Received 28 March 2018
Revised 23 December 2018
27 April 2020
Accepted 13 May 2020

A data mining approach for lubricant-based fault diagnosis

James Wakiru and Liliane Pintelon
Centre for Industrial Management/Traffic and Infrastructure, KU Leuven, Leuven, Belgium
Peter Muchiri
Dedan Kimathi University of Technology, School of Engineering, Nyeri, Kenya, and
Peter Chemweno
Department of Design, Production and Management, University of Twente, Enschede, The Netherlands

Abstract

Purpose – The purpose of this paper is to develop a maintenance decision support system (DSS) framework using in-service lubricant data for fault diagnosis. The DSS reveals embedded patterns in the data (knowledge discovery) and automatically quantifies the influence of lubricant parameters on the unhealthy state of the machine using alternative classifiers. The classifiers are compared for robustness from which decision-makers select an appropriate classifier given a specific lubricant data set.

Design/methodology/approach – The DSS embeds a framework integrating cluster and principal component analysis, for feature extraction, and eight classifiers among them extreme gradient boosting (XGB), random forest (RF), decision trees (DT) and logistic regression (LR). A qualitative and quantitative criterion is developed in conjunction with practitioners for comparing the classifier models.

Findings – The results show the importance of embedded knowledge, explored via a knowledge discovery approach. Moreover, the efficacy of the embedded knowledge on maintenance DSS is emphasized. Importantly, the proposed framework is demonstrated as plausible for decision support due to its high accuracy and consideration of practitioners needs.

Practical implications – The proposed framework will potentially assist maintenance managers in accurately exploiting lubricant data for maintenance DSS, while offering insights with reduced time and errors.

Originality/value – Advances in lubricant-based intelligent approach for fault diagnosis is seldom utilized in practice, however, may be incorporated in the information management systems offering high predictive accuracy. The classification models' comparison approach, will inevitably assist the industry in selecting amongst divergent models' for DSS.

Keywords Lubricant condition monitoring, Maintenance decision support, Classification, Oil analysis, Data mining, Machine health

Paper type Research paper

1. Introduction

1.1 Background

Industrial set up that use rotating, and reciprocating equipment such as engines, gearboxes and compressors have adopted condition monitoring techniques under the condition-based maintenance (CBM) strategy including vibrational analysis, lubricant condition monitoring (LCM) and thermography (Wakiru *et al.*, 2019). Condition monitoring offers significant benefits such as reduction of maintenance costs by eliminating potential failures and traditionally, often required equipment shutdown for inspection during operations (Zhu *et al.*, 2017). LCM involves in-service oil analysis and interpretation of the results which assist in maintenance decision by indicating the condition and health of the oil and the machine being lubricated. The performance of a lubricant is primarily influenced by its deterioration level and further altered by the operational conditions of the equipment. Many organizations collect hundreds of samples, analysed and manually classified by in-house analysts to indicate whether the condition of the machine is either satisfactory or not. If the sample is

Journal of Quality in Maintenance Engineering
Vol. 27 No. 2, 2021
pp. 264-291

1355-2511
DOI 10.1108/JQME-03-2018-0027

within acceptable specifications, the equipment can remain operable using the same lubricant. In contrast, action needs to be taken, either to improve the equipment condition, or renew the state of the lubricant or equipment.

The process of manually classifying and evaluating the oil samples experiences several challenges. Firstly, the manual process is time-consuming due to a considerable number of parameters which needs to be evaluated; therefore, it is prone to human errors and delays. Secondly, the analysts frequently have limited knowledge of the parametric behaviour of the lubricant and operational knowledge of the equipment. Furthermore, their interpretation may be limited to their experience, ideas and the extent to which they can suggest changes to the maintenance regimes. Hence, deriving useful decisions from knowledge discovery based on manual analysis and interpreting the results of used oil data is instead sub-optimal and not straightforward. Thirdly, besides the analysis considering univariate parametric interpretation, the analyst's interpretation is generally limited to a particular sample, without considering historical data. Multiple parametric interpretations and interactions of the parameters are rarely considered (Wakiru *et al.*, 2017a); hence, ignoring such parametric interactions may yield an invalid interpretation of the parametric behaviour of the lubricant data (Kumar and Kumar, 2016). Lastly, errors in the sampling procedure may influence the results of the used oil analysis, and if no benchmarking is done, substantial information may go unnoticed or erroneously dismissed. For instance, a high value depicted by a parameter like silicon may be attributed to dirt ingression to the sample due to wrong sampling procedures. In this study, we use data mining methods to explore significant correlation and patterns in used oil analysis data for maintenance decision support and more substantially build classification models. As a result, the maintenance dataset can be comprehended more intuitively, which will assist in predicting how new events like oil deterioration will act based on the classification metric selected. In literature, selection of the most appropriate model to use while building the expert system in the field of lubricants has been made using both quantitative and qualitative measures by very few authors and studies, as reviewed in Section 2.

Owing to these limitations, this research seeks to derive decision support enhancement from the analysis of used oil analysis data. The enhancement is achieved by developing an

LCM	Lubricant condition monitoring	RF	Random forest	OEM	Original equipment manufacturer
CBM	Condition-based maintenance	KNN	K-nearest neighbours	PC	Principal component
DSS	Decision support system	HFO	Heavy fuel oil	AU	Approximately Unbiased
UOA	Used oil analysis	CA	Cluster analysis	MLP	Multilayer Perceptron
DT	Decision tree	TBN	Total base number	ERT	Extremely randomized trees
LR	Logistics regression	RUL	Remaining useful life	BN	Bayes network
SVM	Support vector machine	AUC	Area under curve	ADA	AdaBoost
NN	Neural network	ACC	Accuracy	BP	Bootstrap probability
XGB	Extreme gradient boosting	LB	Logit Boost	PLS	Partial least squares
PCA	Principal component analysis	LASSO	Least absolute shrinkage and selection operator	SGBRT	Stochastic gradient boosted regression trees
RDA	Regularized discriminant analysis	PNN	Probabilistic neural network	SIMCA	Soft independent modelling class analogy

Table 1.
List of abbreviations

integrated framework to analyse and extract knowledge from identified patterns in the data, to enable appropriate maintenance intervention. The identified patterns are then used as input while advancing the predictive model for the system, where quantifying the influence of lubricant parameters to sample classification is explored. The decision support system is expected to offer a framework that, firstly, exposes meaningful patterns embedded in the used oil analysis (UOA) data and their inferences towards maintenance decision support. Secondly, identify critical lubrication parameters to consider in the fault-diagnosis of the equipment. Thirdly, based on knowledge discovery of the lubricant data, seek to establish the general mechanical fault causes, and lastly among others, assist in the selection of the appropriate classification model to employ under the plant's operational context and practitioner's preferences. Table 1 provides the list of abbreviations used in this study.

The remainder of the paper is organized as follows: In Section 2, we present a state-of-the-art review of the problem, data mining techniques and classifiers. In Section 3, we give details of the methodology the paper will follow. Next, we illustrate the results and provide a discussion in Section 4. In Section 5, we provide managerial implications of the results. Finally, in Section 6, we conclude the work with directions for subsequent research.

2. Review of related literature

2.1 Lubricant and functions of lubrication

The primary functions of lubricants applied in machines include removing contaminants, reducing wear and friction, protection against corrosion and rust, and cleaning the system being lubricated (Wakiru *et al.*, 2019). During the use in equipment or system, the lubricant undergoes degradation that affects its quality and performance, either losing its essential properties such as viscosity through intrinsic (e.g. shear) or extrinsic (thermal degradation, contamination) processes, which may lead to severe equipment failure. Therefore, oil analysis is carried out to monitor the condition of the lubricant to prescribe interventions that abate equipment failure. A comprehensive review of lubricating oil conditioning sensors and oil parameters can be found here (Zhu *et al.*, 2017). The analysis of the UOA results in data generated from the equipment through predominantly statistical methods, assists in maintenance decision support by exposing embedded knowledge in the data (Wakiru *et al.*, 2019). Utilizing traditional statistical methods, for instance, trend analysis, while pursuing maintenance decision support from this type of maintenance data, cannot respond to the demands for the analysis of the extensive data generated. These methods are unable to track historical patterns, explain the complex phenomenon and ultimately offer prediction (Raza *et al.*, 2010). Hence, the application of intelligent data mining techniques, which possesses the potential of discovering hidden and useful information from data. The primary methods discussed in this context include correlation analysis, principal component analysis and cluster analysis that offer data visualisation and dimension reduction in the next Section 2.2. The classification models include DT, RF, SVM, XGB, LR, ERT, NN and KNN are discussed in the subsequent Section 2.3.

2.2 Data visualizations and dimension reductions

Correlation analysis is a statistical technique that ascertains the strength of association between two discrete variables which can signify a predictive association applicable in practice. Pearson correlation method is used when both variables follow a normal distribution, while the Spearman correlation method is used where one or both variables are not following a normal distribution. Moreover, Spearman's correlation coefficient possesses merits over Pearson's method such as, being more robust to outliers and can equally be used when one or both variables are ordinal (Carla Bittencourt Reis *et al.*, 2009).

Multiple correlations represent another method based on the Pearson correlation, where a given variable can be predicted applying a linear function of a set of other variables and used to assess dependent variables in multiple regression analysis. Pearson correlation was applied in the testing of lubricant wear data by Adnani *et al.* (2013), while multiple correlations were used along with other methods while investigating the dependency of total base number (TBN) and selected wear metals (Vališ *et al.*, 2016a, b). From studies reviewed, the Spearman method was seldom used while the studies did not constitute the basis on which the Pearson correlation method was selected. Notwithstanding, some authors recognize the limitation of correlation analysis, in that it does not go further enough in looking at data in a multivariate way. This aspect intrinsically will not highlight interactions of the parameters; hence, the proposition of using cluster analysis by various authors discussed next (Aggarwal, 2015).

Cluster analysis (CA) represents a statistical method that classifies several variables (clusters) according to their similarities. Some studies utilizing clustering technique in the lubrication field include (Vališ *et al.*, 2016a, b). While CA has gained widespread usage, it has some shortcomings. For instance, it is a heuristic technique. Thus, clusters can be developed still where there may be no similarity patterns in the data (Discenzo *et al.*, 2006). Moreover, the results' uncertainty caused by sampling error has been unevaluated, hence the use of methods to assess the uncertainty has been fronted like the use of bootstrap analysis (Suzuki and Shimodaira, 2006). Clustering can sometimes constitute groups based on noise in the data originating from sampling error or even sample procedure error hence establishing reliability challenges. Hence, due to these flaws, several authors (e.g. Balabin and Safieva, 2011; Wang and Hussin, 2009) propose PCA which is used for correlation as well as dimensional reduction and variable selection as discussed next.

Principal component analysis (PCA) is a technique that constitutes new variables uncorrelated with each other called principal components (PCs) that are linear combinations of the primary variables. Therefore, it reduces an extensive set of variables to a narrow set that nonetheless contains most of the vital information in the original extensive set, which remains a challenge while using cluster and correlation analysis. PCA has been used for dimension reduction in diverse areas, for instance; it was used in variables reduction for CBM model Lin *et al.* (2006) and in oil analysis prediction models (Jun *et al.*, 2008). Despite the use, PCs and variables interpretability present a challenge conventionally hence rotation methods like Orthogonal and oblique rotation are introduced (Tabachnick and Fidell, 2007). PCA rotation has been used to analyse further initial PCA results to expose the pattern of loadings in a more simplistic, conspicuous way, and enable robust understanding and interpretation.

Owing to the flaws mentioned above of each method, some authors propose an integrated framework for exploring hidden patterns in the maintenance data. Customarily, variables derived from the exploration are used as input to predictive models that may assist in the maintenance decision making process. Data mining incorporates several generalized tasks, like association rule mining, regression, clustering, classification and detection of anomalies. One notable predictive model is a classification which classifies data according to some predictor variables where several models may be feasibly applied in practice, as discussed next.

2.3 Classification models

Classification is a machine learning technique used to predict a category or class membership for data by allocating items in a collection to target class/category for each case in the data, hence a supervised learning approach. A state of the art review of classification algorithms is advanced (Zhang *et al.*, 2017). Several comparative studies are advanced, such as in the application of NN and statistical techniques Paliwal and Kumar (2009) and NN, LR and SVM (Raza *et al.*, 2010).

We group the classification algorithms investigated in this work into two groups related explicitly to LCM context. The first group includes logistics regression (LR), support vector machine (SVM), random forest (RF) and neural network (NN) which have been used in LCM as default classifiers. The second group includes established algorithms seldom used in the lubrication field like K-nearest neighbours (KNN), decision trees (DTs), extreme gradient boosting (XGB) and extremely randomized trees (ERT). In this section, we review the mentioned binary classification techniques which the author views as accessible to maintenance practitioners.

Support vector machine (SVM) refers to a classifier that employs linear combinations of distinct attributes of the data to make classification decisions. It can deal with a substantial number of features as it attempts to arrive at a globally optimized solution, hence, avoiding overfitting (Moosavian *et al.*, 2014). It is claimed to perform well on small training samples that may as well be non-linear in structure. SVM can handle large feature space and offer a right generalization property to classification, making feature selection less critical. SVM has been used in the LCM field by (Chowdhury *et al.*, 2016; Eitrich and Lang, 2006; Phillips *et al.*, 2015), and in other condition monitoring fields as discussed by Moosavian *et al.* (2014).

NN is an information processing model patterned on the human brain function and structure by learning knowledge from data with known inputs and outputs. NN can be applied as supervised learning requiring the external input of the previous knowledge about the target and unsupervised neural network, which is self-learning as corroborated by Jardine *et al.* (2006). NN has been used in the oil analysis to classify the sample condition (Phillips *et al.*, 2015). The primary limitations of NN represent the lack of explanation of the model parameters, the complexity of the training process and requirement of important training data (Raza and Liyanage, 2009; Phillips *et al.*, 2015).

LR represents a model that illustrates the nature of the relationship between the mean response and one or more predictor variables which are categorical. LR has been used in studies incorporating used oil analysis (UOA) (Caesarendra *et al.*, 2010). All the studies the researcher found to have manipulated selected UOA variables principally based on one or a maximum of two categories of lubricant analysis (among wear metals, additives, dilution or physiochemical properties). In contrast, the basis and methodology of selection of the explanatory or dependent variables in the LR model are seldom discussed.

RFs are a class of ensemble algorithms that propagate several trees as the base estimate and aggregate them to predict and solve classification or regression problems (Scornet, 2016). RF is like DT, except that a collection of un-pruned decision trees is combined to provide better classification accuracy with reduced over-fitting by averaging the result, unlike DT. Other features of RF embody determination of variable importance measures and offering high predictions even with training noisy data (Satishkumar and Sugumaran, 2016). RF has been used in classification models in the UOA field (De Rivas *et al.*, 2017).

DTs utilize a tree-like structure to divide a data set into branches and leaves while developing the decision tree incrementally (Nasridinov *et al.*, 2013). DT is computationally insensitive to missing data, uses predictor variables inapplicable to data outliers. The method is adaptable to various types of datasets, while the algorithm utilizes a recursive partitioning approach. Wakiru *et al.*, (2018) used DT to classify oil samples while DT has likewise been used in other predictive maintenance applications such as classifying automotive fault conditions (Shafi *et al.*, 2018).

XGB is a recently advanced gradient tree boosting an accurate and scalable ensemble algorithm due to its inherent characteristics that support parallel processing, regularization and early stopping (Xia *et al.*, 2017). XGB classified as “black box” model, has been applied in other condition monitoring field such as predicting the RUL of a wind turbine gearbox (Lu *et al.*, 2017). XGB has been applied significantly in other sectors such as manufacturing quality control (e.g. Flath and Stein, 2018) health (e.g. Semiz *et al.*, 2018), agro-ecosystem (e.g. Zhao *et al.*, 2018) and banking (e.g. Carmona *et al.*, 2018). The authors did not come across an article applying XGB in UOA field.

KNN is a non-parametric algorithm that generates predictions for a sample by finding the k nearest samples and assigning the most represented class among them using a distance function. It is seldom utilized in LCM but was used in condition monitoring to classify machine condition using vibration signals (Moosavian *et al.*, 2014; Safizadeh and Latifi, 2014; Yu, 2011).

ERT, just like RFs, are an ensemble model. In addition to sampling features at each stage of splitting the tree, it also samples the random threshold at which to complete the splits. The additional randomness may enhance the ability of the model to generalize (compared to RF) and may yield better results. ERT was used along with other classifiers in power curve modelling while monitoring the performance of offshore wind turbines, where both wind and the turbine blade characteristics were employed (Janssens *et al.*, 2016). Similar to XGB and KNN, authors did not get an application of ERT in the UOA field.

The methods discussed above likewise include ensemble techniques which combine results of multiple classifiers increasing prediction accuracy using either bagging and boosting techniques (Aggarwal, 2015). Among the bagging is RF while the boosting are XGB and ERT used in this study. Other classification algorithms include deep learning, Sparse representation-based classification and Extreme Learning Machine which can be classified as relatively new, while Gradient Boosting Decision Trees, Naïve Bayes and AdaBoost as established classifiers (Zhang *et al.*, 2017). However, we do not employ these algorithms due to their relative unavailability as standard toolboxes, which limits access to the practitioners. A comprehensive review of the classification algorithms such as DT, LR, NN, SVM and RF as used in lubrication condition monitoring is presented by (Wakiru *et al.*, 2019).

The results in this section indicate that several algorithms have significantly been used in the UOA field such as DT, ANN and LR. At the same time, the author could not find an application utilising KNN, XGB and ERT algorithms in the UOA field. However, a considerable amount of literature has been published that employs binary classification of data derived from other condition monitoring techniques as illustrated in Table 2, where the application on vibration analysis data is illustrated as significantly utilised.

While much of the current literature on classification models, pay attention to the model performance, they disregard critical aspects such as comparing various algorithms and considering user preferences. However, the classification models are challenged with overestimation attributed to the learning and prediction derived from the data. This challenge is mitigated by validation of the model as discussed in Section 2.4, while Section 2.5 moves on to review studies that have undertaken model comparison.

2.4 Model validation

To address model overestimation, internal validation processes like split validation, Cross-validation and Bootstrapping method are used to evaluate the model performance using part of the data used to train the model. Using external validation, the prediction model's performance is evaluated utilizing data that was unused to train the model.

Article reference	Classifier	Field	Data used
Cerrada *et al.* (2016), Patel and Giri (2016)	RF	Machine condition	Vibration signals
Muruganatham *et al.* (2013), Li *et al.* (2014)	ANN	Machine condition	Vibration analysis
Li *et al.* (2014)	SVM	Machine condition	Vibration analysis
Lazakis *et al.* (2018)	ANN	Machine condition	Temperature analysis
Campora *et al.* (2018)	ANN	Machine condition	Functional signals

Table 2.
Binary classification studies under other condition monitoring techniques

2.5 Model comparisons

We found a scarcity of literature that attempts to compare the effectiveness of different classification techniques considered in this study, particularly concerning classifying UOA data (LCM). From our review of various studies, we found various classification models applied to the LCM field as discussed in Section 2.3 but few studies specifically under LCM comparing the separate models used in binary classification as represented in Table 3.

However, it is important to note that a sizeable number of studies was found that have compared the different classifiers while utilising other condition monitoring techniques data, as seen in Table 4.

Reviewing Tables 3 and 4, SVM and ANN seem to be applied significantly compared to the other classifiers in both LCM and other condition monitoring based studies. The studies presented so far provide evidence that both classification and comparison of the algorithms in the UOA field are not as extensively done compared to other condition monitoring techniques. Furthermore, research on the subject has been primarily restricted to limited comparisons of classifier's accuracy and predictive power, as seen in Table 3. Moreover, some aspects such as the model's goodness of fit, efforts put in building the model and model's predictive power, which play a significant role in model robustness and applicability in maintenance decision support, have been uninvestigated in the various comparative studies (e.g. Balabin and Safieva, 2011; Phillips *et al.*, 2015).

2.6 Insights from a review of related literature and study motivation

Due to the various strengths and weaknesses of the various data mining techniques, there is a trend towards the use of an integrated framework as indicated in Section 2.2 and corroborated by (Wakiru *et al.*, 2019). An integrated framework enhances synergy from the various individual techniques as well as seek the concurrence of various embedded data/information patterns by developing a wholesome picture of the embedded patterns.

Classification technique is an essential aspect utilized in LCM while developing a predictive model. Two significant themes emerge from the studies discussed so far: Firstly, relatively few classification models have been applied to oil analysis data, while limited studies have compared the various classification techniques, where only three studies were found demonstrating a comparison of the classifiers while applying oil analysis data. Additionally, the studies do not illustrate the basis of comparing the algorithms that consider both performance and user preferences. These findings, while preliminary, suggest there is a need to expand the classification models' scope in application and comparison employing

Table 3.
Binary classification model comparison studies -LCM related

Article reference	Classifiers compared[a]	Parameters	Criteria utilized
De Rivas *et al.* (2017)	SVM and RF	Total acid number	Root mean square (RMSE), Mean absolute error (MAE)
Balabin *et al.* (2011)	SVM, KNN, PNN, MLP, RDA, SIMCA, PLS	Viscosity	Classification error
Phillips *et al.* (2015)	ANN, LR and SVM	Iron, Sodium, Lead, Copper, oxidation	Classification accuracy, Ease of use, Ability to expose how changes of variables influence classification, Consequences of misclassification

Note(s): [a]RMSE, root mean square error; MAE, mean absolute error; PNN, probabilistic neural network; MLP, multilayer perceptron; RDA, regularized discriminant analysis; SIMCA, soft independent modelling class analogy; PLS, partial least squares

Article reference	Classifiers compared	Field	Data used
Raza *et al.*, (2010)	ANN, SVM and LR	Machine condition	Pressure time series signals
Samanta (2004), Yang *et al.* (2005), Pandya *et al.* (2014), Li *et al.* (2014)	ANN and SVM	Machine condition	Vibration signals
Seera *et al.* (2017)	SVM, RF and MLP	Machine condition	Vibration signals
Demetgul (2013)	SVM and DT	Machine condition	Optic and pressure signals
Cai *et al.* (2010)	LR, ANN, SVM and KNN	Machine condition	Normal and fault data
Shafi *et al.* (2018a)	DT, SVM, KNN and RF	Machine condition	Normal and fault data
Pereira *et al.* (2018)	SVM, DT, RF, XGB and KNN	Machine condition	Operational data
Kumar *et al.* (2018)	SVM, ANN, LR, RF, BN, ADA and LB	Machine condition	Remaining useful life
Janssens *et al.* (2016)	RF, ERT, SGBRT and KNN	Machine condition	Operational data

Table 4.
Binary classification model comparison studies – other condition monitoring techniques

used oil analysis data. Secondly, relatively few selected lubricant parameters are used in the classification model, which is not fully representative, lacks a selection basis and retains the potential of generating biased results as also corroborated by (Sharma, 1996). Hence, to ensure vivid representation and consideration of critical parameters, a systematic process of selecting appropriate lubricant parameters to employ in the classification model is needed.

Consequently, there is a need for a multi-model framework integrating separate models to check concurrence, where the output of these models may provide useful decision support regarding the classifiers. Incorporating a balanced qualitative and quantitative criterion for selection would greatly assist practitioners to invest in appropriate techniques. Such as approach is so far unproposed in the literature for analysing used oil data.

The first contribution of this paper is to develop an integrated framework for maintenance data (UOA data) analysis using various data mining techniques. Explicitly, we focus on knowledge discovery using data visualisation and dimension reduction techniques such as correlation, cluster analysis and PCA. The proposed approach explores the embedded, meaningful parametric patterns in the used oil analysis data and evaluates their implications for maintenance decision support.

The second contribution entails developing a systematic classification framework that considers the selection of variables and application of various classifiers (both black box and white box) for used oil analysis samples classification. Most of the present studies in the literature (e.g. Chowdhury *et al.*, 2016) focus on well-established classifiers without considering the recent machine learning advancement concerning classification. Moreover, the studies, as described in Section 2.5, utilise selected parameters as input to the models without indicating the variables selection criteria used. The lack of a systematic criterion for selecting optimal parameters for oil analysis compromises the optimal results because in often cases, the interactive effects of the variables are not considered as discussed above. However, new classifiers have been compared in different condition monitoring sectors like vibration, but their evaluation is limited to their specific domain.

The third contribution of this paper aims to develop a scalable comparison and selection criterion for different classification models applied to UOA data. Despite various researchers (e.g., Phillips *et al.*, 2015; De Rivas *et al.*, 2017) addressing the problem of comparing classifiers using the oil analysis data, there remains a paucity in a criterion that will help the maintenance practitioners select the appropriate classification model(s). The selection of a suitable classification model will incorporate the use of defined qualitative and quantitative criteria developed in consultation with practitioners, which will help the maintenance team select a suitable model based on the type of data they have and their expectations towards enhancing maintenance decision making. By incorporating both quantitative and qualitative criterion, we show a maintenance manager should balance the trade-off between interpretability, accuracy derived by different classifiers and own preferences while selecting an appropriate classifier algorithm as also corroborated by (Wakiru *et al.*, 2019).

This paper aims to propose such an integrated framework to analyse maintenance data (UOA data) using an integrated data mining formalism and to evaluate the extent to which this integrated framework is useful for maintenance decision support using classification models.

3. Methodology

3.1 Case study and lubricated equipment

This study employs a case study approach where, research data was obtained from a thermal power plant that uses heavy fuel oil (HFO) driven medium speed engines, whose speed and a cylinder bore of 750 rpm and 320 mm respectively, to drive generators for power generation. This type of engines are exposed to wear challenges due to the HFO characteristics of high sulphur content (1%) which could subsequently lead to high acidity in the engines, hence corrosive wear on the cylinder liner, piston rings, injector pump, crankshaft and camshaft bearings. Among other factors such as water ingression, aggravates this problem by the formation of carbonaceous deposits within the combustion section (piston ring groove, cooling gallery, injector pump plunger and nozzles) that can lead to engine seize and significant wear of the components. The plant carries out routine UOA on a month-by-month basis as part of its preventive maintenance strategy. It maintains a database of oil analysed and classified by the analysts in an independent laboratory. However, despite routinely undertaking UOA, chronic wear-related failures are easily undetected in advance. At the same time, such plants lack a systematic framework to discover equipment health knowledge embedded in the UOA data. Moreover, while evaluating the LCM data, they seldom utilize historical data which could offer significant insights into the preventive maintenance strategy to be implemented. The methodology, as illustrated in Figure 1, consists of several steps discussed in the subsequent section.

3.2 Data collection and pre-processing

This study utilizes empirical used oil analysis data that covered the period from 2011 to 2016 with 21 lubricant test parameters analysed against recommended standards (see Table 5). The plant employs an off-line sampling strategy, where engine oil samples are drawn monthly and taken for analysis in an independent laboratory. The data pre-processing incorporated several steps: the first step involved integrating the data into one file to enable analysis and inspection because the UOA sample analysis data were in separate files representing individual sample results. With the involvement of expert consultations, the data was cleaned in the second step, where data consistency is harmonized by inspecting the data, expunging errors like duplicated results. During this step, missing values are dealt with, either by disavowing the observations, performing some imputation techniques like replacing the missing value with median or mode for continuous or categorical data, respectively. Lastly, outliers, which represent observations that lie abnormal distance from other values, were validated using boxplots. In this case, the invalid outliers are considered as missing values.

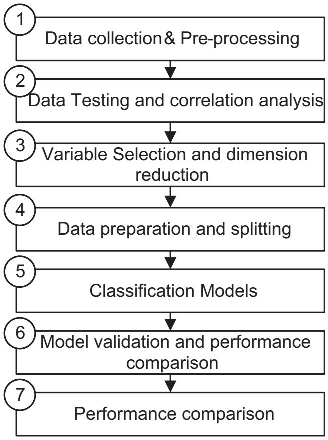

Figure 1. Schematic representation of the methodology

Viscosity 40°C	Calcium	Sodium	Chromium	Tin	Total base number
Flashpoint	Water	Molybdenum	Lead	Pentane insoluble	
Zinc	Carbon	Nickel	Copper	Viscosity 100°C	
Silicon	Iron	Magnesium	Vanadium	Aluminium	

Table 5. Analysed variables in the case LCM program

3.3 Data testing and correlation analyses

In this respect, pre-processed data were tested for normality to confirm the correct correlation analysis method to be used. From research, various studies conclude the Shapiro–Wilk test as a robust normality test for all sample sizes and distribution types, for example (Razali and Wah, 2011).

3.4 Variables selections and feature reductions

It was essential to select essential variables to be applied as an input to the classifier models, due to a substantial number of analysed variables. Hence, a dimension reduction technique was incorporated into the classifier model.

3.4.1 Principal component analysis. PCA was used for establishing patterns in data to explore the similarities and dissimilarities in the data set, using correlation coefficients between the primary variables and the PCs. These patterns indicate the extent to which the original variables are essential in creating new variables. This is equally represented by the factor loading plot where vectors of relatively comparable length and narrow-angle of separation infer strong correlation. The Tabachnick and Fidell criteria were used to select the rotation method to use, which helps in making the PCs more visible and more comfortable to interpret.

3.4.2 Cluster analysis. Concerning analysing the data, we utilize a hierarchical clustering technique using the average linkage method. To deal with the uncertainty, we utilize the pv-clust method embedded in the R software to calculate two variant probability values or p-values that is AU (approximately unbiased) and BP (bootstrap probability) for each cluster using bootstrap resampling techniques.

3.4.3 Feature reduction. RF, PCA, least absolute shrinkage and selection operator (LASSO) and correlation with target approaches were tested for suitability in the subject study based on the model predictive accuracy and predictive power (AUC) to select the appropriate technique to be incorporated in the classifiers.

A brief discussion with the managerial implication is done for the knowledge extracted using both PCA and CA. Admissible variables are selected using the cut-off criteria, where these oil parameters are seen to be significant regarding uncovering the embedded information in the specific power plant UOA data. Feature or dimension reduction will utilize classification algorithms with the ability to automatically select important input features with the guidance of the number of suitable features selected.

3.5 Data scoring and splitting

The data were scored by analysts evaluating each of the selected parameters against a limit or threshold set by the Original Equipment Manufacturers (OEM) and moderated by the power plant. The data represented 1,103 sample observations where 197 (17.9%) were scored as 1 (PASS) and the rest as 0 (FAIL). The observed rate was attributed to the fact that score 0 included samples that were at the borderline and exhibited cautionary advice to the maintenance team. Random split validation on training and testing data was used, where data was split in the ratio 80:20 that is, 80% of the data were used for training the model while 20% of the data was used in testing the model. This split validation approach was viewed to suffice the objective of the study due to ease of modelling in most of the software and graphical user interfaces maintenance managers can access.

3.6 Classification models and hyperparameter optimisation

The scored data for the selected variables were used to build different binary classification models, as explained in the following section. We used R, Rapid miner and Dataiku DSS software.

While making the *RF, LR, XGB and ERT* models, two standard parameters were tuned with validation data. The first parameter is the number of classification trees to be trained to build the ensemble of classification trees, which the range is (100, 400); the other parameter maximum depth of individual trees rage is (1, 10). Common to *LR*, *RF* and *ERT* comprise the minimum samples per leaf which range is (1, 10) and the number of features to sample which range from 1 to the number of features in the dataset (16). Other parameters represent the learning rate for *XGB*, in the range of (0.1, 0.3) and penalty parameter for *LR* (C or regularization term) to the weights whose range is (0.1, 100). The DT model was built by splitting the variables into several branches recursively until the termination and classification are reached. With a complexity parameter (cp) of 0.01, the minimum split was varied from 1 to 10 while the maximum range is (3, 30). When testing *NN,* the number of hidden layer nodes was changed from 1 to the number of features in the dataset (16) while maximum iterations as 200. When testing *SVM***,** we use radial basis function (rbf), polynomial and sigmoid kernels while tuning the penalty parameter (C) and parameter gamma (γ). While reviewing *KNN*, the number of neighbours to check for each sample k ranged $\{1,3,5\}$, while the distance used to search for the neighbours varied between Euclidian and Manhattan. ($p = 1$ or 2 respectively). During the development of the models, grid search method was utilized for algorithms that required tuning of more than one parameter.

3.7 Model validation and performance comparisons

In this study, a random split validation method was applied to all the eight estimated models, as discussed in Section 2.4. While comparing the performances, the study utilized quantitative (model predictive power, classification accuracy and goodness-of-fit) and qualitative (ease of interpretability, modelling effort and significant parameter towards classification exposure) performance measures or criteria as elaborated in the following section. These criteria were developed in consultation with maintenance practitioners.

(1) Model predictive power

The area under the receiver operating curve (AUC) is utilized in classification models to indicate the predictive power of the model, hence the comparison of this parameter for the respective model's guides to a more efficient model for prediction, where a high AUC value indicates superior predictive power.

(2) Classification accuracy (ACC):

Each model classifies the data during the testing phase and returns a classification table known as a confusion table. The classification accuracy represents the proportion of the combined number of accurate predictions, which will indicate the model accuracy.

(3) Model Goodness of fit:

Gini coefficient represents the ratio between the area between the receiver operating curve and the diagonal line and the area of the above triangle. Gini coefficient above 60% infers a sustainable model.

(4) Ease of interpretability:

The output from the models will also be examined on the ease and straightforwardness of the interpretation. In this case, the model should be capable of indicating how changes in lubricant parameters would influence the oil analysis sample classification. This aspect is crucial to maintenance personnel which requires the output for maintenance decision support as relates to the UOA parameters and their effect on the machine and lubricant health condition.

(5) Modelling effort:

This criterion will evaluate the effort required in modelling and results generation as well as results analysis. Some models need processing the input data while others may additionally require more effort in tuning (few or moderately more hyperparameters), adjustment, calculations to obtain the desired output at various scales.

(6) Significant parametric exposure:

This part of the criteria seeks to check the relevance and applicability of a model to explain the lubricant parameter's influence on the classification of the sample. A vital variant in maintenance decision derived from the oil analysis by practitioners.

4. Results and discussion

The following section outlines the results and discussion following the set methodology illustrated in Figure 1.

4.1 Data testings and correlation analyses

4.1.1 Data pre-processing scoring and testing. The data pre-processing incorporated several steps: the first step involved integrating the data into one file since each sample results was separate. This was achieved by considering the dates and running hours of the respective samples and where clarity was needed, expert consultations were made. The data was cleaned in the second step, where data consistency was harmonized by inspecting the data, and we expunged errors like duplicated results. As earlier alluded, missing values were also investigated in this step, where molybdenum was withdrawn from the database as it had over 32% missing values. This action was taken after literature and expert consultation where it was considered uninfluential, hence each sample eventually contained twenty parameters.

While inspecting the data, we implemented graphical representations such as histograms and box-plots, *z*-scores and benchmarking of sample results to comprehend the extent of the outliers (how far an observation is from the mean in terms of standard deviation); however, no invalid outliers were noted. The ultimate step of data transformation included standardization, where data was scaled because some parameters such as calcium and water exhibited different measurement scales. Normality testing was performed, and 25% of the data was found to follow a normal distribution, while the rest were non-parametric. Other data collected included the type of lubricant used in the engines whose fundamental properties included Viscosity at 40°C (135cSt), Viscosity at 100°C (14cSt), TBN (40 mg KOH/g), Flashpoint (220°C), Viscosity index (100) and pour point (−18°C), which correlates to the industry requirements for medium-speed engines (300–900 rpm) running on HFO. As expected, due to the operations and potential effects as illustrated in Section 3.1, a moderately viscous lubricant with a considerable alkaline reserve (TBN) is used.

4.1.2 Correlation analysis. Figure 2 illustrates a correlogram developed exhibiting the correlation of the oil parameters. The significant correlations (p-value < 0.05) have respective correlation coefficient indicated, where higher values close to 1 indicate strong correlations while those close to 0 indicate weak correlations, and insignificant correlations are blank.

For example, the correlation between the parameters Nickel-Vanadium had a p-value of $2.2 \times 10^{-16} < 0.05$, hence significant with a Spearman's correlation coefficient of 0.77, signifying a positive correlation. The potential sources of vanadium and nickel in the engine include alloy metals and heavy fuel oil contamination. Since both nickel and vanadium are present, there is a high indication of fuel contamination (Wakiru *et al.*, 2017a). This correlation

	LEAD	COPPER	ZINC	FINA WATER	CHROMIUM	IRON	ALUMINIUM	SILICON	FLASH POINT	MAGNESIUM	SODIUM	PENTANE INSOLUBLES	VISCOSITY @ 40	VISCOSITY @ 100	CARBON CONTENT	NICKEL	VANADIUM	TIN	TBN
COPPER	0.54																		
ZINC	0.29	0.38																	
FINA WATER	0.26	0.33	0.65																
CHROMIUM	0.37	0.56	0.38	0.43															
IRON	0.35	0.54	0.41	0.53	0.73														
ALUMINIUM	0.16	0.4	0.49	0.49	0.67	0.77													
SILICON	0.25	0.45	0.36	0.36	0.56	0.66	0.75												
FLASH POINT	0.25	0.4	0.51	0.7	0.53	0.61	0.66	0.63											
MAGNESIUM	0.23	0.36	0.5	0.5	0.52	0.66	0.8	0.8	0.76										
SODIUM	0.15	0.27	0.4	0.58	0.51	0.68	0.74	0.68	0.74	0.83									
PENTANE INSOLUBLES			0.14	0.26	0.28	0.3	0.4	0.31	0.44	0.46	0.53								
VISCOSITY @ 40	-0.27	-0.19		0.19	0.32	0.42	0.47	0.26	0.26	0.3	0.46	0.38							
VISCOSITY @ 100	-0.34					0.29	0.32					0.25	0.86						
CARBON CONTENT	-0.39	-0.33	-0.12	-0.19			0.17		-0.13			0.08	0.5	0.61					
NICKEL			0.32	0.32	0.53	0.57	0.57	0.24	0.27	0.29	0.42	0.27	0.68	0.68	0.51				
VANADIUM	-0.18	-0.13	0.09		0.29	0.33	0.22	-0.05	-0.11	-0.09	0.06	-0.05	0.49	0.61	0.57	0.77			
TIN	-0.24	-0.43	-0.66	-0.66	-0.56	-0.57	-0.69	-0.6	-0.79	-0.71	-0.63				0.06	-0.33	0.05		
TBN	0.01		0.1				0.25	0.38	0.14	0.45	0.3				0.1	-0.06	-0.1		
CALCIUM	-0.12	-0.14	0.15	-0.32	-0.13	-0.15			-0.28		-0.14		-0.08		0.29	0.05	0.18		0.51

Figure 2.
Correlation correlogram for lubricant samples

and possible insight are confirmed by the correlation of nickel and viscosity at 40°C, which is 0.68 (relatively strong). With heavy fuel oil ingression, viscosity increases; hence, one can conclusively interpret the correlation representing fuel contamination exposed. The analysis, interpreted along with literature and expert review, uncovers several embedded patterns in the data that eventually assist in maintenance decision support. From this one example, it shows two-dimensional correlation can be used in multiple ways, though time-intensive and requiring technical insights, to discover hidden patterns in the UOA data of the plant. The discovered pattern informs appropriate maintenance decision actions towards addressing the causes of fuel dilution such us faulty injector pump, injector nozzles, poor atomization and fuel condensation during idling. Contrary to this, the analysts in the industry rely on one parameter trending, which cannot conclusively confirm the nature and severity of the problem (Fitch, 2007).

4.2 Variable selections and dimension reduction

PCA and CA were used to select the appropriate variables from the UOA dataset, as outlined in the next section.

4.2.1 Principal component analysis (PCA). The optimum number of five PCs was obtained employing the non-graphical solutions and scree test, corroborated by eigenvalues greater or equal to 1 since the data had been standardized and the scree plot technique employed.

Oblique rotation was considered to best suit the analysis, following the Tabachnick and Fidell criterion and the assumption that factors are correlated. A meaningful interpretable result is provided, which identifies the variables with significant influence in constituting a principal component. The factors or PCs representation in the factor loading ends up being different. However, parameters grouping remain similar with better visualization, as seen in Table 6 using 0.5 as a cut-off, after rotation. PC1 can be interpreted as flashpoint and related contaminants in the lubricant, while PC2 as viscosity related, moreover is corroborated by the example in the correlation analysis in Section 4.1.2. PC3 as the wear-related attributed to being constituted by wear metals, while PC4 as alkalinity concentration of the lubricant, also corroborated by Wakiru *et al.* (2017a) and PC5 as water-related. Table 6 shows the factor loading values and the PC's constituents after oblique rotation method of PCA. Using a cut-off of 0.5 loading criteria, the parameters like viscosity at 100°C, tin, pentane insoluble and aluminium fall outside the scope.

Despite this being visual and arbitrary, it is a crucial classification of multiple variables extending beyond correlation analysis explained in Section 4.1.2. PCA confirms the interaction of more than two variables in influencing a principal component which enhances multivariate pattern recognition of the oil parameters.

	Loading	PC		Loading	PC
Flash point	0.78	1	Lead	0.71	3
Magnesium	0.73	1	Copper	0.79	3
Silicon	0.59	1	TBN	0.74	4
Sodium	0.82	1	Calcium	0.78	4
Viscosity @40°C	0.63	2	Zinc	0.81	5
Carbon/soot	0.64	2	Water	0.51	5
Nickel	0.90	2	Viscosity @ 100°C		
Vanadium	0.87	2	Tin		
Iron	0.71	3	Pentane Insoluble		
Chromium	0.69	3	Aluminium		

Table 6. Factor loadings alongside principal components and cluster constituents

4.2.2 Cluster analysis. Cluster analysis of the twenty UOA parameters was carried out using the hierarchical average method with the computation of the probability values for the clusters. K-means Bootstrapping was used to determine the optimal number of clusters. In this case, seven indices which were the highest compared to the rest, proposed three as an optimal number of clusters, hence adopted. Figure 3 illustrates the hierarchical clustering dendrogram, where one can visualize three clusters because three branches occur at about the same height.

The first cluster with AU p-value of 0.99 and BP p-value of 0.4, contains TBN and calcium, can be interpreted to depict the alkalinity of the lubricant. This deduction is further corroborated using literature that calcium remains an ingredient of detergent and dispersant, which are additives used in enhancing TBN in the lubricant (Kocsis *et al.*, 2017). The second cluster with p-values of 1 and 1 for AU and BP respectively, contains zinc, water, iron, aluminium, sodium, flash point, magnesium, silicon, chromium, lead and copper, depicting both wear metals and the lubricant properties. The third cluster with p-values 0.99 and 0.60 respectively, consists of viscosity at 40°C, vanadium, nickel and carbon content, might be depicted as the fuel contamination related cluster. Oxidation due to the elevated operating temperature of the equipment or denser contaminants like insoluble soot (carbon), water and heavy fuel oil potentially increases the value of viscosity. Since fuel contains nickel and vanadium as part of its ingredients as alluded by CIMAC (2011), these elements may lead to a potentially valid reason for the increase in the viscosity especially when there is fuel ingression in the lubricant.

To the contrary, several associations were not easy to distinguish and substantially elucidate from the functional perspective of the represented UOA parameters. For example, in cluster 2, the association between flash point and aluminium is depicted to be statistically correlated, yet in the functional sphere, the two parameters do not correlate.

This combination of correlation analysis, PCA and CA findings provides some support for the conceptual premise that infers remarkable alkaline level depletion and fuel contamination in the lubricant. Firstly, addressing the pattern exposed between Calcium and TBN, indicating depletion of the lubricant's alkalinity reserve may be attributed to several factors. Such factors range from engine status like blow-by, operational conditions like temperature and fuel quality. The challenges of alkaline depletion can be addressed by changing the lubricant if the depletion levels fall beyond 50% of new oil and addressing causes like fuel ingression that increase acidity level hence adversely affect the TBN (CIMAC, 2011; Kocsis *et al.*, 2017). Carrying out a partially forced recharge of the engine oil to replenish the alkali reserve of the lubricant, could offer a potential solution when the alkaline level is above 63% of the new oil (in this case 25 mg KOH/g) also corroborated by (CIMAC, 2008). Partially forced recharge should be a short-term recommendation since there is evidence of a correlation between additive depletion and fuel dilution as addressed in the following point. Secondly, addressing the source of the fuel ingression to mitigate fuel contamination, includes interventions such as repairing or servicing the injector pump and nozzle. Additionally, routinely carrying out inspection and or rectification of leaking high-pressure fuel lines, leaking oil/fuel heat exchangers which are potential sources of fuel dilution, would address the problem. Finally, as a last resort when additive depletion is recurrent, change of the lubricant to one containing higher alkaline levels, for instance, 50 mg KOH/g, in this case, would address this problem. However, this decision would require consultation with the OEM and several compatibility tests carried to ensure non-compromise with the engine and lubricant functionalities and performance. This combination of findings, taken collectively, suggest that the engine optimal drain interval (to ensure optimal TBN level) and fuel ingression should be addressed. The maintenance team can monitor the TBN levels to derive optimal drain intervals, using techniques such as linear regression to predict the time the level will deviate from the threshold using extrapolation. However, consideration of other factors

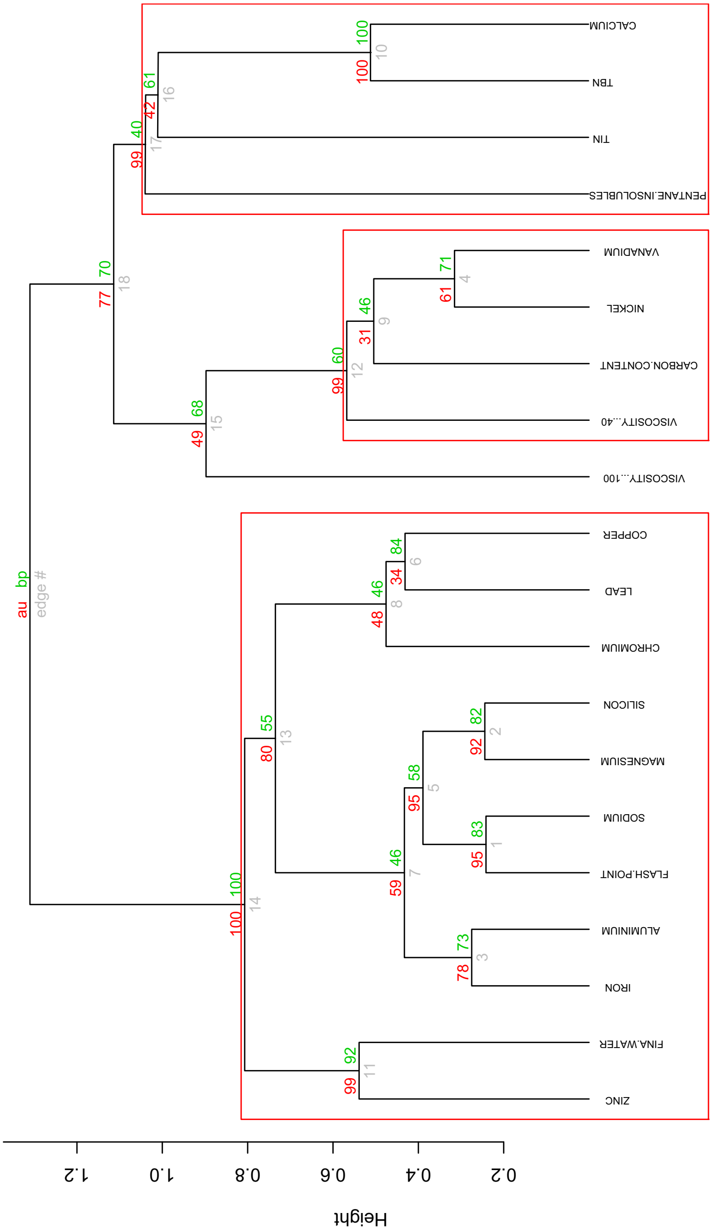

Figure 3.
Cluster dendrogram for correlation using average distance with AU/BP values (%)

such as wear levels, other lubricant parameters that interact with TBN such as sulphur level in the fuel, specific oil consumption of the engine and operating conditions should be considered while formulating a trade-off decision regarding the drain interval.

4.2.3 Selection of variables and feature reduction. During the selection of the critical variables to be used in the predictive model, to begin with, the exclusion criteria remain the 0.5 cut-off value for the PCA, corroborated in the literature (Bro and Smilde, 2014; Jolliffe, 2002). Accordingly, viscosity at 100°C, tin, pentane insoluble and aluminium were eliminated. Secondly, the cluster height on the dendrogram for the same parameters, except aluminium, retained higher heights and thirdly, literature and expert assessment to intuitively analyse the variables. The sixteen variables selected included viscosity at 40°C, flash point, total base number, magnesium, calcium, zinc, silicon, sodium, water, carbon, iron, chromium, lead, copper, nickel and vanadium.

In summary, the three approaches, correlation analysis, CA and PCA converge with concurrences in the correlations, albeit correlation analysis demands more effort to envision the grouped correlations.

This study incorporates a feature reduction technique while developing the classification model. Thus, RF, PCA, LASSO and correlation with target approaches were tested for suitability in the subject study using sixteen features. Table 7 shows classifier's performance utilizing the various methods.

From the exercise, RF obtained better performance in both AUC and ACC, hence selected as suitable feature reduction technique to be incorporated while developing the classification models.

4.3 Scoring of data

Based on the expert assessment and OEM set thresholds/limits of each parameter, the sample was manually classified as "PASS" or "FAIL" by the analyst. However, it was observed that the satisfactory sample rate was significantly low when compared against the samples that were categorised as failed. The inclusion of samples that signal caution in the "FAIL" category represents the likely cause for the difference between the two classification rates; in this case, the parameters are at the borderline of the thresholds. It is proposed that the future classification of sample results should embed a three-tier level, "PASS", "CAUTION" and "FAIL", to address this unique situation. After consolidating the data, scaling/normalisation of the variables was done due to the variation of the scales used. For example, values of viscosity at 40°C of 152cSt, Calcium of 12,445ppm and silicon of 30ppm, after rescaling from a value of 0–1, were 0.2288, 0.6704 and 0.8902 respectively. The classification models were built incorporating RF feature reduction. Hyperparameter optimisation was done as documented in Section 3.6.

Technique	Measure	SVM	NN	LR	RF	DT	XGB	KNN	ERT
RF	AUC	0.951	0.951	0.942	0.997	0.961	0.989	0.936	0.947
	ACC	0.929	0.900	0.900	0.976	0.957	0.971	0.900	0.905
PCA	AUC	0.949	0.940	0.935	0.933	0.836	0.929	0.940	0.934
	ACC	0.905	0.881	0.900	0.900	0.810	0.905	0.914	0.905
Correlation with target	AUC	0.954	0.935	0.948	0.986	0.959	0.981	0.913	0.940
	ACC	0.914	0.886	0.890	0.962	0.957	0.957	0.886	0.900
No reduction	AUC	0.953	0.942	0.926	0.972	0.961	0.999	0.910	0.940
	ACC	0.929	0.900	0.895	0.929	0.957	0.995	0.871	0.886

Table 7.
Feature reduction techniques performance results for comparison

4.4 Classification models

Table 8 illustrates the performance of the various classification models employing the case data. Among the "white box" approaches, DT outperformed LR while among "black box" approaches; XGB performed better followed by RF considering all the three measures.

4.4.1 Logistics regression. The model returned ACC of 0.8857 and AUC of 0.9464 as seen in Table 8. Table 9 presents an overview of the LR model with each parameter's coefficient, odds ratio and *p*-value. The *p*-values indicate variables like zinc, sodium and nickel predict the classification result significantly at a 5% level of significance. The variable coefficients illustrate the relation an oil parameter has with the classification outcome. From the results, the particular variable with the highest importance on the classification of the oil as "FAIL" is sodium. The odds ratio for sodium is 0.8973, which infers that an increase of sodium by 1unit (ppm) will decrease the odds of the oil sample outcome classification as "PASS" by 10.27%. This finding will prompt the decision to mitigate the ingression of Silicon, for instance, investigate to confirm proper functionality of the air filtration system, changing the air filters and addressing the origin of dust considering the operational environment. In the long run, the plant may engage the off-line centrifuge system to eliminate such contaminants during the scheduled centrifuging process. A similar analysis can be followed for other oil parameters to understand the impact selected parameters have on the oil classification, with easy interpretability. Modelling using LR offered valuable insights into the oil parameters relating to the classification outcome. However, the interpretation of the coefficients and odds ratio involves moderate knowledge hence some effort required.

4.4.2 Decision tree. The model yielded a classification accuracy of 0.8905 and AUC of 0.9507. Figure 4 illustrates the schematic representation of a decision tree generated. The nodes represent the oil parameter attributes through which, by tracing them, one can reach the leaf nodes in the bottom of the tree, which depict the classification of the oil sample.

Interpretation of the decision rules is straightforward; for example, we classify the sample whose UOA results are shown in Table 10 using the DT from the top node. The sample passes the thresholds of sodium, vanadium, calcium, viscosity at 40°C. Still, it fails due to zinc content being less than 299ppm. Hence, one can explore causes of zinc content decrease, that include increased metal to metal contact causing depletion of anti-wear additives and anti-

Measure	SVM	NN	LR	RF	DT	XGB	KNN	ERT
AUC	0.9528	0.8149	0.9464	0.9906	0.9507	0.9993	0.9020	0.9840
ACC	0.9333	0.7905	0.8857	0.9857	0.8905	0.9905	0.8905	0.9290
Gini	0.9056	0.6298	0.8928	0.9812	0.9014	0.9986	0.8040	0.9680

Table 8. Classification model's comparison based on model performance characteristics

Variable	Coefficient	Odds ratio	*p*-value	Variable	Coefficient	Odds ratio	*p*-value
Intercept	1.5299		0.6899	Calcium	0.0001	1.0001	0.2984
Viscosity 40°C	−0.0261	0.9742	0.1575	Water	−13.3763	0.0000	0.1570
Flash point	0.0124	1.0124	0.4230	Carbon	1.1233	3.0748	0.0812
Zinc	0.0081	1.0082	0.0034	Iron	0.0200	1.0202	0.5121
Silicon	0.0256	1.0259	0.1777	Chromium	−0.1362	0.8727	0.2390
Sodium	−0.1084	0.8973	0.0000	Lead	0.0446	1.0456	0.6565
TBN	−0.0044	0.9956	0.9007	Copper	−0.0225	0.9778	0.5337
Nickel	−0.0582	0.9435	0.0035	Vanadium	−0.0103	0.9897	0.1128
Magnesium	−0.0210	0.9793	0.1913				

Table 9. Binary LR model summary of oil parameters

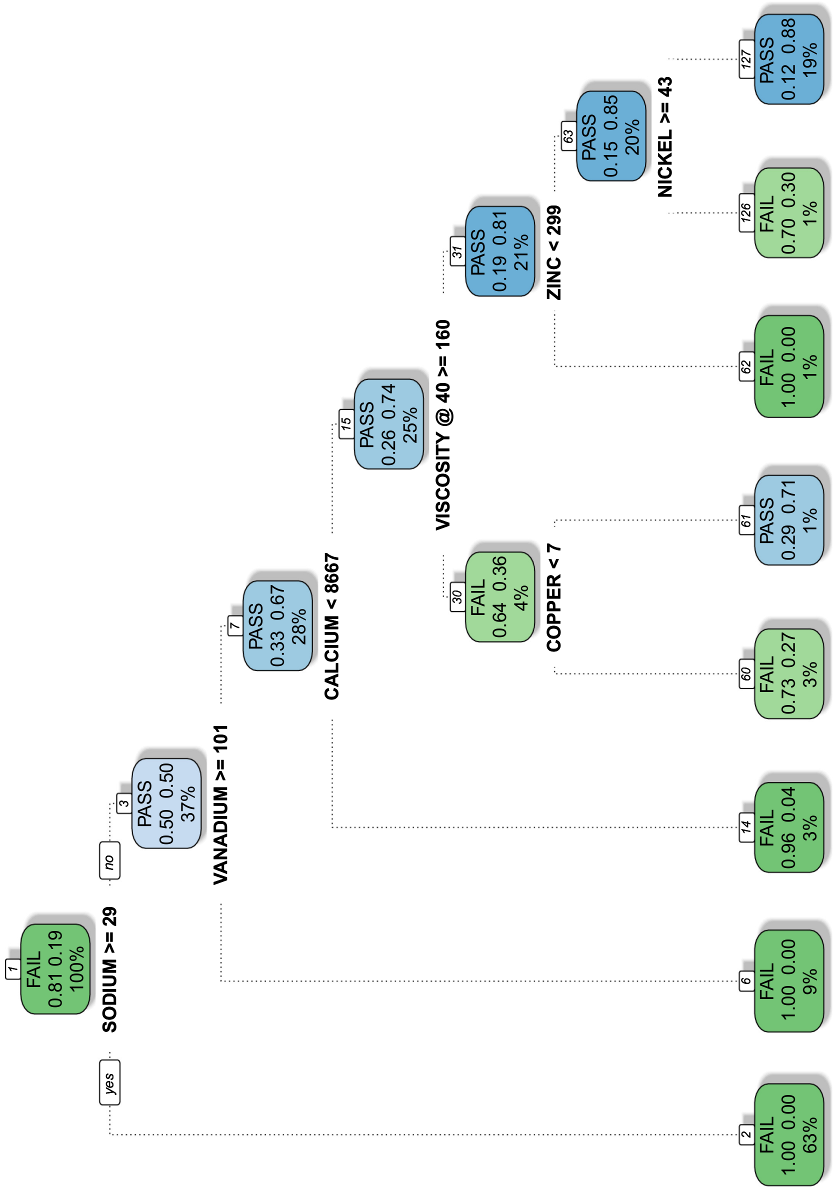

Figure 4.
Schematic representation of binary decision tree model

oxidation inhibitors, probably caused by high-temperature operations or fuel dilution on the lubricant, hence leading to lower lubricity challenges. To address such a challenge, while considering the patterns revealed, the sources of potential fuel dilution like injector pump leaks or malfunctioning injector nozzle should be checked and or rectified. Replacement of injector pump gaskets and seals would address the potential leakage cause if the pump were in good condition. Other potential areas to address include the viscosity of the lubricant in use, where changing to a higher viscous lubricant will reduce the metal to metal contact, whose friction raises the oil temperature. However, with relatively few variables, caution must be applied, as the findings might be uncomprehensive if other variables are unconsidered.

4.4.3 Random forest. The model achieved a classification accuracy of 0.9857 and AUC of 0.9906, as shown in Table 8. Despite high accuracy and goodness of fit, the model lacks simple and straightforward interpretability and exposure of the effects caused by different oil parameters in the model, making it hard to understand the embedded knowledge pattern in the data analysed. Reviewing all the trees generated is time-intensive and unable to offer coherency in the individual classification.

4.4.4 Neural network. The model achieved an ACC of 0.7905 and AUC of 0.8149. Building the NN model required data pre-processing by transformation using normalization. As the number of hidden layers decreased, the predictive accuracy of the model improved. The model lacked a simplistic and straightforward approach to interpret the results. Moreover, the model did not explicitly exhibit the influence of the various parameters towards the sample classification but showed the weightings of the parameters on the plot which most of the times is not readily retrievable while building the model.

4.4.5 Support vector machine. The model exhibits ACC of 0.9333 and AUC of 0.9528. Visualization of the various parameters based on various kernel types like sigmoid and linear can be shown in two-dimensional. Despite the lack of direct and straightforward interpretation, the moderate effort shows two-dimensional interpretability of the effect of the variables towards sample classification, while data must be transformed using normalization.

4.4.6 Extreme gradient boost. The model exhibits ACC of 0.9905 and AUC of 0.9993. Despite the satisfactory performance of the algorithm, interpretation of the influence of parameters to the classification is unattainable. Nevertheless, XGB possesses the capability of generating the variables of importance.

4.4.7 K nearest neighbour. The model exhibits ACC of 0.8905 and AUC of 0.9020, like SVM and NN, it does not evaluate the importance of the variables in the classification of the outcome.

4.4.8 Extremely randomized trees. The model exhibits ACC of 0.9290 and AUC of 0.9840 while it evaluates the importance of the variables in the classification of the outcome like RF.

RF, XGB, and ERT evaluated the importance of the variables in the classification of the outcome; it infers sodium represents the most significant variable in the classification followed by nickel, vanadium, calcium, viscosity and iron. Sodium sources in lubricant include water and or coolant leaks, while nickel and vanadium are derived from alloys or HFO. This is similarly depicted by the concurrence of LR and DT picking the same variables

Variable	Value	Variable	Value	Variable	Value	Variable	Value
Viscosity 40°C	144	Calcium	9,859	Sodium	5.1	Chromium	0.15
Flash point	14.7	Water	0.1	TBN	34	Lead	2
Zinc	238	Carbon	0.11	Nickel	7.7	Copper	6.5
Silicon	12	Iron	12.2	Magnesium	22.6	Vanadium	12.2

Table 10. Example of sample results for classification

as the most significant oil parameter in the classification of the oil sample as "FAIL". Further in-depth investigation of these variables by the maintenance team could potentially reduce the lubricant failure and enhance the life of both the machine and lubricant.

4.5 Classification models' comparisons

From the results in Table 8, XGB attained the most substantial predictive power (AUC), followed by RF, ERT, SVM, DT, LR, KNN and NN. High predictive power infers the model can accurately predict the classification of the sample hence offers an effective and efficient diagnosis to prompt maintenance intervention. While evaluating ACC, XGB model exhibits the highest value followed by RF, SVM, ERT, DT, KNN, LR, and NN. XGB produces a prediction model, with stage-wise approach optimizing an arbitrary differential loss function in the form of an ensemble of other models (typically decision trees with weak prediction). The utilization of ensemble ensures not only XGB retains high accuracy, but also it is fast and outperforms other techniques generally. While reviewing the goodness of fit criterion using Gini, XGB had the highest, followed by RF, ERT, SVM, DT, LR, KNN and NN. This goodness of fit results infers the model employing XGB is correctly specified and fits the data used. This test reveals the discrepancy between observed values and the values expected under the model in question, hence determines whether sample data are consistent with a hypothesised distribution. Hence, the engineer can confirm the classification model fits the set of observations. Overall, DT a white box technique performs comparatively well compared to the black box techniques. Among the black-box techniques, XGB and RF exhibit better performance. However, NN's Gini value is at the threshold as alluded in Section 3.7.

Overall interpretability of the different classification models was with varied ease. Regarding interpretability ease, DT had better and straightforward interpretability, followed by LR, which requires some slight efforts to interpret the odds ratio and *p*-values. The remaining models were complex to interpret. Nonetheless, RF, XGB and ERT generated the significant variables quantitatively that may be utilised to offer interpretation insights, like the critical variables that profoundly influence the lubricant sample to be classified as failed. Despite offering this interpretation, no further exposition is drawn from the significant variables.

Modelling SVM, RF, LR, XGB, KNN and ERT required moderate effort due to the number of parameters involved in the tuning exercise, for instance, they all had more than two hyperparameters often search grid method had to be applied. Contrary to this, DT and NN had fewer parameters for optimisation during the tuning.

While reviewing the ease of the estimated models to expose the various lubricant parameter's impact or influence on the classification of the oil samples, DT and LR showed remarkable results. DT offered traceability of respective oil parameter towards the classification of the sample, while LR depicted the relationship of an oil parameter to the sample classification. RF, XGB and ERT provided only the estimation of the variable's importance to the classification, while SVM, KNN and NN did not exhibit variables interpretation. These classifiers are considered as "black box" which have limited ability to explicitly expose the causal relationships between the explanatory variables or lubricant parameters hence no probability of class memberships is available (Phillips *et al.*, 2015). Due to this challenge, one cannot easily obtain predictions from the model analysis save whether the machine state and condition is okay or not (healthy/not healthy) and which dependent variables are significant. However, in data sets where they return significant predictive power and accuracy, the aspect could be traded off if the other aspects are unnecessary.

Table 11 illustrates the scores for each of the evaluated model considering the researcher's criteria set for comparing the model's performance. DT and LR based classification model from this study satisfy the researcher's established criteria, despite the moderate effort in modelling. From the criteria set, it is worth also to note that selection of one model over the other ultimately has a trade-off in one or more aspects, and no single classifier algorithm yields a more significant result for every dataset also corroborated by Tomar and Agarwal (2013). An example of selecting LR over RF would imply a loss of 10% accuracy and 4.42% predictive power for good and easy interpretability.

On the contrary, depending on the user preference, LR will offer better interpretability of the impact of variables changes on the probability of the machine and lubricant condition being healthy or not. However, in instances where performance and accuracy of the model is the overarching objective, an ensemble algorithm XGB would offer the best alternative. Comparison of the separate models is likely to be primarily dependent on the nature of a specific data set, and one cannot conclude whether one model will be superior to the other model in each data set, also corroborated by Raza *et al.* (2010).

5. Implications for decision support in practice

The developed decision support system under the case context offers several recommendations following the expectations raised in Section 1.2. To begin with, the plant and analysts need to consider broadening the current two-tier ("PASS" and "FAIL") classification to a three-tier level (introduce "CAUTION") as discussed in Section 4.3. The broad-spectrum will offer categorized intervention requirements like inspection of possible ingression of dust via the air filtration system when Silicon is classified as "CAUTION". On the other hand, partially forced recharge, change of air filters and off-line centrifuging of the lubricant could be employed when the silicon level is classified as "FAIL". Secondly, sodium, vanadium, calcium and nickel are exposed as critical lubricant parameters exhibiting strong influence towards the samples being classified as failed. Addressing such exposed critical parameters deviations will reduce lubricant related faults and failures significantly. Furthermore, fuel dilution, water and dust (silicon) ingression into the engine are exposed in the case context as the primary mechanical fault causes. Ultimately, the use of Decision tree and Logistics regression model offers the most appropriate support considering both quantitative and qualitative aspects derived from the developed criterion. In contrast, XGB offers the most superior accuracy. However, there are additional essential implications we derive from this study, as highlighted in the following part.

As indicated in Section 4.1.1, data integrity represents an essential facet for successful DSS results. The challenges of missing data and consistency could be addressed by clear procedures and protocols that define sampling frequency, sampling point and critical parameters to be tested. A predictive approach utilising shorter oil sampling interval, in our analysis would offer the opportunity for timely intervention, enhance maintenance planning and scheduling (source lubricant and spares, e.g. worn bearings, injector pumps) to reduce unplanned downtime significantly. Incorporation of on-line condition monitoring undoubtedly offers more accuracy and benefits. Moreover, un-procedural sample handling potentially could undermine the quality of the sample, hence negatively impact the data integrity. The un-procedural sampling could be addressed by skills enhancement like

Criteria		DT	RF	SVM	LR	NN	XGB	KNN	ERT
Quantitative	High predictive power	✓	✓	✓	✓	✓	✓	✓	✓
	Data classification accuracy	✓	✓	✓	✓	✗	✓	✓	✓
	Model goodness of fit	✓	✓	✓	✓	✗	✓	✓	✓
Qualitative	Ease interpretability	✓	✗	✗	✓	✗	✗	✗	✗
	Low modelling effort	✗	✗	✗	✗	✓	✗	✓	✗
	Significant parametric exposure	✓	✗	✗	✓	✗	✗	✗	✗

Table 11. Classification models' comparison employing the researcher's criteria

training and investment on oil sampling equipment (vacuum pump, sample containers) and testing equipment.

Moreover, setting up of appropriate procedures and installation of equipment for charging the system with new lubricant is fundamental to ensure the quality of the lubricant in the system, in this case, the engine, since this is a salient means that ingression of contaminants occurs. An in-house or on-site laboratory would suffice for large-scale plants, depending on the sampling frequency, while portable oil testing kits can be employed for immediate interventions purposes while awaiting extensive laboratory test results. Despite the perceived moderate cost of oil analysis tests (depending on the type and number of parameters to be tested), plants should be able to ensure rapid sample turnaround time and enhance in-house knowledge. Ultimately, the plant will improve their bottom-line because of increased problem detection, which implies significant maintenance costs reduction if implemented regularly and articulately. The installation of on-line condition monitoring using sensors would address not only the sample and data integrity, but also ensure continuous or significantly increased sampling/testing, accuracy and prompt decision making. However, in such plants, caution should be taken to evaluate the cost-benefit analysis on the predictive method to be adopted where increased sampling could offer better economics for ageing equipment because installation of online sensors may be expensive and require excessively more configurations.

Statistical analysis methods like trending have been utilized significantly in the industry which offers uncomprehensive and incomplete decision support as alluded in Section 1.1. This is attributed to the reliance of the customarily used one factor at a time analysis that retains non-consideration of the interaction between lubricant parameters, an aspect also corroborated by (Wakiru *et al.*, 2017b). Methods like Correlation, CA and PCA offers more insights to DSS. However, the use of trend analysis and Correlation can be employed as a first-hand diagnosis analysis due to ease of application and accessibility. The mentioned techniques are easy to learn and use because a generic code, tailor-made to the parameters considered by distinct plants can be developed. Moreover, incorporation of a graphical user interface (GUI) to ease interpretability and offer first-hand key insights for prompt intervention can be advanced, as the extensive analysis (such as exhibited by correlation, PCA and CA) which is time intense follows later.

Likewise, this study advanced a scalable framework that considers all the lubricant parameters while selecting the relevant variables for the classification process. This systematic selection methodology ensures that each set of datasets is evaluated, and significantly relevant parameters are employed while developing the model. This aspect will significantly reduce not only the time involved in modelling but also ensure comprehensive results that factors both the historical aspect of the machine being lubricated and the interaction characteristics of the oil parameters. An all-inclusive model developed from this systematic variable selection offers robust, reliable classification results.

While considering the classification of the oil analysis results, this will help the maintenance manager to discover the various factors to consider in developing a robust model and mitigate the limitations as alluded earlier, where studies considering UOA remain narrow in focus, dealing merely with a single or at most five oil parameters as illustrated in Table 3, in Section 2.5. These findings contribute in several ways to our understanding of classification algorithms as applied to UOA. In the first place, the performance of the algorithm will depend on the feature or variable selection and reduction used, as seen in Section 4.2.3, where it is essential to select the variables to use systematically. In the second place, the use of white and black box classification models offers various advantages and disadvantages separately. Moreover, the use of a hybrid approach would add value to the maintenance engineer. However, due to various constraints expected such us time, accessibility and skills, a selection criterion like advanced in this study would assist the engineer in selecting and focusing on the appropriate algorithm(s), hence, ensure the development of expertise and user preferences are met. This selection criterion offers two fundamental factors to consider in practice; this is qualitative and quantitative factors. In addition to extensively used quantitative factors like model accuracy and predictive power, it has been shown that qualitative aspects remain equally crucial while selecting the appropriate classifier from a user perspective in terms of effort and skill, comprehensibility and interpretability also corroborated by (Martens *et al.*, 2011). This combination of findings provides some support for the conceptual premise that despite the low statistical and data mining experience as would be expected among engineers, appropriate model selection adopting such a framework is insightful and easy to use.

6. Conclusion and future work

The primary goal of the current study was to develop an integrated framework to analyse maintenance data (UOA data) using an integrated data mining formalism and to evaluate the extent to which this integrated framework is useful for maintenance decision support using classification models. The relevance of integration is supported by the current findings, where meaningful parametric patterns were exposed while applying several data mining techniques. The revealed patterns by correlation, PCA and CA, besides assisting in maintenance decision support, additionally offer a more thorough understanding of the associations. PCA was more significantly used to confirm the number of appropriate oil parameters utilised in the binary classification models, while the study developed a systematic classification framework. In the developed framework, pre-processing of the UOA data scored by the analysts has been shown as a salient aspect in data preparation before use as input in the binary classification models. The framework has demonstrated the value of selecting essential variables to be employed in the classification models using a clearly defined basis explicitly and tangibly. LR, XGB, KNN, ERT, RF, DT, SVM and NN models were estimated using the same data set and validation process. The study employed the developed model selection criterion and revealed the selection of a suitable model involves picking one which best suits the dataset, preferences, needs and expectations and further, not necessarily the most accurate model. The present study has advanced some way towards enhancing the understanding and application of data mining techniques in fault diagnosis and classification while using maintenance data, in this case, oil analysis data. These results besides strengthening the view that performance of a model is dependent on the dataset and the user preferences, as well offer an automated framework for evaluating lubricant samples overcoming various challenges associated with the manual process. Such automated frameworks generated from real-plant data are important in complementing maintenance decision support. In our view, this offers the maintenance managers insight on the potential of data mining as well as classifiers in the strategic and tactical actions related to maintenance decision support.

One limitation of this study represents the use of one case study data in the analysis. To address this aspect, the employment of the developed DSS framework considering used oil analysis results from other types of equipment like gearboxes or hydraulic would constitute part of the subsequent research. Moreover, comparing the various classifier's dependency on the number of classes where, as alluded in Section 4.3, three classes "Fail", "Caution" and "Pass" could be incorporated as the prediction variables (multinomial classification) will be employed and a comparison to validate the dependencies performed in the future research. A further study with more focus on the number of classes and dependencies therein is therefore suggested. Ultimately, subsequent research would benefit from setting thresholds or weights to the researcher's criteria of model evaluation and selection while exploring lubricant data from other fields which was unexplored by this study.

References

Adnani, S.A., Hashemi, S.J., Shooshtari, A. and Attar, M.M. (2013), "The initial estimate of the useful lifetime of the oil in diesel engines using oil analysis", *Tribology in Industry*, Vol. 35 No. 1, pp. 61-68.

Aggarwal, C. (2015), *Data Mining: The Textbook*, Springer International Publishing, Switzerland, doi: 10.1007/978-3-319-14142-8.

Balabin, R.M. and Safieva, R.Z. (2011), "Biodiesel classification by base stock type (vegetable oil) using near-infrared spectroscopy data", *Analytica Chimica Acta*, Vol. 689 No. 2, pp. 190-197.

Bro, R. and Smilde, A.K. (2014), "Principal component analysis", *Analytical Methods*, Vol. 6 No. 9, pp. 2812-2831.

Caesarendra, W., Widodo, A. and Yang, B.S. (2010), "Application of relevance vector machine and logistic regression for machine degradation assessment", *Mechanical Systems and Signal Processing*, Elsevier, Vol. 24 No .4, pp. 1161-1171.

Cai, Y., Chow, M., Lu, W. and Li, L. (2010), "Evaluation of distribution fault diagnosis algorithms using ROC curves", IEEE PES General Meeting, Providence, RI, pp. 1-6.

Campora, U., Cravero, C. and Zaccone, R. (2018), "Marine gas turbine monitoring and diagnostics by simulation and pattern recognition", *International Journal of Naval Architecture and Ocean Engineering*, Vol. 10 No. 5, pp. 617-628.

Carla Bittencourt Reis, A., Cabral Seixas Costa, A.P. and Teixeira de Almeida, A. (2009), "Planning and competitiveness in maintenance management", *Journal of Quality in Maintenance Engineering*, Vol. 15 No. 3, pp. 259-270.

Carmona, P., Climent, F. and Momparler, A. (2018), "Predicting failure in the U.S. banking sector: an extreme gradient boosting approach", *International Review of Economics and Finance*, Vol. 61, pp. 304-323.

Cerrada, M., Zurita, G., Cabrera, D., Sánchez, R.-V., Artés, M. and Li, C. (2016), "Fault diagnosis in spur gears based on genetic algorithm and random forest", *Mechanical Systems and Signal Processing*, Vol. 70-71, pp. 87-103.

Chowdhury, S.R., Kumar, R., Kaur, R., Sharma, A. and Bhondekar, A.P. (2016), "Quality assessment of engine oil: an impedance spectroscopy based approach", 2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering, UPCON 2016, Institute of Electrical and Electronics Engineers Inc., pp. 608-612.

CIMAC (2008), *Guidelines for the Lubrication of Medium Speed Diesel Engines*, The International Council on Combustion Engines, CIMAC Recommendation No. 29, Frankfurt, Germany.

CIMAC (2011), *Used Engine Oil Analysis -User Interpretation Guide on Combustion Engines*, The International Council on Combustion Engines, CIMAC Recommendation No. 30, Frankfurt, Germany.

De Rivas, B.L., Vivancos, J.-L., Ordieres-Meré, J. and Capuz-Rizo, S.F. (2017), "Determination of the total acid number (TAN) of used mineral oils in aviation engines by FTIR using regression models", *Chemometrics and Intelligent Laboratory Systems*, Elsevier, Vol. 160 October 2016, pp. 32-39.

Demetgul, M. (2013), "Fault diagnosis on production systems with support vector machine and decision trees algorithms", *International Journal of Advanced Manufacturing Technology*, Vol. 67 Nos 9-12, pp. 2183-2194.

Discenzo, F., Chung, D., Kendig, M. and Loparo, K. (2006), "Intelligent fluid sensor for machinery diagnostics, prognostics, and control", United States Patent and Trademark Office, available at: https://www.google.com/patents/US7581434 (accessed 29 April 2017).

Eitrich, T. and Lang, B. (2006), "Efficient optimization of support vector machine learning parameters for unbalanced datasets", *Journal of Computational and Applied Mathematics*, Vol. 196 No. 2, pp. 425-436.

Fitch, J. (2007), "The truths about oil analysis data trending", Noria Publication on Machinery Lubrication, Tulsa, OK, available at: http://www.machinerylubrication.com/Read/1075/oil-analysis-data-trending (accessed 1 July 2017).

Flath, C.M. and Stein, N. (2018), "Towards a data science toolbox for industrial analytics applications", *Computers in Industry*, Vol. 94, pp. 16-25.

Janssens, O., Noppe, N., Devriendt, C. and Van De Walle, R. (2016), "Data-driven multivariate power curve modelling of offshore wind turbines", *Engineering Applications of Artificial Intelligence*, Vol. 55, pp. 331-338.

Jardine, A.K.S., Lin, D. and Banjevic, D. (2006), "A review on machinery diagnostics and prognostics implementing condition-based maintenance", *Mechanical Systems and Signal Processing*, Vol. 20, pp. 1483-1510.

Jolliffe, I.T. (2002), "Principal components analysis", *Springer Series in Statistics*, pp. 1-694.

Jun, H.B., Conte, F.L., Kiritsis, D. and Xirouchakis, P. (2008), "A predictive algorithm for estimating the quality of vehicle engine oil", *International Journal of Industrial Engineering: Theory Applications and Practice*, Vol. 15 No. 4, pp. 386-396.

Kocsis, M.C., Briggs, T. and Anderson, G. (2017), "The impact of lubricant volatility, viscosity and detergent chemistry on low-speed pre-ignition behavior", *SAE International Journal of Engines*, Vol. 10 No. 3, pp. 2017-010685.

Kumar, S. and Kumar, M. (2016), "Assessing remaining useful life of lubricant using Fourier transform infrared spectroscopy", *Journal of Quality in Maintenance Engineering*, Vol. 22 No. 2, pp. 202-214.

Kumar, A., Shankar, R. and Thakur, L.S. (2018), "A big data driven sustainable manufacturing framework for condition-based maintenance prediction", *Journal of Computational Science*, Vol. 27, pp. 428-439.

Lazakis, I., Raptodimos, Y. and Varelas, T. (2018), "Predicting ship machinery system condition through analytical reliability tools and artificial neural networks", *Ocean Engineering*, Vol. 152, November 2017, pp. 404-415.

Li, W., Zhu, Z., Jiang, F., Zhou, G. and Chen, G. (2014), "Fault diagnosis of rotating machinery with a novel statistical feature extraction and evaluation method", *Mechanical Systems and Signal Processing*, Vol. 50-51, pp. 414-426.

Lin, D., Banjevic, D. and Jardine, a K.S. (2006), "Using principal components in a proportional hazards model with applications in condition-based maintenance", *Journal of the Operational Research Society*, Vol. 57 No. 8, pp. 910-919.

Lu, P., Liu, H., Serratella, C. and Wang, X. (2017), "Assessment of data-driven, machine learning techniques for machinery prognostics of offshore assets", *Offshore Technology Conference, Offshore Technology Conference*, Houston, Texas, USA, pp. 1-22.

Martens, D., Vanthienen, J., Verbeke, W. and Baesens, B. (2011), "Performance of classification models from a user perspective", *Decision Support Systems*, North-Holland, Vol. 51 No. 4, pp. 782-793.

Moosavian, A., Ahmadi, H., Sakhaei, B. and Labbafi, R. (2014), "Support vector machine and K-nearest neighbour for unbalanced fault detection", *Journal of Quality in Maintenance Engineering*, Vol. 20 No. 1, pp. 65-75.

Muruganatham, B., Sanjith, M.A., Krishnakumar, B. and Satya Murty, S.A.V. (2013), "Roller element bearing fault diagnosis using singular spectrum analysis", *Mechanical Systems and Signal Processing*, Vol. 35 No. 1, pp. 150-166.

Nasridinov, A., Ihm, S.-Y. and Park, Y.-H. (2013), *A Decision Tree-Based Classification Model for Crime Prediction*, Springer, Dordrecht, pp. 531-538.

Paliwal, M. and Kumar, U.A. (2009), "Neural networks and statistical techniques: a review of applications", *Expert Systems with Applications*, Elsevier, Vol. 36 No. 1, pp. 2-17.

Pandya, D.H., Upadhyay, S.H. and Harsha, S.P. (2014), "Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform", *Soft Computing*, Vol. 18 No. 2, pp. 255-266.

Patel, R.K. and Giri, V.K. (2016), "Feature selection and classification of mechanical fault of an induction motor using random forest classifier", *Perspectives in Science*, Vol. 8, pp. 334-337.

Pereira, S., Baptista, M. and Henriques, E.M.P. (2018), "Data-driven quality prognostics for automated riveting processes", 2018 IEEE Aerospace Conference, IEEE, Big Sky, MT, pp. 1-15.

Phillips, J., Cripps, E., Lau, J.W. and Hodkiewicz, M.R. (2015), "Classifying machinery condition using oil samples and binary logistic regression", *Mechanical Systems and Signal Processing*, Elsevier, Vol. 60, pp. 316-325.

Raza, J. and Liyanage, J.P. (2009), "Application of the intelligent technique to identify hidden abnormalities in a system", *Journal of Quality in Maintenance Engineering*, Vol. 15 No. 2, pp. 221-235.

Raza, J., Liyanage, J.P., Al Atat, H. and Lee, J. (2010), "A comparative study of maintenance data classification based on neural networks, logistic regression and support vector machines", *Journal of Quality in Maintenance Engineering*, Vol. 16 No. 3, pp. 303-318.

Razali, N.M. and Wah, Y.B. (2011), "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests", *Journal of Statistical Modeling and Analytics*, Vol. 2 No. 1, pp. 21-33.

Safizadeh, M.S. and Latifi, S.K. (2014), "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell", *Information Fusion*, Vol. 18, pp. 1-8.

Samanta, B. (2004), "Mechanical Systems and Signal Processing Gear fault detection using artificial neural networks and support vector machines with genetic algorithms", *Mechanical Systems and Signal Processing*, Vol. 18, pp. 625-644.

Satishkumar, R. and Sugumaran, V. (2016), "Vibration-based health assessment of bearings using random forest classifier", *Indian Journal of Science and Technology*, Vol. 9 No. 10, doi: 10.17485/ijst/2016/v9i10/85303.

Scornet, E. (2016), "On the asymptotics of random forests", *Journal of Multivariate Analysis*, Elsevier, Vol. 146, pp. 72-83.

Seera, M., Wong, M.L.D. and Nandi, A.K. (2017), "Classification of ball bearing faults using a hybrid intelligent model", *Applied Soft Computing Journal*, Vol. 57, pp. 427-435.

Semiz, B., Hersek, S., Pouyan, M.B., Partida, C., Blazquez, L., Selby, V., Wieselthaler, G., Omer, I. and Liviu, K. (2018), "Use of ventricular assist device acoustical signatures to detect device thrombosis" *Journal of Cardiac Failure*, Vol. 24 No. 8, p. S19.

Shafi, U., Safi, A., Shahid, A.R., Ziauddin, S. and Saleem, M.Q. (2018), "Vehicle remote health monitoring and prognostic maintenance system", *Journal of Advanced Transportation*, Hindawi, Vol. 2018, pp. 1-10.

Sharma, S. (1996), *Applied Multivariate Techniques*, Kent, T. and Sellers, P. (Eds), John Wiley & Sons, New York, NY.

Suzuki, R. and Shimodaira, H. (2006), "Pvclust: an R package for assessing the uncertainty in hierarchical clustering", *Bioinformatics*, Vol. 22 No. 12, pp. 1540-1542.

Tabachnick, B.G. and Fidell, L.S. (2007), *Using Multivariate Statistics*, Bacon, P.A. (Ed.), 5th ed., Upper Saddle River, NJ.

Tomar, D. and Agarwal, S. (2013), "A survey on data mining approaches in healthcare", *International Journal of Bio-Science and Bio-Technology*, Vol. 5 No. 5, pp. 241-266.

Vališ, D., Žák, L. and Pokora, O. (2016a), "System condition estimation based on selected tribodiagnostic data", *Quality and Reliability Engineering International*, Vol. 32 No. 2, pp. 635-645.

Vališ, D., Žák, L., Pokora, O. and Lánský, P. (2016b), "Perspective analysis outcomes of selected tribodiagnostic data used as input for condition-based maintenance", *Reliability Engineering and System Safety*, Vol. 145, pp. 231-242.

Wakiru, J., Pintelon, L., Chemweno, P. and Muchiri, P. (2017a), "Analysis of lubrication oil contamination by fuel dilution with the application of cluster analysis", *XVII International Scientific Conference on Industrial Systems*, pp. 252-257.

Wakiru, J., Pintelon, L., Chemweno, P. and Muchiri, P. (2017b), "A lubricant condition monitoring approach for maintenance decision support - a data exploratory case study", *2nd Maintenance Forum 2017*, pp. 69-82.

Wakiru, J., Pintelon, L., Chemweno, P. and Muchiri, P.N. (2018), "A decision tree-based classification framework for used oil analysis applying random forest feature selection", *Journal of Applied Sciences, Engineering and Technology for Development*, Vol. 3 No. 1, pp. 90-100.

Wakiru, J., Pintelon, L., Muchiri, P.N. and Chemweno, P.K. (2019), "A review on lubricant condition monitoring information analysis for maintenance decision support", *Mechanical Systems and Signal Processing*, Elsevier, Vol. 118, pp. 108-132.

Wang, W. and Hussin, B. (2009), "Plant residual time modelling based on observed variables in oil samples", *Journal of the Operational Research Society*, Vol. 60 No. 6, pp. 789-796.

Xia, Y., Liu, C., Li, Y. and Liu, N. (2017), "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring", *Expert Systems with Applications*, Elsevier, Vol. 78, pp. 225-241.

Yang, B.-S., Hwang, W.-W., Kim, D.-J. and Chit Tan, A. (2005), "Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines", *Mechanical Systems and Signal Processing*, Vol. 19, pp. 371-390.

Yu, J.B. (2011), "Bearing performance degradation assessment using locality preserving projections", *Expert Systems with Applications*, Elsevier, Vol. 38 No. 6, pp. 7440-7450.

Zhang, C., Liu, C., Zhang, X. and Almpanidis, G. (2017), "An up-to-date comparison of state-of-the-art classification algorithms", *Expert Systems with Applications*, Elsevier, Vol. 82, pp. 128-150.

Zhao, Y., Vergopolan, N., Baylis, K., Blekking, J., Caylor, K., Evans, T., Giroux, S., Justin, S. and Lyndon, E. (2018), "Comparing empirical and survey-based yield forecasts in a dryland agro-ecosystem", *Agricultural and Forest Meteorology*, Vol. 262, pp. 147-156.

Zhu, X., Zhong, C. and Zhe, J. (2017), "Lubricating oil conditioning sensors for online machine health monitoring – a review", *Tribology International*, Elsevier, Vol. 109, pp. 473-484.

Corresponding author
James Wakiru can be contacted at: jamesmutuota.wakiru@kuleuven.be