synthetic oils, water-based lubricants, and gas-based lubricants. This broadening of the concept has been accompanied by a deeper understanding of lubricant performance and its limitations. The typical composition of a lubricant is illustrated in Figure 1. Lubricating oil is a generic term that encompasses a wide range of products, characterized by hundreds of base chemicals and additives. The most common lubricating oils are derived from crude oil distillate fractions, although synthetic and plant-based lubricating oils are also used. Petroleum hydrocarbon distillates, which are commonly used as base oils, typically consist of paraffinic or naphthenic compounds. These base oils are obtained through conventional refining of petroleum and are considered mineral-based. In contrast, base oils obtained through more complex processes, such as chemical synthesis, are considered synthetic. Semi-synthetic lubricants, on the other hand, combine mineral and synthetic base oils. Some commonly available synthetic oils include Polyalpha-olefin (PAO), synthetic esters, and Polyalkylene glycols (PAG) (Ismail et al., 2024).


**Figure 1.** Lubricant - Base Oil and Additives

Synthetic oil represents the pinnacle of lubrication technology, offering unparalleled performance and protection for modern rotating machines and engines (Pichler et al., 2023). In contrast to mineral oil, which is extracted from crude oil, synthetic oil is carefully crafted in laboratories through a process of chemical synthesis. This controlled environment enables the precise manipulation of the oil's composition, resulting in enhanced stability and performance attributes. The optimized composition of synthetic oil ensures superior chemical stability, consistent viscosity, and reduced sludge formation, even in the most extreme operating conditions. The primary function of a lubricant is to form a thin oil film under operating conditions, which separates contacting surfaces and provides effective lubrication. However, lubricants play a multifaceted role, extending beyond mere lubrication to include cooling, cleaning, protecting contacting surfaces, sealing, and power transmission. These diverse functions are crucial to the smooth operation and longevity of rotating machines and engines. The major functions, properties, and types of a typical lubrication system are illustrated in Figure 2, providing a comprehensive overview of the complex role that lubricants play in modern machinery.

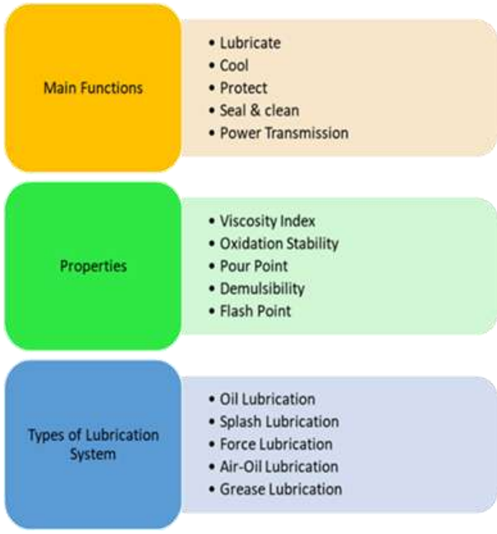


**Figure 2.** Major functions, properties & types of systems

Lubricants possess a diverse array of properties that govern their physical and chemical behavior, making it essential to understand these characteristics in order to choose the most suitable lubricant for a specific application. Key properties such as viscosity, viscosity index, oxidation stability, demulsibility, and flash point are particularly important, as they have a significant impact on the performance and longevity of the equipment or system being lubricated (Singh et al., 2021).

The selection of a lubrication method is a critical consideration in lubrication, as it is largely dictated by the specific lubrication needs of the application. Various lubrication methods exist, each with its unique advantages and limitations, and are designed to cater to distinct demands and constraints. These constraints include parameters such as load-carrying capacity, temperature tolerance, speed ranges, and compatibility with surrounding environments (Abbas & Metwalli, 2025). To achieve optimal results and prolong component lifespan, it is essential to carefully consider these variables when selecting a lubrication technique. There are several types of lubrication systems, including Oil Lubrication Systems, also known as Loss Lubrication Systems (Parez, 2025). In these systems, oil or liquid grease creates a protective film around moving parts, which is replenished periodically by an automated system equipped with an electric oil pump. Another category is Splash Lubrication Systems, commonly found in smaller gasoline engines, where lubricating oil collects in an oil reservoir and is splashed onto the cylinder walls when the engine operates. Recirculating Oil/Force Lubrication Systems serve both lubrication and cooling purposes for bearings and gearboxes, while Air-Oil Lubrication Systems utilize compressed air streams infused with minute amounts of oil to convey precise doses of lubricant to target locations. This arrangement is suitable for massive machinery and manufacturing equipment, offering extended bearing lifespans and enhanced productivity due to minimized interruptions. In the realm of Grease Lubrication Systems, Dual Line and Progressive Systems emerge as dominant configurations. Dual Line Lubrication Systems boast modularity, facilitating flexible expansions and adaptations suited to sprawling industrial complexes housing vast arrays of machinery requiring simultaneous lubrication. In contrast, Progressive Lubrication Systems are better-suited for compact to mid-scale apparatus necessitating constant lubrication flows.

Ensuring the integrity of in-service lubricant conditions is vital for safeguarding critical rotating equipment and preventing costly downtime. To achieve this, companies utilize advanced Lubrication Condition Monitoring (LCM) systems, which oversee thousands of vital machines containing vast quantities of lubricants across all operational facilities (Whitby, 2021). These LCM systems automate the sampling process, adhering to established maintenance schedules for enrolled equipment, and dispatch collected samples to specialized laboratories for thorough analysis. LCM serves as a multifaceted tool, functioning as an early warning system for emerging problems, as well as a means for diagnosing and predicting faults under condition-based maintenance protocols. The detailed information obtained from lubricant testing makes LCM an essential condition monitoring technique, providing unparalleled insights into the complex relationships between machinery and lubricant states. By leveraging this valuable resource, our research utilizes lubrication data from diverse machines to develop cutting-edge artificial intelligence-driven machine learning models. We employ various algorithms to compare their effectiveness and accurately predict machinery health, enabling the anticipation of potential failures, elevation of asset reliability, and amplification of facility availability. Ultimately, our goal is to establish a foundation for efficient and sustainable operations by harnessing the power of LCM and machine learning. By doing so, we aim to optimize maintenance schedules, reduce downtime, and improve overall equipment performance, thereby contributing to the development of more reliable and resilient industrial systems.

The remainder of the of this manuscript are structured in the following manner: A comprehensive review of existing literature pertaining to the utilization of lubrication data for predicting equipment health is presented in Section 2. Section 3 provides an in-depth examination of the research methodology employed, including data collection protocols, performance evaluation metrics, and the application of various machine learning algorithms. The results of the study, including an assessment of model performance, are thoroughly discussed in Section 4. Finally, Section 5 offers a conclusive summary of the overall investigation, highlighting key findings and proposing potential avenues for future research to further advance the field.

## 2. LITERATURE REVIEW

To provide context for the research effort focused on predicting machinery health through lubrication condition data, a systematic literature review was conducted to synthesize the existing body of knowledge in this domain. This comprehensive review aimed to identify the current state-of-the-art in lubrication condition monitoring, predictive maintenance, and the applications of artificial intelligence (AI) and machine learning in asset management. Machine learning, a subfield of AI, involves the development of algorithms and statistical models that enable computers to learn from data without being explicitly programmed. According to (Shahzad, 2022), AI refers to the development of computer systems that can perform tasks that would normally require human intelligence, such as learning, reasoning, problem-solving, perception, and language understanding. By leveraging these advanced technologies, the research seeks to explore the potential of lubrication condition data in predicting machinery health and optimizing maintenance schedules. The literature review establishes a comprehensive understanding of the current landscape of lubrication condition monitoring and predictive maintenance, as well as the applications of AI and machine learning in asset management. By analysing existing research and methodologies, this study aims to identify knowledge gaps and develop innovative approaches to predicting machinery health and reducing downtime, ultimately enhancing the reliability and performance of industrial assets.

Numerical analysis and machine learning techniques have been explored to optimize the hydrodynamic design of thrust bearings. A dataset of varying loads and rotational speeds was used as input to train regression models that predict the bearing's behaviour. Four types of regression models were developed and compared for their accuracy in predicting the bearing's operating conditions, including simple and multivariable, linear, polynomial, and Support Vector Machine (SVM) models (Moder et al. 2018). A novel approach to monitoring the hydromechanical system of a rotary machine featuring fluid-film bearings has been presented. By feeding measurement data into an artificial neural network, the system's state can be accurately recognized, with the trained neural network achieving a recognition accuracy of over 98% (Kornaeva et al., 2020). This high level of accuracy enables the detection of even minor deviations in the system's behaviour, making it a valuable tool for predictive maintenance. A predictive analytics methodology has been developed to assess the performance state of marine journal bearings. This involves simulating the Reynolds equation to produce a comprehensive dataset of bearing operation states, which is then used to train, test, and validate multiple machine learning algorithms (Rossopoulos & Papadopoulos, 2021). Feature selection analysis is also performed to accurately categorize bearing operational states and support proactive maintenance strategies. Induction motors (IMs) rely heavily on bearings, making timely fault diagnosis essential for dependable operation. A multifaceted approach has been devised to address this requirement, involving the careful selection of salient features and their incorporation into three separate classification algorithms: Complex Decision Tree (CDT), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM) (Choudhary et al., 2020). Comparative analyses have shown that SVM exhibits superior performance relative to CDT and LDA, enabling adaptive recognition of bearing faults in IMs and reducing the likelihood of unanticipated system downtime. This research explores the speed and accuracy of decision-making processes in Plant Health Management (PHM), with a focus on ML-powered anomaly detection. By leveraging historical data to train linear regression models, predictions are generated based on lagged values in time-series data. The predicted outcomes are then evaluated using a scoring model, revealing remarkably high prediction accuracy scores. This methodology demonstrates the feasibility of equipment condition prediction using ML technology, highlighting its potential for improving maintenance decision-making (Soleh et al., 2019). A novel approach to fault diagnosis is introduced through the Radial Concave Deviation (RCD) method, which utilizes particle boundary signals to analyse wear particle features. The RCD output enables the determination of additional feature parameters, such as shape and size, which are then used to identify debris features and types through various classification methods, including linear and quadratic discriminant analysis, naive Bayes, and CART. The results of this study enable maintenance teams to diagnose faults accurately, reducing downtime and improving overall system reliability (Yuan et al., 2016). The importance of oil analysis in mechanical systems is also highlighted, as most systems are oil-lubricated and oil analysis provides valuable insights into machine condition. This research explores the use of random forests, neural networks, and logistic regression models trained on oil analysis data to classify machine conditions. The results show that Random Forests (RF) outperform other classifiers, and feature importance interpretations align with industry expertise, demonstrating RF's potential as a diagnostic tool in predictive maintenance (Keartland & Van, 2020). Furthermore, classical methods for predicting component reliability and failure often struggle with limited data and selecting the right distribution, failing to account for the impact of time series patterns on component behaviour. To address this, a hybrid approach is proposed, combining classical methods with machine learning techniques, specifically Support Vector Regression (SVR) and Least Square Support Vector Regression (LSSVR). The results show that the classical-LSSVR combination yields higher accuracy than SVR alone, enhancing the prediction of component failure and providing a more robust approach to maintenance decision-making (Arumsari et al., 2021). A supervised machine learning framework is presented in this paper, which utilizes a Random Forest Classifier (RF) to determine the operating conditions of lubricant oil in diesel engines. The approach achieves a high level of accuracy, with a recall of 97.9%, precision of 99.5%, and F1-score of 98.7%, and is shown to be effective in identifying oil operating conditions. The study also highlights the importance of engineered features derived from raw data in improving condition determination (Malaguti et al., 2021). In another study, an Artificial Neural Network (ANN) model is used to predict the age of lubrication oil based on select monitoring indicators. The model demonstrates strong predictive performance, with an R-squared value of 0.8176, an MSE of 1191, and an MAD of 26. The results suggest that lubrication oil parameters remain within acceptable limits even beyond the recommended service duration, providing valuable insights for predicting the remaining useful life of the oil (Mohammad & Muhammad, 2021).

A comprehensive literature review revealed key themes in predicting machinery health using lubrication condition data, including the importance of integrating real-time sensor data and machine learning approaches. The review highlighted the successful application of supervised, unsupervised, and reinforcement learning techniques in diagnosing anomalies and detecting faults. However, it also identified a gap in fusing lubrication condition data with other modalities, emphasizing the need for interdisciplinary collaborations. The review emphasized practical concerns such as data quality, noise reduction, and scalability, and provided a framework for further research, charting a course toward novel contributions and tangible impacts on industrial practice.

## 3. RESEARCH METHODOLOGY

In this research study, we conducted a comprehensive investigation (Figure 3) involving the collection of lubrication datasets from diverse machinery sources. This approach ensured the inclusion of various types of data, ultimately yielding more informative and useful research findings. The collected datasets were then utilized to train and validate an artificial intelligence model. The primary objective of this research endeavour was to perform a comparative analysis, assessing the efficacy of various machine learning models in predicting asset health based on lubrication data. By evaluating the performance metrics of these models, we aimed to identify those that exhibited superior predictive accuracy. This, in turn, would enable the development of strategies for optimized maintenance, timely intervention, and efficient fault detection in industrial settings.

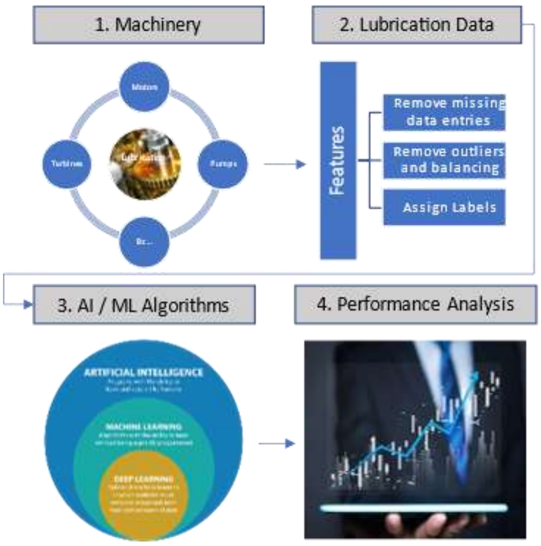


**Figure 3.** Research steps and methodology

## 3.1 Data Collection and Processing

This investigative study involved the acquisition of lubrication data from a fleet of rotating machines operating within a facility, which encompassed a total of 13 key features carefully selected to provide a comprehensive characterization of lubrication conditions as shown in Table 1.

**Table 1.** Details of features and description

| SN | Attribute Name | Description |
|---|---|---|
| 1 | Calcium (PPM) | Contaminant from detergent additive in oil |
| 2 | Chromium (PPM) | Wearing of alloyed components. The condemnation limit for Chromium is 5 ppm |
| 3 | Color | Represent degradation of oil |
| 4 | Copper (PPM) | Primarily due to wearing of Babbitt bearing. limit for copper is 10 ppm |
| 5 | Iron (PPM) | Wearing of ferrous components. The limit for Iron is 30 ppm |
| 6 | Lead (PPM) | Wearing of Babbitt bearing. The set condemnation limit for lead is 10ppm |
| 7 | Magnesium (PPM) | Source is detergent additive in oil. The set limit is 50% of Magnesium in new oil |
| 8 | Moisture (PPM) | Multiple sources, the moisture should be below saturation point of the associated lube oil. The limit is 500 ppm. |
| 9 | Phosphorous (PPM) | Source is AW/EP additive in lube oil. The set contamination list is 50% of Phosphorous in fresh oil |
| 10 | Silicon (PPM) | Sources are inhibitor additive in lube oil or external contaminants. set limit is 10 ppm |
| 11 | Solids | Identifies varnish potential in lube oil. The contamination limit is 5 |
| 12 | Total Acid Number (mg KOH/g) | Degree of lube oil degradation, the contamination limit is 0.2 mg KOH/g |
| 13 | Viscosity (cSt) at 40 deg C | Represent the degradation, contamination or mixing with another lube oil |

After collecting the data, a rigorous review and cleansing process was conducted to eliminate outlier values and annotate the data to reflect the corresponding machine conditions, thereby ensuring the quality and integrity of the data. The resulting dataset consists of 501 individual machinery records, covering a range of critical equipment such as compressors, pumps, and other vital machinery assets. This robust dataset provides a solid foundation for subsequent analysis and modelling endeavours.

## 3.2 Algorithm performance criterion

To understand and compare the performance of the algorithms, we employed three individual metrics, which are explained below, and also utilized the F1 score to consolidate precision and recall into a single metric, providing a comprehensive view of the model's performance (Naidu et al., 2023; Belavagi & Muniyal, 2016). The algorithms applied in this study include logistic regression (LR), Bagging, random forest (RF), GBM, Adaboost, Xgboost, decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN), and Naive Bayes. In the context of machine learning and model evaluation, accuracy is a fundamental metric used to assess a model's performance during both the training and testing phases. Accuracy measures the frequency of correct predictions made by the model compared to the total number of attempts. Specifically, training accuracy evaluates how well the model fits the training dataset after learning the patterns and relationships within it. In contrast, testing accuracy assesses the model's performance on unseen data, providing insight into its ability to generalize.

$$Accuracy = \frac{Correctly\ predicted\ observations}{Total\ number\ of\ observations} \tag{1}$$

In addition to other metrics, precision is another key metric used to compare and evaluate the model's performance. Precision is a measure of how often a machine learning model correctly predicts the positive class. It can be calculated by dividing the number of true positives (correct positive predictions) by the total number of instances that the model predicted as positive, including both true positives and false positives. A higher precision value indicates a better model, as it suggests that the model is more accurate in its predictions and consistently predicts the target class correctly.

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \tag{2}$$

The metric, Recall, measures how well the model detects all instances of the positive class. It is also known as sensitivity or true positive rate.

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (NP)} \tag{3}$$

Accuracy, precision, and recall are all crucial metrics for evaluating the performance of a machine learning (ML) model. Since none of these metrics provides a comprehensive picture of model quality, it is essential to consider them jointly or select the one that best suits your specific scenario. Understanding the values of these metrics is vital, as a model with high precision but low recall may miss many actual positive cases due to its strict prediction criteria. On the other hand, a model with low precision but high recall may catch many actual positive cases, but it may also generate a large number of false alarms, as it predicts positive outcomes for almost every instance. The balance between precision and recall often depends on the problem being addressed, as the consequences of missing an actual positive case can be more severe than incorrectly predicting a positive outcome when it is not present. For instance, in medical diagnosis, high recall is critical to ensure that no diseases go undetected, even if some healthy individuals are wrongly diagnosed. In contrast, high precision is preferred in spam filtering to prevent genuine emails from being mistakenly marked as spam. By examining these metrics, developers can gain insights into a classifier's performance across various aspects and tune its parameters to achieve optimal results based on the application's specific requirements.

In machine learning and data science, F1 score or F-measure is critical (Chowdhury & Schoen, 2020). Research paper classification using supervised machine learning techniques. In 2020 intermountain engineering, technology and computing (IETC) (pp. 1-6). IEEE. F measure is a statistical measure used to evaluate the performance of a classification model, especially in cases where there is an imbalance between classes. The F1-score is the harmonic mean of precision and recall, providing a balanced measure of both.

$$F1\ Score = 2\ X\ (\frac{Precision\ *\ Recall}{Precision\ +\ Recall}) \tag{4}$$

F1-Score is a combined measure and allows a complete view of the model's overall performance. The use of the F1-score is prevalent because it provides a clear, singular metric that encapsulates the trade-off between precision and recall, making it easier to compare different models' performances, especially in imbalanced datasets where accuracy alone may not give a full picture of a model's effectiveness.

## 3.3 Machine learning Algorithms

In this paper, we utilized ten algorithms to compare their performance in predicting machine condition or health based on oil analysis data, where a value of "0" indicates that the rotating equipment is in poor condition and a value of "1" indicates good condition. Notably, all machine learning algorithms were assessed without hyperparameter tuning in this study. Although hyperparameter tuning, which involves selecting the optimal hyperparameters and settings before training a model, can significantly impact the model's performance, it was not performed in this investigation. This is a challenging problem, as the number of possible hyperparameter combinations can be extremely large, and the relationship between hyperparameters and model performance is often complex.

This study conducts a comprehensive examination of the predictive capabilities of ten distinct algorithms in forecasting the health condition of machinery through the analysis of lubricant data. By employing a binary classification framework, the research categorizes machinery into two distinct states: a score of 0 indicates equipment in a degraded or faulty condition, while a score of 1 denotes machinery operating at optimal levels. A key aspect of this investigation is its evaluation methodology, wherein each of the ten machine learning models is assessed using its default parameter settings, without any hyperparameter fine-tuning. This deliberate choice allows the researchers to evaluate the intrinsic effectiveness and adaptability of these algorithms in real-world scenarios, without the benefit of customized calibration. As a result, the findings provide valuable insights into the native strengths and weaknesses of each model, offering practitioners and researchers a clearer understanding of what to expect from these tools in practical applications. The use of lubricant analysis data as the basis for prediction adds an additional layer of complexity and relevance, as lubricants play a critical role in the maintenance and operation of rotary equipment, serving both as a medium for reducing friction and as a potential indicator of underlying issues within the machine. Therefore, developing accurate predictive models that can effectively interpret such data holds significant promise for enhancing preventive maintenance strategies, reducing downtime, and improving overall asset reliability. By comparing the performance of multiple algorithms across this specific domain, this study aims to contribute meaningfully to the ongoing discussion surrounding the application of machine learning techniques in industrial diagnostics and prognostics.

Machine learning can be categorized into several types, including supervised, unsupervised, and semi-supervised learning, which are distinguished by their purposes and the way the underlying machine is trained (Berry et al., 2020). In this study, we employed supervised machine learning, which utilizes programmed algorithms that learn and optimize their operations by analysing input data to make predictions within an acceptable range. As new data is fed into these algorithms, they tend to make more accurate predictions. The process of supervised machine learning involves using a labelled training dataset to train the underlying algorithm, which is then applied to an unlabelled test dataset to categorize it into similar groups (Uddin et al., 2019). Supervised learning algorithms are well-suited for two types of problems: classification problems and regression problems. In classification problems, the output variable is discrete, meaning it is categorized into distinct groups or categories, such as "red" or "black", or "0" and "1". In contrast, regression problems involve a continuous

output variable, such as the risk of machinery failure, which is a real value. In the following subsections, we provide a brief overview of the commonly used supervised machine learning algorithms for prediction.
Some Key Insights of the algorithms are:

### 3.3.1 Logistic Regression

Logistic regression is a supervised learning technique utilized for classification tasks, particularly suited for predicting categorical outcomes. This statistical approach forecasts the likelihood of a specific event occurring, based on one or multiple input variables. (Hosmer et al., 2013) It is most commonly applied to binary classification scenarios, where the outcome can have only two possible values, such as 0 or 1, yes or no, or true or false. (Pramesti, et al., 2016) used logistic regression to study faults in induction motors through high frequency signal of motor electric current. The logistic regression equation is given by:

$$p = \frac{1}{(1 + e^{-z})}) \qquad (5)$$

where:
p: predicted probability,
z: linear combination of inputs (x) and weights (w),
e: base of the natural logarithm.

### 3.3.2 Bagging

Bootstrap Aggregating, commonly referred to as Bagging, is a powerful ensemble learning method that enhances the robustness and precision of machine learning models. First introduced by Leo Breiman, this technique involves combining multiple models to improve overall performance (Breiman, 1996). By leveraging bagging, you can substantially boost the reliability and accuracy of your predictions, particularly when working with intricate or noisy datasets, ultimately leading to more trustworthy and effective machine learning outcomes.

### 3.3.3 Random Forest

A Random Forest is a robust ensemble learning technique that leverages the collective power of multiple decision trees to enhance the accuracy and reliability of predictions for both classification and regression problems (Breiman, 2001). By aggregating the outputs of numerous decision trees, Random Forests mitigate the risk of overfitting and significantly improve the overall performance of the model, resulting in more precise and dependable predictions compared to using a single decision tree.
Overall, Random Forests offer a powerful tool for tackling various supervised learning challenges, providing excellent predictive performance and interpretability through feature importance scores.

### 3.3.4 Gradient Boosting Machine

The Gradient Boosting Machine (GBM) is a highly effective and versatile ensemble learning algorithm, widely utilized for both classification and regression tasks. By iteratively combining multiple simple models, GBM creates a robust and accurate predictive model, making it a popular choice in both industry and academic settings (Trizoglou et al., 2021). Its exceptional performance, flexibility, and relatively straightforward implementation have contributed to its widespread adoption. While GBM offers impressive predictive power and interpretability, its success relies on careful hyperparameter tuning, and its computational demands can be significant, requiring careful consideration and optimization to unlock its full potential.

### 3.3.5 AdaBoost

AdaBoost, or Adaptive Boosting, is a renowned ensemble learning algorithm introduced by Yoav Freund and Robert Schapire in 1996, designed to tackle classification and regression tasks. Its primary objective is to aggregate multiple weak models, transforming them into a robust and accurate predictive model (Freund, & Schapire, 1996). Although AdaBoost is characterized by its simplicity and respectable performance, it may encounter challenges with specific data types or scenarios, limiting its effectiveness. Nevertheless, it remains a valuable and reliable component in the machine learning arsenal. Notably, AdaBoost often demonstrates high recall during the training phase, making it a useful technique for identifying key patterns and relationships within the data, despite potential limitations in certain applications.

### 3.3.6 XGBoost

XGBoost is a highly optimized, open-source library that leverages the Gradient Boosting framework to deliver exceptional efficiency, flexibility, and portability. By implementing a range of machine learning algorithms, XGBoost enables rapid and accurate solutions for diverse problems, including classification, regression, ranking, and custom objectives (Zhang et al., 2018). Its unique blend of speed, adaptability, and performance has solidified XGBoost's position as a leading solution for a wide range of machine learning applications, driving advancements and innovation in the field. With its versatility and exceptional capabilities, XGBoost has become the preferred choice for many machines learning practitioners, allowing them to tackle complex challenges with ease and precision.

### 3.3.7 Decision Tree

A Decision Tree is a supervised learning algorithm utilized for both classification and regression tasks, which operates by constructing a hierarchical, tree-like structure of decisions (Uddin et al., 2019). Within this model, internal nodes signify a specific attribute or feature being evaluated, while the branches emanating from these nodes represent the possible outcomes of this evaluation. Ultimately, the leaf nodes at the terminus of the tree denote the predicted class label or target value, providing a clear and interpretable representation of the decision-making process. This intuitive framework enables Decision Trees to effectively categorize data or predict continuous values, making them a fundamental and widely used technique in machine learning. Decision Trees serve as a fundamental component in many machines learning pipelines, offering transparency, efficiency, and effectiveness in solving a wide array of problems.

### 3.3.8 Support Vector Machine

A Support Vector Machine (SVM) is a robust and versatile supervised learning algorithm, widely employed for both classification and regression tasks (Burges, 1998). Its key strength lies in its capacity to effectively manage high-dimensional data, identifying the optimal hyperplane that maximally distinguishes between classes. SVM's exceptional ability to navigate complex, non-linear relationships and its resilience to noisy data have cemented its position as a preferred choice in numerous real-world applications. By finding the most optimal boundary between classes, SVM enables accurate predictions and classifications, making it a powerful tool in the realm of machine learning, with a broad range of applications across various industries and domains (Orrù et al., 2020).

### 3.3.9 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a supervised learning algorithm utilized for both classification and regression tasks, which operates by assessing the similarity between a new instance and existing data points to predict the target variable (Raviya & Gajjar, 2013). Performance Evaluation of different data mining classification algorithm using WEKA. Indian Journal of Research, 2(1), 19-21. This algorithm is characterized by its simplicity and effectiveness, making it a popular choice for various applications. KNN excels in handling complex, non-linear boundaries and offers a straightforward implementation, which contributes to its appeal. However, its performance is heavily dependent on the selection of the optimal k value, and it can be vulnerable to the influence of noisy data, necessitating careful consideration and tuning to achieve optimal results. Despite these limitations, KNN remains a valuable and widely used technique in machine learning, providing a robust and intuitive approach to classification and regression tasks.

### 3.3.10 Naïve Bayes

Naive Bayes is a class of probabilistic machine learning algorithms that leverage Bayes' theorem to predict the likelihood of an event based on prior knowledge of related conditions. (Bhavsar & Ganatra, 2012). Contrary to its name, which might suggest simplicity or lack of sophistication, Naive Bayes classifiers have proven to be remarkably effective and versatile, finding widespread application in various domains, such as text classification, spam filtering, and personalized recommendation systems. By harnessing the power of probabilistic reasoning, Naive Bayes algorithms are able to make accurate predictions and classifications, often with impressive results, despite their relatively straightforward implementation and assumptions. As a result, they remain a popular and reliable choice in many areas of machine learning and data science.

Despite its simplicity and potential drawbacks, Naive Bayes remains a viable option for many classification tasks, especially when dealing with textual data or when computational resources are limited.

## 4. MODELLING PERFORMANCE RESULTS AND REVIEW

This section covers the modelling performance results. Table 2 below includes, accuracy, recall, precision and F1-Score, which is harmonic mean of precision and recall, providing a balanced measure of both.

**Table 2**. Performance results of different algorithms

| Algorithms | Training Recall | Testing Recall | Training Accuracy | Testing Accuracy | Training Precision | Testing Precision | F1 Score (Train) | F1 Score (Test) |
|---|---|---|---|---|---|---|---|---|
| LR | 0.775 | 0.920 | 0.762 | 0.792 | 0.756 | 0.730 | 0.765 | 0.814 |
| BA | 0.990 | 0.880 | 0.988 | 0.832 | 0.985 | 0.800 | 0.987 | 0.838 |
| RF | 0.995 | 0.900 | 0.995 | 0.851 | 0.995 | 0.818 | 0.995 | 0.857 |
| GBM | 0.980 | 0.860 | 0.950 | 0.842 | 0.925 | 0.827 | 0.952 | 0.843 |
| AB | 0.935 | 0.860 | 0.895 | 0.812 | 0.866 | 0.782 | 0.899 | 0.819 |
| XGB | 0.990 | 0.900 | 0.993 | 0.861 | 0.995 | 0.833 | 0.992 | 0.865 |
| DT | 0.990 | 0.800 | 0.995 | 0.832 | 1.000 | 0.851 | 0.995 | 0.825 |
| SVM | 0.705 | 0.700 | 0.715 | 0.743 | 0.719 | 0.761 | 0.712 | 0.729 |
| KNN | 0.885 | 0.880 | 0.828 | 0.812 | 0.794 | 0.772 | 0.837 | 0.822 |
| NB | 0.760 | 0.880 | 0.682 | 0.762 | 0.658 | 0.710 | 0.705 | 0.786 |

The performance of the various machine learning models was evaluated, and the results are shown in Figure 4. Linear Regression demonstrated a commendable balance between training and testing performance, with slightly superior recall and accuracy during the testing phase. In contrast, Bagging, an ensemble approach, displayed strong recall during the training phase, but its performance declined significantly during the testing phase, suggesting potential overfitting, as evidenced by a reduction in accuracy. The Random Forest algorithm, on the other hand, exhibited consistent performance across both training and testing datasets, achieving high recall, precision, and accuracy, which indicates robustness in its predictive capabilities. GBM achieved excellent recall and precision in both phases, coupled with strong accuracy metrics, signifying effective generalization capabilities. However, AdaBoost, while exhibiting high recall during training, demonstrated a decline in testing recall and a reduction in testing accuracy, indicating the presence of overfitting tendencies. Similarly, the XGBoost algorithm attained perfect recall in the training phase, but its testing metrics, particularly accuracy, declined significantly, suggesting potential overfitting.
The Decision Tree model showcased robust training performance, but its testing metrics revealed a significant drop in both recall and accuracy, indicative of overfitting tendencies. The SVM algorithm achieved perfect recall, but at the expense of lower accuracy, suggesting an excessive focus on correctly predicting one class while compromising overall model performance. In contrast, the KNN model demonstrated moderate performance, achieving balanced recall and precision, with a slight decrease in accuracy during the testing phase. Lastly, Naive Bayes, despite exhibiting lower recall during training, demonstrated strong performance in testing, concerning both recall and precision, indicating effective generalization despite its inherent simplicity.

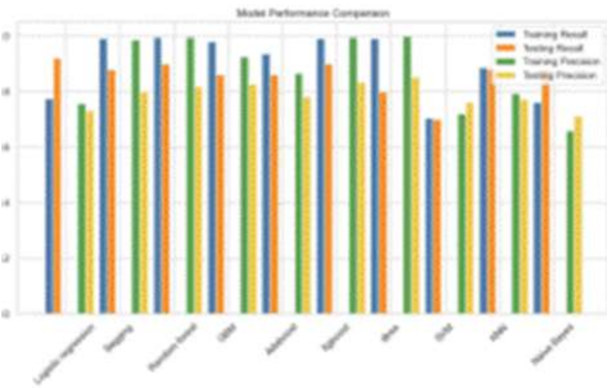

**Figure 4.** Comparison of Models Performance

Overall, the Random Forest and GBM models exhibited the most balanced and high performance across all evaluated metrics, demonstrating their robustness and generalization capabilities. In contrast, Bagging and XGBoost showed signs of overfitting, which raises concerns about their ability to generalize well to new, unseen data. Additionally, the SVM model achieved perfect recall, but at the cost of sacrificing accuracy, indicating a potential imbalance in its class predictions. On a positive note, the Linear Regression model demonstrated a commendable balance between training and testing performance, exhibiting slightly superior recall and accuracy during the testing phase, which suggests its potential for reliable predictions.

### 4.1 Best Performance: Random Forest Model Performance

The Random Forest model, trained with optimized hyperparameters, demonstrated exceptional performance on both the training and testing datasets. With an accuracy of 0.988, the model correctly classified approximately 98.8% of the training samples, showcasing its ability to accurately represent the training data. Notably, the recall and precision scores were identical at 0.988, indicating that the model effectively balanced sensitivity and specificity, and successfully learned to distinguish between the classes with high accuracy. When evaluated on the testing dataset, the model maintained a high level of performance, achieving an accuracy of 0.950, and recall and precision scores of 0.951 and 0.955, respectively. These results suggest that the model generalized well to unseen data, although a slight decrease in performance compared to the training results was observed, as expected when evaluating a model on new, unseen data.

### 4.2 Hyperparameter Configuration

The Random Forest model was optimized using the training parameters shown in Table 3. These parameters are used to control the complexity and behaviour of the Random Forest model, balancing accuracy and computational efficiency.

**Table 3**. Training parameters to optimize performance

| SN | Attribute Name | Value | Description |
|---|---|---|---|
| 1 | n_estimators | 400 | Specifies the decision trees in the forest, which can improve accuracy but increase computational time |
| 2 | min_samples_split | 2 | Samples required to split an internal node, with lower values potentially leading to overfitting |
| 3 | min_samples_leaf | 2 | Samples required at a leaf node, with lower values potentially leading to overfitting |
| 4 | max_leaf_nodes | None | This parameter defines the maximum number of leaf nodes in the tree. If none, no limit imposed |
| 5 | max_features | Auto | This indicates the maximum number of features to consider at each split. 'Auto' means that the square root of the total number of features will be utilized |
| 6 | max_depth | 10 | Represent the depth, with higher values potentially leading to overfitting |

| | | | |
|---|---|---|---|
| 7 | criterion | Gini | Measures the quality of a split, where 'gini' refers to Gini impurity, a measure of how well the classes is separated |
| 8 | bootstrap | False | Suggests no bootstrap sampling is used. |
| 9 | random_state | 1 | Used to shuffle the data and ensure reproducibility of results |

### 4.2.1 Principal Component Analysis (PCA)

Out of the entire dataset, only nine instances were predicted incorrectly, with five of those errors occurring within the testing set utilizing PCA as shown in Figure 5. These errors were primarily associated with accumulated results that were very close to one another, thereby complicating the model's predictive capability. The proximity of these values likely contributed to the model's difficulty in distinguishing between the classes, highlighting a challenge in accurately predicting outcomes in closely related data points.

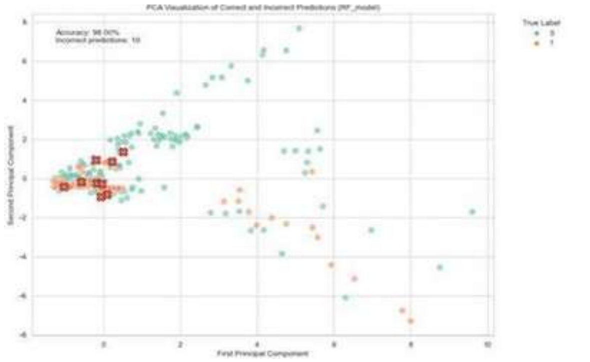

**Figure 5.** PCA visualization of prediction

### 4.2.2 Heatmap Analysis

This section presents an analysis of the output from the heatmap in conjunction with the feature importance derived from the best-performing model, the Random Forest classifier.

**Figure 6.** Heatmap – feature importance

The heatmap (Figure 6) analysis reveals that the strongest correlations with the label are predominantly negative, particularly with the variables Solids, Iron, and Silicon. This observation suggests that higher concentrations of these components are associated with lower label outcomes, potentially indicating detrimental effects on the label's measurement or the overall process outcome. The absence of significant positive correlations implies a generally inverse relationship between many variables and the label, further emphasizing the need for careful monitoring of these components.

### 4.2.3 Feature Importance - Random Forest Model

Figure 7 presents an overview of the prediction performance. Notably, the model successfully avoided misclassifying any good equipment as bad. However, it did exhibit a false positive rate, where approximately 5% of bad equipment were incorrectly predicted as good, warranting further attention. It is likely that enhancing data quality, use of ensemble methods, and optimizing hyperparameters through better tuning will lead to improved model performance and increased prediction accuracy.

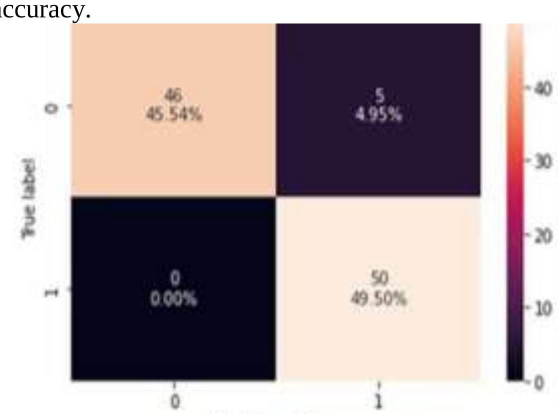


**Figure 7.** Predicted vs true labels

In addition to the heatmap analysis, the feature importance metrics generated by the Random Forest model provide further insight into the relationships between the input features and the label. The importance scores highlight the influence of various features on the model's predictions, allowing for a comprehensive understanding of which variables are most impactful.

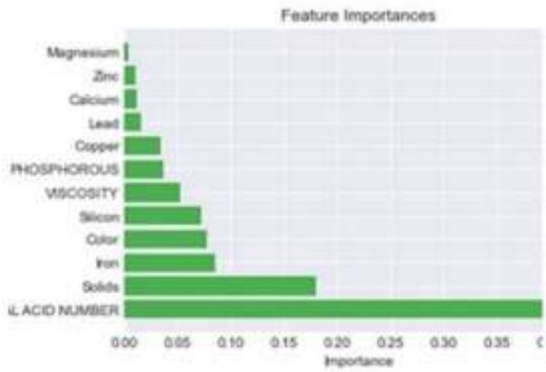


**Figure 8.** Feature importance

The following are key takeaways from the feature importance (Figure 8) analysis:

Solids and total acid number: The model assigns high importance to these features, consistent with the heatmap findings, which suggests its detrimental impact on the label outcome. Iron: Similar to Solids, Iron also exhibits high feature importance, reinforcing the negative correlation identified in the heatmap. Silicon and color: These variables also show significant importance, corroborating the heatmap results and indicating its potential adverse effects on the label. Other Variables: The remaining features exhibit comparatively lower importance scores, which may reflect weaker or less direct influences on the label outcome.

In summary, both the heatmap and the Random Forest feature importance analysis converge on the conclusion that certain variables, particularly Solids, Iron, color and Silicon, have a strong negative correlation with the label. These insights can inform further investigations and operational decisions aimed at optimizing machine health based on oil analyses.

#### 4.2.4 Equipment condition prediction tool

Utilizing all the learnings, a virtual tool has been developed to predict machine conditions using oil features, allowing maintenance personnel to make informed decisions and take proactive measures to ensure optimal equipment performance. The tool provides real-time data analysis, enabling maintenance teams to:

1. Make data-driven decisions about equipment maintenance and repair
2. Schedule proactive maintenance to reduce downtime and increase equipment availability
3. Identify potential problems early to improve equipment reliability and extend lifespan
4. Optimize workflows, prioritize tasks, and allocate resources more effectively

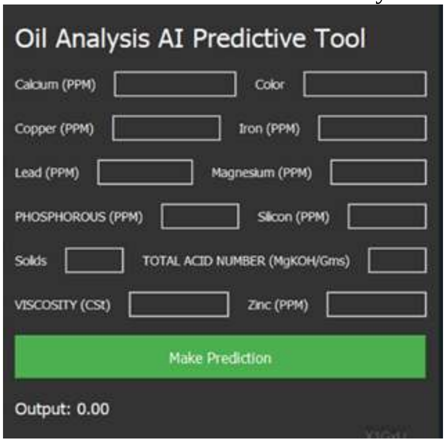


**Figure 9.** Equipment Condition Prediction Tool

To further enhance the tool's capabilities, future development plans include:

1. Integrating with oil condition online systems for real-time monitoring and analysis
2. Providing tool information in the control room for efficient action planning and execution
3. Continuously updating and refining the tool's algorithms and models using advanced analytics and machine learning

By pursuing these opportunities, the virtual tool can become an even more effective predictive maintenance system, driving business value and improving equipment reliability (Figure 9).

## 5. CONCLUSION

In this paper, we utilized ten algorithms to compare their performance in predicting machine health, leveraging the integration of AI and ML in oil analysis. This innovative approach has revolutionized the way we predict the health of rotating equipment, enabling organizations to gain deeper insights into machinery condition and make timely interventions. By harnessing advanced algorithms, companies can reduce the risk of equipment failure, enhance operational efficiency, and minimize downtime and associated costs. Notably, our study achieved a remarkable accuracy of 0.988 on the testing dataset using the Random Forest model, demonstrating its effectiveness in learning complex patterns from oil analysis data and differentiating between varying states of equipment health. This high level of precision is crucial in industries where equipment reliability is paramount, as it fosters confidence in decision-making and aids in resource allocation. The adoption of AI not only streamlines the predictive maintenance process but also enables a shift from reactive to proactive asset management, which is vital for optimizing performance and maintaining a competitive edge in the dynamic energy sector.

In conclusion, the application of AI and ML in oil analysis is essential for predictive maintenance of rotating equipment. The impressive performance of our model highlights the potential of these technologies to transform traditional practices and deliver significant operational benefits. As industries increasingly adopt AI, the prospects for improving equipment health management will continue to expand, paving the way for smarter, more efficient operations.

**Acknowledgement:**
Authors are grateful to Equipment Consulting Services and CS Technical Support Departments management for providing us the necessary, tools, resources and facilities to conduct this research.

## References

Abbas, M., & Metwalli, S. (2025). Multi-Objective Optimization and Supervised Machine Learning for Synthesis of Rolling Element Bearing and Lubricant Selection. *Journal of Mechanical Design*, 147(12), 123501.

Arumsari, N., Pamuji, F. A., & Puspita, B. D. (2021, August). Application of Machine Learning Method to Predict Reliability in Lubricating Oil System Components. In IOP Conference Series: Materials Science and Engineering (Vol. 1175, No. 1, p. 012012). IOP Publishing.

Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Computer Science*, 89, 117-123.

Bhavsar, H., & Ganatra, A. (2012). A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering* (IJSCE), 2(4), 2231-2307.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Burges, C. J. (1998). A tutorial on support vector machine for pattern recognition. *Data mining and knowledge discovery*, 2(2), 955-974.

Choudhary, A., Goyal, D., & Letha, S. S. (2020). Infrared thermography-based fault diagnosis of induction motor bearings using machine learning. *IEEE Sensors Journal*, 21(2), 1727-1734.

Chowdhury, S., & Schoen, M. P. (2020, October). Research paper classification using supervised machine learning techniques. In 2020 intermountain engineering, technology and computing (IETC) (pp. 1-6). IEEE.

Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In icml (Vol. 96, pp. 148-156).

Ismail, R. N. R., Azmi, W. H., & Zawawi, N. N. M. (2024). A Tribological Analysis of PAO-Based Hybrid SiO2-TiO2 Nanolubricants. In Journal of Physics: Conference Series (Vol. 2688, No. 1, p. 012020). IOP Publishing.

Keartland, S., & Van Zyl, T. L. (2020, January). Automating predictive maintenance using oil analysis and machine learning. In 2020 International SAUPEC/RobMech/PRASA Conference (pp. 1-6). IEEE.

Kornaeva, E. P., Kornaev, A. V., Kazakov, Y. N., & Polyakov, R. N. (2020). Application of artificial neural networks to diagnostics of fluid-film bearing lubrication. In IOP Conference Series: Materials Science and Engineering (Vol. 734, No. 1, p. 012154). IOP Publishing.

Lawal, S. A., Choudhury, I. A., & Nukman, Y. (2013). A critical assessment of lubrication techniques in machining processes: a case for minimum quantity lubrication using vegetable oil-based lubricant. *Journal of Cleaner Production*, 41, 210-221.

Malaguti, R., Lourenço, N., & Silva, C. (2021, September). A Well Lubricated Machine: A Data Driven Model for Lubricant Oil Conditions. In EPIA Conference on Artificial Intelligence (pp. 549-560). Springer, Cham.

Moder, J., Bergmann, P., & Grün, F. (2018). Lubrication regime classification of hydrodynamic journal bearings by machine learning using torque data. *Lubricants*, 6(4), 108.

Mohammad Nazari, N., & Muhammad, M. (2021, August). Developing a Lubrication Oil Age Prediction Model. In World Congress on Engineering Asset Management (pp. 411-421). Springer, Cham.

Naidu, G., Zuva, T., & Sibanda, E. M. (2023, April). A review of evaluation metrics in machine learning algorithms. In Computer science on-line conference (pp. 15-25). Cham: Springer International Publishing.

Orrù, P. F., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R., & Arena, S. (2020). Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. *Sustainability*, 12(11), 4776.

Perez, R. X. (2025). Introduction to Machinery Lubrication. Process Machinery Handbook: For Field Personnel, *Decision Makers, and Students*, 445-466.

Pichler, J., Maria Eder, R., Besser, C., Pisarova, L., Dörr, N., Marchetti-Deschmann, M., & Frauscher, M. (2023). A comprehensive review of sustainable approaches for synthetic lubricant components. *Green Chemistry Letters and Reviews*, 16(1). DOI:10.1080/17518253.2023.2185547

Pramesti, W., Damayanti, I., & Asfani, D. A. (2016, July). Stator fault identification analysis in induction motor using multinomial logistic regression. In 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 439-442). IEEE.

Raviya, K. H., & Gajjar, B. (2013). Performance Evaluation of different data mining classification algorithm using WEKA. Indian Journal of Research, 2(1), 19-21.

Rossopoulos, G. N., & Papadopoulos, C. I. (2021). A journal bearing performance prediction method utilizing a machine learning technique. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 13506501211055710.

Shahzad, Umair. (2022). A comparative analysis of artificial neural network and support vector machine for online transient stability prediction considering uncertainties. *Australian Journal of Electrical and Electronics Engineering.* 19. 1-16. 10.1080/1448837X.2021.2022999.

Singh, A., Verma, N., Mamatha, T. G., Kumar, A., Singh, S., & Kumar, K. (2021). Properties, functions and applications of commonly used lubricant additives: A review. *Materials Today: Proceedings*, 44, 5018-5022.

Soleh, M., Riyadi, A., & Asnel, A. P. (2019, October). Development of Predictive Maintenance Methodology Utilizing Machine Learning Technology to Support Plant Health Management. In 2019 International Conference on Technologies and Policies in Electric Power & Energy (pp. 1-6). IEEE.

Trizoglou, P., Liu, X., & Lin, Z. (2021). Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines. *Renewable Energy*, 179, 945-962.

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1-16.

Uppar, R., Dinesha, P. & Kumar, S. A critical review on vegetable oil-based bio-lubricants: preparation, characterization, and challenges. Environ Dev Sustain 25, 9011–9046 (2023). DOI: 10.1007/s10668-022-02669-w

Whitby, R. D. (2021). Lubricant analysis and condition monitoring. CRC Press.

Yuan, W., Chin, K. S., Hua, M., Dong, G., & Wang, C. (2016). Shape classification of wear particles by image boundary analysis using machine learning algorithms. *Mechanical Systems and Signal Processing*, 72, 346-358

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access*, 6, 21020-21031.

**Qadeer Ahmed**
Equipment Consulting Services Department, Saudi Aramco
Dhahran, Saudi Arabia
qahmed50@gmail.com
qadeer.ahmed@aramco.com
**ORCID:** 0000-0002-9744-9896

**Hasanur J. Molla**
Equipment Consulting Services Department, Saudi Aramco
Dhahran, Saudi Arabia
hasanur.molla@aramco.com
**ORCID:** 0009-0006-2653-6390

**Ziad M. Aloqifi**
CS Technical Support,
Saudi Aramco
Dhahran, Saudi Arabia
ziad.oqyfy@aramco.com
**ORCID:** 0009-0001-6810-1006

**Dahham M. Anazi**
Equipment Consulting Services Department, Saudi Aramco
Dhahran, Saudi Arabia
dahham.anazi@aramco.com
**ORCID:** 0009-0008-2349-4273

**Saad H. Dossary**
Equipment Consulting Services Department, Saudi Aramco
Dhahran, Saudi Arabia
saad.dossary.38@aramco.com
**ORCID:** 0009-0002-8215-3467