An Intelligent Predictive Maintenance Framework with Autonomous Feature Selection based on Hybrid Fuzzy Set and Rough Set Theories for Multiclass Fault Classification*

[1]A. Buabeng, [1]A. Simons, [2]N. K. Frempong, [1]Y. Y. Ziggah
[1]University of Mines and Technology, Tarkwa, Ghana
[2]Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Buabeng, A., Simons, A. Frempong, N. K. and Ziggah, Y. Y. (2022), "An Intelligent Predictive Maintenance Framework with Autonomous Feature Selection based on Hybrid Fuzzy and Rough Set Theories for Multiclass Fault Classification", *Ghana Journal of Technology,* Vol. 6, No. 2, pp. 9 - 19.

Abstract

In this paper, an intelligent Predictive Maintenance (PdM) framework with an efficient and automated selection of relevant and most informative features has been proposed for fault classification. This was achieved by using the hybrid of Fuzzy Set and Rough Set Theories as a feature selection technique for pre-processing and the selection of features that contained only relevant fault characteristics. The selected features were then served as input for training the Support Vector Machine (SVM) classifier for the classification of the condition of four major hydraulic components (accumulator, cooler, pump and valve). To ascertain the performance of the proposed framework, a comparative study with five different and well-established machine learning classifiers was evaluated using nine different performance metrics. The result from the analysis proves the versatility of the proposed framework in classifying the various conditions of the hydraulic components whiles reducing the computational cost. When compared with prior works, a significant average improvement of over 26% in test accuracy was obtained for both accumulator and pump conditions whiles similar results were seen for cooler and valve conditions.

Keywords: Fuzzy Rough Feature Selection, Hydraulic System, Ordered Weighted Average, SVM

1 Introduction

One of the major challenges encountered in the field of computational intelligence is the selection of relevant and most informative features from a pool of knowledge that contains an optimum level of variability for predicting a given outcome. The challenge is intensified in the field of Predictive Maintenance (PdM) which is characterised by a high dimensional dataset (in the order of tens of thousands) generated from different process sensors at varying sampling rates. The usage of the high dimensional dataset for inferential purposes or to develop PdM frameworks often proves to be challenging. This is because the associated computational requirements (cost and storage) increase exponentially as the data size increases (Guha *et al*., 2022). Additionally, selecting a representative subset is practically impossible since there exist combinatorially many feature combinations. That is, $N!\left[n!(N-n)!\right]$ with n features and from a collection of N total features (Jensen and Shen, 2009). Ideally, it is acknowledged that the inclusion of more features for the development of a PdM framework is expected to increase the chances of providing enough information for distinguishing between class outputs. However, this assumption does not always hold since the inclusion of too many features may increase the level of uncertainties (noise), thus leading to spurious outcomes (Jensen and Shen, 2009). Likewise, the usage of many or little features (not optimal) for the design and development of a PdM framework subsequently leads to low efficiency and high computational cost (Abdulhafiz and Khamis, 2013).

In literature, enormous research has been directed towards the devising of approaches for the selection of a smaller subset from a pool of knowledge that is representative enough for a given task. This concerns the domain of dimension reduction which aims to produce a representative subset. Feature selection, unlike other dimensionality reduction techniques, preserves the underlying structure of the resulting feature subset. Hence, its application in numerous disciplines such as image processing (Bruzzone and Persello, 2009), text (Bharti and Singh, 2015), data classification (Mohapatra and Chakravarty, 2015; Schneider *et al.*, 2017) among others.

Despite the promising solution offered by feature selectors, recent advancements in science have led to the development of techniques for addressing the uncertainties and imprecision in data recorded for the development of intelligent models. Notable among these techniques are in the domain of Fuzzy Set Theory (FST) (Zadeh, 1965) and the Rough Set Theory (RST) (Pawlak, 2012). The FST utilises the concept of partial set membership which facilitate reasoning in an imprecise manner (Zadeh, 1965) thus, making FST very efficient for addressing the imprecision of datasets. However, FST is limited in application as it requires prior knowledge of

**Manuscript received December 14, 2021*
Revised version accepted March 18, 2022

membership functions for various fuzzy sets. This limitation is a significant drawback in FST as any decision made by the user may possibly be faulty or based on their subjective judgement. The RST on the other hand utilises the internal structure (granularity) to exploit the facts hidden in a given crisp or discrete datasets (Pawlak, 2012). In addition, RST requires no prior knowledge for computation and can be used as a feature selection technique for finding a representative subset from a pool of knowledge. Hence, its successful application in diverse fields (Huang, 1992). However, since RST was originally proposed for addressing uncertainties in crisp or discrete datasets, it cannot be implemented on real-valued datasets characterised with uncertainties (Pawlak, 2012; Riza *et al.*, 2014), which is typical in the field of predictive maintenance.

Despite the shortfalls in both FST and RST, the two domains are analogous as they seek to address uncertainty with the only difference being the type of uncertainty addressed and the procedure involved. Therefore, exploiting or hybridising the strength of these domains, Fuzzy-Rough Set Feature Selection (FRFS) will provide a high degree of flexibility for addressing several data imperfections, intelligently extract the most informative feature subset in both discrete and real-valued industrial data characterised by uncertainties for the development of PdM frameworks without the need for user-supplied input. The resulting feature subset from FRFS can then serve as input for training supervised machine learners.

To validate the proposed FRFS approach in PdM, this paper emphasises on a hydraulic system that was presented by Helwig *et al*. (2015a) as a test benchmark in the field of monitoring hydraulic systems. The work of Helwig *et al*. (2015a) employed the scatter-based feature extraction and Pearson's correlation-based selection techniques for selecting the 20 most highly correlated features for classification. The 20 selected features were served as input for the classification of accumulator, cooler, pump and valve conditions. The work of Helwig *et al*. (2015a) was extended by Helwig *et al*. (2015b) to allow for the detection of typical faults and the compensation of sensor faults in the hydraulic system. Similar to the approach used by Helwig *et al*. (2015a), the study utilised the same feature extraction and selection techniques for selecting features for classification. To further explore the research area, Schneider *et al*. (2017) automated the reduction in dimension by applying four complimentary feature extraction methods and three features selection algorithms. The feature selection methods were Pearson's correlation-based, the Recursive Feature Elimination Support Vector Machines (RFESVM) and the Univariate Relief feature selection. The approach improved the classification of accumulator conditions while the others (cooler, pump and valve) were comparatively similar to the results obtained by Helwig *et al*. (2015a).

Along the line, Chawathe (2019) also sought to improve the classification accuracy of prior research works by investigating a trade-off between the number of features and accuracy. This was realised through the extraction of 68 features based on some distribution functions. The resulting set of features was then selected and grouped into four different categories based on feature-selection-free (all 68 features), One Rule (OneR), J48 Decision Tree (DT) and Information gain. The four categories of features were then used as input for training seven (7) different classifiers. Although some classifiers resulted in higher accuracies than prior research works, others also performed worse. Furthermore, Quatrini *et al*. (2020) also developed a predictive model for the degradation of major components of the hydraulic system by extracting 102 relevant features. Using Pearson's correlation selection criteria, the 102 extracted features were grouped into three (3) categories of inputs based on feature-selection-free (all 102 features), 60 and 80 features. These categories of inputs were then used for training five Machine Learning (ML) classifiers. The approach achieved high accuracies as the reduction in features allowed for optimal combination of inputs that linearly separates the groups while minimising inter-group distance. Finally, König and Helmi (2020) predicted the different levels of degradation of major components of the hydraulic system. As opposed to the feature extraction and selection techniques adopted by prior works, the study used the deep learning technique of Convolutional Neural Networks (CNN). Hence, the developed deep learning-based condition monitoring system for the hydraulic system was achieved without explicitly engineering the features. Although the techniques utilised in prior works have resulted in satisfactory results along the line, the feature selection approaches such as the Pearson's correlation and others adopted are subjective because they require some level of input from the user. These user inputs and decisions which are based on subjective judgement may be faulty, leading to suboptimal model specification. CNN on the other hand is noted to be highly expensive in terms of computation, thus, requires high hardware resources such as memory and processors.

In view of the foregoing discussion, the paper proposes a PdM framework for the efficient classification of fault conditions characterised by uncertainties based on the features selected through hybrid fuzzy rough set theories. The proposed framework advances the field of PdM by enhancing the feature selection technique used for optimal selection of relevant and most informative features from a pool of industrial datasets without the need for user-supplied input which is predominantly subjective and may lead to model inefficiencies.

2 Resources and Methods Used

2.1 Hydraulic System Dataset

The utilisation of hydraulic systems in most engineering setups is very crucial due to their numerous industrial applications such as manufacturing, transportation, and machinery. With the recent advancement in science and today's highly competitive environment, the role of hydraulic systems in complex and integrated machinery is in high demand and are generally operated in extreme and challenging conditions. These dynamic conditions expose the hydraulic systems to progressive deterioration which impact the availability and reliability states of equipment, increase operational and maintenance related cost (Sheng *et al.*, 2011). For these reasons, a robust condition monitoring of hydraulic systems as a requirement for predictive-based maintenance is imperative since it can improve productivity, reduce the maintenance cost of industrial equipment and offer a significant improvement in their health condition. This, ultimately will increase asset utilisation and prevent them from deteriorating, considering that reliability and high safety standards are of paramount importance.

Considering the rising demand and increasing complexity of industrial systems, the number of installed sensors and their sampling rate are constantly growing (Schneider *et al.*, 2018). As such, the processing of high-dimensional data (signals) from industrial equipment to detect a fault or monitor the condition of a component or the entire system is usually not a common place for human operators. For instance, a fault in a single hydraulic component, say pump, can affect the normal operational performance of the entire hydraulic system (Chawathe, 2019; König and Helmi, 2020). This calls for the development of automated and robust frameworks capable of predicting the degradation states of the hydraulic system as well as the specific components.

In this paper, emphasis is placed on a hydraulic system which was presented by Helwig *et al.* (2015a) as a test benchmark in the field of monitoring hydraulic systems. The hydraulic system dataset recorded from monitoring the condition of the hydraulic system is publicly available at the University of California, Irvine (UCI) Machine Learning repository. The dataset can be accessed via http://archive.ics.uci.edu/ml/datasets/Condition+monitoring+of+hydraulic+systems. The hydraulic system dataset utilised in this paper consists of 2 205 instances and 43 680 features recorded from 17 process sensors with varying sampling rates. The 17 processes comprised of six pressure sensors, four temperature sensors, two volume flow sensors, five different sensors for recording motor power, vibration, cooling efficiency, cooling power and system efficiency. A detailed discussion of the 17 process sensors can be found at the source (Helwig *et al.*, 2015a). The dataset also contains fault scenarios depicting the variations in fault conditions of four major components such as hydraulic accumulator, cooler, internal pump leakage and valve. The accumulator is designed for storing energy, absorbing shocks and pulsations, the cooler for ensuring optimal oil temperature, the pump for distributing fluids, and the valve for regulating the flow of fluids in the hydraulic system. The details of the four major components are shown in Table 1.

2.2 Feature Engineering

To develop an efficient and robust Predictive Maintenance (PdM) framework, one major issue that needs to be addressed is the dimension of features (i.e., 43 680 features from the 17 process sensors), which is relatively high. Consequently, conventional Machine Learning (ML) techniques will suffer from tractability, scalability, high time complexities and more importantly classification performance issues (Herrmann *et al.*, 2012; Bach, 2017). Also, feeding the dataset directly into conventional ML techniques fails to detect features that contain the most characteristic fault information required for the efficient classification of the fault conditions (Chawathe, 2019). Hence, a well-established strategy is to extract some statistics-based features or transforms that represent the characteristic properties of the hydraulic dataset from the 43 680 features.

2.2.1 Feature Extraction

Seven different statistical time-domain features such as the mean, median, variance, standard deviation, skewness, kurtosis and position of maximum were extracted from the hydraulic system dataset. To ensure uniformity, these statistical estimates were obtained after partitioning each sensor into various time intervals as follows: the six Pressure Sensor (PS): PS1 (13), PS2 (14), PS3 (18), PS4 (25), PS5 (19), PS6 (19); the four Temperature Sensor (TS): TS1 (7), TS2 (7), TS3 (8), TS4 (15); volume Flow Sensor (FS): FS1 (18), FS2 (18); motor Efficiency Power Sensor (EPS): EPS1 (13), Vibration Sensor (VS) 15; Cooling Efficiency (CE) sensor 13; Cooling Power (CP) sensor 13; System Efficiency (SE) sensor 23. This resulted in a complete feature vector of 1 806 features.

However, these 1 806 extracted features may contain insignificant and redundant features with uncertainties which when used may negatively affect the performance of the proposed PdM framework to discriminate between class output effectively and efficiently (Binsaeid *et al.*, 2009). Hence, the Fuzzy Rough Set Feature Selection (FRFS) is utilised in this paper to determine the most optimal feature subsets that capture relevant fault characteristics. The FRFS technique, as opposed to the correlation-based feature selection technique adopted by prior works such as Helwig *et al.* (2015a) and Helwig *et al.* (2015b), does not require prior knowledge or the user to subjectively specify the number of features to select as wrong judgement may influence the efficacy of the model being developed.

2.2.2 Fuzzy Rough Set Theory

The fuzzy rough set theory can be thought of as a generalisation to the rough set theory where the approximation from fuzzy sets; lower and upper fuzzy approximations, is derived in a crisp rough set space. That is, fuzziness is integrated into rough sets by defining the lower and upper approximations of the set when U, the non-empty set of finite fuzzy sets becomes rough because of the equivalence relation.

Suppose that the subsets of features from A, $P \subseteq A$, with equivalence relation over U denoted as $\mathrm{IND}(P)$, the equivalence class can be expressed as fuzzy sets $F = \{F_1, F_2, \ldots, F_h\}$ if the class to which F_i for all $i \in \{1, 2, \ldots, h\}$ attribute are ambiguous. Thus, the fuzzy P- lower and P- upper approximations of X are expressed as Equations (1) and (2).

$$\mu_{\underline{P}X}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad \forall i \qquad (1)$$

$$\mu_{\overline{P}X}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i \qquad (2)$$

where X is the fuzzy concept to be approximated with x denoting an item in X and F_i being the fuzzy equivalence class belonging to $\mathrm{IND}(P)$. The ordered pair $\left(\underline{P}X, \overline{P}X\right)$ is the fuzzy rough set.

These definitions deviate a bit from the lower and upper approximations under the crisp rough set due to the inability to explicitly access the membership of individual objects to the approximations. As a result, the fuzzy lower and upper approximations are redefined by employing the concept of $\sup$ and $\inf$ as shown in Equations (3) and (4).

$$\mu_{\underline{P}X}(x) = \sup_{F \in U/\mathrm{IND}(P)} \times \min\left(\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}\right) \qquad (3)$$

$$\mu_{\overline{P}X}(F_i) = \sup_{F \in U/\mathrm{IND}(P)} \times \min\left(\mu_F(x), \sup_{y \in U} \min\{\mu_F(y), \mu_X(y)\}\right) \qquad (4)$$

It can be observed from Equations (3) and (4) that every $y \in U$ is taken into account with instances where their corresponding $\mu_F(y) \neq 0$. A detailed discussion on the usage of the $\min$ and $\max$ operators is seen in Radzikowska and Kerre (2002), where a comparative study of fuzzy rough sets is represented by specific implication and t-norm.

Table 1 Details of Hydraulic Components

Monitored Condition	Unit	States	Class Output	Cases
Accumulator	bar	Optimal Pressure	130	599
		Slightly Reduced Pressure	115	399
		Severely Reduced Pressure	100	399
		Close to Total Failure	90	808
Cooler	%	Full Efficiency	100	741
		Reduced Efficiency	20	732
		Close to Total Failure	3	732
Pump	-	No Leakage	0	1221
		Weak Leakage	1	492
		Severe Leakage	2	492
Valve	%	Optimal Switching Behaviour	100	1125
		Small Lag	90	360
		Severe Lag	80	360
		Close to Total Failure	73	360

2.2.2 Fuzzy Rough Set Feature Selection (FRFS)

The feature selection or reduction ability of rough set theory is perhaps a significant factor owning to its successful application in diverse disciplines. This unique ability can be exploited in the fuzzy rough set via the concept of fuzzy lower approximation for reducing datasets of real-valued features. Referring to the extension principle of Zadeh (1978), the membership of an object $x \in U$ belongs to a fuzzy positive region as expressed in Equation (5).

$$\mu_{\mathrm{POS}_P(Q)}(x) = \sup_{X \in U/\mathrm{IND}(Q)} \mu_{\underline{P}X}(x) \tag{5}$$

From Equation (5), it can be deduced that the object x fails to belong to the positive region if the equivalence class it belongs to is not a member of the positive region. The fuzzy rough dependency degree function is expressed from the definition of the positive region as shown in Equation (6).

$$\gamma'_P(Q) = \frac{\left|\mu_{\mathrm{POS}_P(Q)}(x)\right|}{|U|} = \frac{\sum_{x\in U} \mu_{\mathrm{POS}_P(Q)}(x)}{|U|} \tag{6}$$

However, for the fuzzy rough feature reduction to be useful in practice, it should be capable of handling high-dimensional datasets by means of estimating the dependencies of various feature subsets with the original dataset. This is relevant as the objects may belong to several equivalence classes. For instance, in the crisp case $U/\mathrm{IND}(P)$ consist of groups of objects that are indiscernible based on the features from P. However, in the case of fuzzy, the cartesian product of features from P is considered in estimating $U/\mathrm{IND}(P)$ where each set in $U/\mathrm{IND}(P)$ is an equivalence class. Hence, the extent to which an object belongs to the equivalence class could be estimated using the combination of constituent fuzzy equivalence classes, $F_i, \forall\, i = 1,2,..,n$ as shown in Equation (7).

$$\mu_{F_1\cap\ldots\cap F_n}(x) = \min\left(\mu_{F_1}(x), \mu_{F_2}(x), \ldots, \mu_{F_n}(x)\right) \tag{7}$$

However, despite the promising results that the classical FRFS techniques offer in the selection of an optimal feature subset, it possesses some deficiencies that could be minimised by incorporating an Ordered Weighted Averaging (OWA) operators in the calculation of upper and lower approximations for increased robustness (Cornelis *et al.*, 2010).

FRFS Based on Ordered Weighted Average (OWA)

The OWA approach to FRFS was proposed by Cornelis *et al*. (2010) and follows the principle of addressing sensitivity to noise and outlying samples. However, for the $\mathrm{FRFS}_{(\mathrm{OWA})}$, an aggregation technique using OWA estimators are employed for estimating the lower and upper approximations as shown in Equations (8) and (9) respectively.

$$\mu_{\underline{R_P}X}(x) = OWA_{W_l}\left\langle I\left\{\mu_{R_P}(x,y),\ \mu_X(y)\right\}\right\rangle \tag{8}$$

$$\mu_{\overline{R_P}X}(x) = OWA_{W_u}\left\langle T\left\{\mu_{R_P}(x,y),\ \mu_X(y)\right\}\right\rangle \tag{9}$$

where $OWA_W(X) = \sum_{i=1}^{n} w_i c_i$, c_i denotes the i^{th} largest value in X and $W = \langle w_i \rangle$ is the weighting vector such that for $m < n$ $\forall\, i = 1,2,\ldots,n$, it is possible to define $W_l = \langle w_i^l \rangle = W^{\min}$ and $W_u = \langle w_i^u \rangle = W^{\max}$ as Equations (10) and (11) respectively.

$$w_{n+1-i}^{l} = \begin{cases} \dfrac{2^{m-i}}{2^{m-1}}, & i = 1,2,\ldots,m \\ 0 & i = m+1,\ldots,n \end{cases} \tag{10}$$

$$w_{i}^{u} = \begin{cases} \dfrac{2^{m-i}}{2^{m-1}}, & i = 1,2,\ldots,m \\ 0 & i = m+1,\ldots,n \end{cases} \tag{11}$$

2.3 Support Vector Machine (SVM)

The SVM is one of the commonly used supervised learners in the field of predictive maintenance. Compared to other learners, it is known for high classification accuracy and efficiency in obtaining global optimum, robust performance and more importantly, high overfitting avoidance capability. As a result, SVM is a highly preferred choice for various classification tasks (Di *et al.*, 2019; Mahmud *et al.*, 2017; Gao *et al.*, 2020).

Given a matrix $X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{n\times p}$ with a corresponding response vector $y_i \in \{-1,+1\}$, where -1 and +1 denote samples from the negative and positive classes respectively. In SVM computation, the optimal hyperplane is defined as Equation (12).

$$w^T \phi(x) + b = 0 \tag{12}$$

where $\phi(x)$ represents a nonlinear mapping function for transforming x to a high-dimensional space, w and b are the weight vector and bias respectively. Thus, the objective of training SVM is to find w and b such that hyperplane divides the classes into distinct partitions and with the largest separation of classes. This problem can be formulated as Equation (13) subject to the constraint (Equation (14))

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{n} \xi_i \tag{13}$$

$$y_i\left(w^T w\phi(x_i)+b\right)\geq 1-\xi_i \quad (14)$$

where $\xi_i \ \forall\ i=1,2,\ldots,n$ are slack variables. Hence, the hyperplane with the maximum margin for separating the classes is realised. However, in situations where the maximum margin fails in finding the optimal separation of the hyperplane, soft margins are utilised such that w and b satisfy the inequalities (Equations (15) to (17)) for all elements in the data.

$$wx_i - b \geq +1-\xi_i \ \text{ if } \ y_i = +1 \quad (15)$$

$$wx_i - b \geq -1+\xi_i \ \text{ if } \ y_i = -1 \quad (16)$$

$$\xi \geq 0 \quad (17)$$

When an error occurs, $\xi_i > 1$ and $\sum_i \xi_i$ become the upper bound on the training error which is controlled by the Lagrangian L_p shown in Equation (18).

$$\begin{aligned} L_p &= \frac{1}{2}\left\|w^2\right\| + C\sum_{i=1}^{n}\xi_i \\ &\quad -\sum_i \alpha_i\left(y_i\left(x_i w - b\right)-1+\xi_i\right)-\sum_i \mu_i\xi_i \end{aligned} \quad (18)$$

where μ_i is the Lagrange multipliers for estimating the positive values of ξ_i.

It is important to note that, the optimal separation of the hyperplane is achieved via a kernel function (i.e., linear, radial basis, polynomial, sigmoid etc.). In this paper, the frequently used Radial Basis Function (RBF) kernel (Du *et al.*, 2016; Wang *et al.*, 2018) which is known for its excellent general performance, wider convergence domain, high-resolution power and requires fewer parameters was adopted. The RBF kernel function is expressed in Equation (19).

$$K_{RBF}\left(x_i, x_j\right) = \exp\left(-\gamma\left\|x-y\right\|^2\right) \quad (19)$$

2.3 Classification Performance Evaluation

To check the reliability of the classifiers, nine evaluation metrics namely Accuracy, Error rate, Sensitivity, Specificity, Precision, F score, Mathews Correlation Coefficient (MCC), Geometric Mean and Area Under Curve (AUC) were used, and are defined as Equations (20) - (28). Here, True Positive (TP) is the number of correct classification counts when there is a fault condition, True Negative (TN) is the number of correct classification counts when there is no fault condition, False Positive (FP) is the number of misclassification counts when there is a fault condition and False Negative (FN) is the number of misclassification counts when there is no fault condition. The performance metrics ranges between 0 and 1. Obtaining a value closer to 1 for these evaluation metrics, except for the error rate (Equation (21)), denotes higher classification performance which is preferable.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (20)$$

$$Error\ Rate = \frac{FP+FN}{TP+FP+TN+FN} \quad (21)$$

$$Precision = \frac{TP}{TP+FP} \quad (22)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (23)$$

$$Specificity = \frac{TN}{TN+FP} \quad (24)$$

$$FScore = \frac{2(Precision * Recall)}{Precision + Recall} \quad (25)$$

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{\begin{array}{l}(TP+FP)(TP+FN)\\ \times(TN+FP)(TN+FN)\end{array}}} \quad (26)$$

$$GM = \sqrt{\frac{TP}{TP+FN} * \frac{TN}{TN+FP}} \quad (27)$$

$$AUC = \frac{1}{N_+ N_-}\sum_{i=1}^{N_+}\sum_{j=1}^{N_-} I(x_i > y_j) \quad (28)$$

where N_+ and N_- is the number of positive and negative instances respectively. $x_i\ (i=1,2,\ldots,N_+)$ are the scores predicted by the model for the N_+ whilst $y_j\ (j=1,2,\ldots,N_-)$ is the scores predicted by the model for the N_-. $I(\bullet)$ is an indicator function satisfying the condition $I(true)=1$ and $I(false)=0$.

3 Results and Discussion

3.1 Fuzzy Rough Set Feature Selection Results

Table 2 shows the number of optimal features selected from the extracted 1 806 statistical time-domain features for the various monitored conditions based on the $FRFS_{(OWA)}$. As observed, feature(s) ranging from 5 to 27 were deemed relevant to have contained the required variability in classifying the monitored conditions. As opposed to prior works (Helwig *et al.*, 2015a; Helwig *et al.*, 2015b; Schneider, 2017), where 20 features were selected using correlation-based feature selection, Table 2 shows that an average of 15 features would have been optimal for training the ML classifiers for classifying the various fault types for each monitored condition.

Table 2 Selected Feature(s) from FRFS Methods

Monitored Condition	Selected Features	
	No.	**ID**
Accumulator	27	F4, F5, F6, F7, F15, F23, F27, F33, F36, F38, F47, F50, F54, F63, F65, F68, F74, F75, F77, F80, F85, F87, F88, F97, F98, F99, F100
Cooler	8	F381, F362, F363, F373, F356, F383, F139, F323
Pump	21	F1030, F1007, F779, F780, F781, F408, F388, F85, F927, F591, F118, F189, F190, F191, F186, F123, F629, F212, F586, F666, F950
Valve	5	F1085, F1146, F1091, F962, F599
Average	**15**	

In this paper, the selected features from $FRFS_{(OWA)}$ served as input for training the considered ML classifiers including SVM, Multi-Layer Perceptron (MLP), k-Nearest Neighbour (k-NN), C4.5, Linear Discriminant Analysis (LDA) and the Logistic Regression (LR). However, before training, the selected features together with their corresponding targets were partitioned into training (70%) and testing (30%) sets. The training set is used for developing the model whiles the testing set is used for validation. Also, due to the stochastic learning characteristics of these classifiers in producing slightly different output after each run, 10 runs of training are computed, and the final test output is presented as the average from the 10 runs. Tables 3 to 6 show the test performance of the selected features based on $FRFS_{(OWA)}$ being used as inputs for classifying the four monitored conditions (accumulator, cooler, pump and valve).

3.2 Classification of Fault Conditions

Table 3 shows the test performance of the $FRFS_{(OWA)}$-based selected features in classifying accumulator conditions. As observed, although the competing classifiers produced satisfactory results above 75% accuracy, SVM and MLP yielded higher classification accuracies and AUC scores of 88.97% (with 11.03% misclassification rates) and 98% respectively. For the remaining metrics, the proposed $FRFS_{(OWA)}$-SVM achieved better results such that when there was a specific fault condition, 87.02% was positively classified (sensitivity) whilst in the case of a different fault condition, only 3.54% were positively classified (96.46% specificity). $FRFS_{(OWA)}$-SVM was 86.68% precise in positively classifying the various fault conditions and showed a strong positive relation (MCC of 0.8329). Also, an aggregate of the overall performance of $FRFS_{(OWA)}$-SVM showed the highest of 0.8683 and 0.9162 for F score and Geometric Mean (GM) respectively. Similar but relatively lesser observations were seen in $FRFS_{(OWA)}$-MLP. This suggests that, based on the 27 features selected with the $FRFS_{(OWA)}$ technique (refer to Table 2), SVM and MLP would be adequate in classifying the conditions of the hydraulic accumulator. Even though the classification of the accumulator is known to be the most challenging among the four monitored hydraulic components (Chawathe, 2019; König and Helmi, 2020), the proposed framework shows a significant average test classification result of 83.36% as compared to the 55.43% obtained by Helwig *et al*. (2015a). This significant improvement could be attributed to the strength of $FRFS_{(OWA)}$ which automatically selects the relevant features without prior knowledge as opposed to the correlation-based approach used in Helwig *et al*. (2015a).

Apart from the accumulator, the next most complicated monitored hydraulic component to classify is the internal pump leakage (Chawathe, 2019; König and Helmi, 2020). Table 4 shows the classification results of the pump conditions using the 21 $FRFS_{(OWA)}$ based selected features (refer to Table 2). Similar to the classification results of the accumulator conditions, SVM and MLP achieved the highest classification test accuracies of 99.55%. KNN and C4.5 obtained relatively comparative classification results of 99.40% and 99.31% respectively. A critical look at the AUC scores in Table 4 suggests SVM and KNN are the most superior in the classification of the internal pump conditions. Nonetheless, MLP produces comparatively similar AUC results. Comparing the average classification test results to the prior work of Helwig *et al*. (2015a) shows a significant improvement from 72.55% to 98.85%. This could be attributed to the ability of $FRFS_{(OWA)}$ to select the most informative features whiles reducing uncertainties in the pump dataset.

Table 3 Classification Performance of FRFS Methods on Accumulator Conditions

Metric	SVM	MLP	KNN	C4.5	LDA	LR	Average
Accuracy	0.8897	0.8897	0.7689	0.8855	0.7961	0.7719	0.8336
Error	0.1103	0.1103	0.2311	0.1145	0.2039	0.2281	0.1664
Sensitivity	0.8702	0.8678	0.7308	0.8548	0.7438	0.7161	0.7972
Specificity	0.9646	0.9643	0.9234	0.9621	0.9323	0.9244	0.9452
Precision	0.8668	0.8677	0.7475	0.8628	0.7606	0.7330	0.8064
F Score	0.8683	0.8677	0.7328	0.8580	0.7466	0.7160	0.7982
MCC	0.8329	0.8320	0.6614	0.8217	0.6852	0.6489	0.7470
GM	0.9162	0.9148	0.8215	0.9069	0.8327	0.8136	0.5564
AUC	0.9800	0.9800	0.9290	0.9080	0.9370	0.9200	0.8676

Table 4 Classification Performance of FRFS Methods on Pump Conditions

Metric	SVM	MLP	KNN	C4.5	LDA	LR	Average
Accuracy	0.9955	0.9955	0.9940	0.9931	0.9728	0.9804	0.9885
Error	0.0045	0.0045	0.0060	0.0069	0.0272	0.0196	0.0115
Sensitivity	0.9932	0.9932	0.9910	0.9904	0.9595	0.9707	0.9830
Specificity	0.9976	0.9981	0.9974	0.9967	0.9874	0.9915	0.9948
Precision	0.9946	0.9934	0.9911	0.9905	0.9622	0.9715	0.9839
F Score	0.9939	0.9932	0.9910	0.9904	0.9606	0.9708	0.9833
MCC	0.9917	0.9914	0.9884	0.9872	0.9486	0.9626	0.9783
GM	0.9954	0.9956	0.9942	0.9936	0.9733	0.9810	0.9742
AUC	1.0000	0.9990	1.0000	0.9940	0.9930	0.9890	0.9889

Table 5 Classification Performance of FRFS Methods on Cooler Conditions

Metric	SVM	MLP	KNN	C4.5	LDA	LR	Average
Accuracy	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000
Error	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000
Specificity	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000
Precision	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000
F Score	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000
MCC	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	1.0000
GM	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
AUC	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6 Classification Performance of FRFS Methods on Valve Fault Classification

Metric	SVM	MLP	KNN	C4.5	LDA	LR	Average
Accuracy	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Error	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Specificity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Precision	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
F Score	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MCC	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
GM	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
AUC	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Unlike the complex nature of the accumulator and pump datasets, the complexity in classifying the cooler and valve conditions was less. This is ascertained by the lesser number of $FRFS_{(OWA)}$ based selected features (8 and 5 respectively) required to produce the perfect test classification results (100.00%) shown in Tables 5 and 6. Moreover, the same perfect test classification results were obtained by prior works of Helwig *et al*. (2015a).

However, the computational complexity (in time and storage) required in the classification of both cooler and valve conditions will be lesser when compared to the work of Helwig *et al*. (2015a). This is because of the lesser number of $FRFS_{(OWA)}$ based selected features of 8 and 5 for cooler and valve respectively, as opposed to the 20 features selected using correlation-based feature selection adopted by Helwig *et al*. (2015a). Due to the less complicated nature in the classification of both the cooler and valve conditions, almost all the classifiers, except LDA achieved 100.00% test results. Hence, any of these classifiers will be adequate for the classification of cooler and valve conditions.

4 Conclusions and Future Work

An intelligent and efficient Predictive Maintenance (PdM) framework based on the hybrid Fuzzy Rough Set Feature Selection (FRFS) has been proposed for multiclass fault classification. The $FRFS_{(OWA)}$ technique automated the PdM frameworks' ability to select the relevant and most informative features for the training of the SVM classifier. The proposed framework was verified on the classification of the condition of four major hydraulic components (accumulator, cooler, pump and valve). After comparing with five different and well-established classifiers, the SVM proved superior in classifying the hydraulic components under consideration in terms of test accuracies, AUC scores and seven other performance metrics. This demonstrates the versatility of the proposed framework in classifying the various conditions of the hydraulic components whiles reducing the computational complexity. Though similar results were seen for cooler and valve conditions, a significant average improvement of over 26% in test accuracy was obtained for both accumulator and pump conditions when compared with prior works. The proposed framework advances the field of PdM by enhancing the feature selection technique used for optimal selection of relevant and most informative features from a pool of industrial datasets without the need for user-supplied input which is predominantly subjective and may lead to model inefficiencies.

For future works, the study will further explore a comparative analysis of new and modified variants of the FRFS technique since the ability to produce optimal features may be task-specific. Also, the concept of using ensembles instead of standalone classifiers will be explored since a considerable number of resources such as time is being invested in the selection of a specific learner for a given task. The usage of ensembles will improve the performance of the PdM framework since the technique combines different sets of hypotheses from multiple learners.

References

Abdulhafiz, W. A. and Khamis, A. (2013), "Handling Data Uncertainty and Inconsistency Using Multisensor Data Fusion", *Advances in Artificial Intelligence*, Vol. 2013, pp. 1 - 11.

Bach, F. (2017), "Breaking the Curse of Dimensionality with Convex Neural Networks", *Journal of Machine Learning Research*, Vol. 18, No. 19, pp. 1 - 53.

Bharti, K. K. and Singh, P. K. (2015), "Hybrid Dimension Reduction by Integrating Feature Selection with Feature Extraction Method for Text Clustering", *Expert Systems with Applications*, Vol. 42, No. 6, pp. 3105 - 3114.

Binsaeid, S., Asfour, S., Cho, S. and Onar, A. (2009), "Machine Ensemble Approach for Simultaneous Detection of Transient and Gradual Abnormalities in End Milling using Multisensor Fusion", *Journal of Materials Processing Technology*, Vol. 209, No. 10, pp. 4728 - 4738.

Bruzzone, L. and Persello, C. (2009), "A Novel Approach to the Selection of Spatially Invariant Features for the Classification of Hyperspectral Images with Improved Generalization Capability", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 9, pp. 3180 - 3191.

Chawathe, S. S. (2019), "Condition Monitoring of Hydraulic Systems by Classifying Sensor Data Streams", *Proceedings of the IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, pp. 898 - 904.

Cornelis, C., Verbiest, N. and Jensen, R. (2010), "Ordered Weighted Average Based Fuzzy Rough Sets", In *Rough Set and Knowledge Technology (RSKT)*, Yu J., Greco S., Lingras P., Wang G., Skowron A. (eds), Springer, Berlin, Heidelberg, pp. 78 - 85.

Di, Z., Kang, Q., Peng, D. and Zhou, M. (2019), "Density Peak-Based Pre-Clustering Support Vector Machine for Multi-Class Imbalanced Classification", *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, pp. 27 - 32.

Du, D., Jia, X. and Hao, C. (2016), "A New Least Squares Support Vector Machines Ensemble Model for Aero Engine Performance Parameter Chaotic Prediction", *Mathematical Problems in Engineering*, Vol. 2016, pp. 1 - 8.

Gao, X., Wei, H., Li, T. and Yang, G. (2020), "A Rolling Bearing Fault Diagnosis Method based on LSSVM", *Advances in Mechanical Engineering*, Vol. 12, No. 1, pp. 1 - 10.

Guha, R., Chatterjee, B., Hassan, S., Ahmed, S., Bhattacharyya, T. and Sarkar, R. (2022), "Py_FS: A Python Package for Feature Selection Using Meta-Heuristic Optimization Algorithms", In: *Computational Intelligence in Pattern Recognitionpp*, Das A. K., Nayak, J., Naik, B., Dutta, S., Pelusi, D. (eds), Springer, Singapore, Vol. 1349, pp. 495 - 504.

Helwig, N., Pignanelli, E. and Schütze, A. (2015a), "Condition Monitoring of a Complex Hydraulic System using Multivariate Statistics", *Proceedings of the Instrumentation and Measurement Technology Conference (I2MTC)*, Pisa, Italy, pp. 210 - 215.

Helwig, N., Pignanelli, E. and Schütze, A. (2015b), "Detecting and Compensating Sensor Faults in a Hydraulic Condition Monitoring System", *Proceedings of Sensor*, Nürnberg, Germany, pp. 641 - 646.

Herrmann, F. J., Friedlander, M. P. and Yilmaz, O. (2012), "Fighting the Curse of Dimensionality: Compressive Sensing in Exploration Seismology", *IEEE Signal Process Mag.*, Vol. 29, No. 3, pp. 88 - 100.

Huang, S. Y. (1992), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Springer Science and Business Media, Germany, 2nd edition, 490 pp.

Jensen, R. and Shen, Q. (2009), "New Approaches to Fuzzy-Rough Feature Selection", *IEEE Transactions on Fuzzy Systems*, Vol. 17, No. 4, pp. 824 - 838.

König, C. and Helmi, A. M. (2020), "Sensitivity Analysis of Sensors in a Hydraulic Condition Monitoring System using CNN Models", *Sensors*, Vol. 20, No. 11, pp. 1 - 19.

Mahmud, M. N., Ibrahim, M. N., Osman, M. K. and Hussain, Z. (2017), "Support Vector Machine (SVM) for Fault Classification in Radial Distribution Network", *Advanced Science Letters*, Vol. 23, No. 5, pp. 4124 - 4128.

Mohapatra, P. and Chakravarty, S. (2015), "Modified PSO Based Feature Selection for Microarray Data Classification", *Proceedings of the 2015 IEEE Power, Communication and Information Technology Conference (PCITC)*, Bhubaneswar, India, pp. 703–709.

Pawlak, Z. (2012), *Rough Sets: Theoretical Aspects of Reasoning about Data*, Springer Science and Business Media, Berlin, Germany, Vol. 9, 252 pp.

Quatrini, E., Costantino, F., Pocci, C. and Tronci, M. (2020), "Predictive Model for the Degradation State of a Hydraulic System with Dimensionality Reduction", *Procedia Manufacturing*, Vol. 42, pp. 516 - 523.

Radzikowska, A. M. and Kerre, E. E. (2002), "A Comparative Study of Fuzzy Rough Sets", *Fuzzy Sets and Systems*, Vol. 126, No. 2, pp. 137 - 155.

Riza, L. S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Śle¸zak, D. and Benítez, J. M. (2014), "Implementing Algorithms of Rough Set Theory and Fuzzy Rough Set Theory in the R Package 'RoughSets'", *Information Sciences*, Vol. 287, pp. 68 - 89.

Schneider, T., Helwig, N. and Schütze, A. (2017), "Automatic Feature Extraction and Selection for Classification of Cyclical Time Series Data", *Tm-Technisches Messen*, Vol. 84, No. 3, pp. 198 - 206.

Schneider, T., Helwig, N. and Schütze, A. (2018), "Automatic Feature Extraction and Selection for Condition Monitoring and Related Datasets", *Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Houston, TX, USA, pp. 1 - 6.

Sheng, C., Li, Z., Qin, L., Guo, Z. and Zhang, Y. (2011), "Recent Progress on Mechanical Condition Monitoring and Fault Diagnosis", *Procedia Engineering*, Vol. 15, pp. 142 - 146.

Wang, K., Wen, X., Hou, D., Tu, D., Zhu, N., Huang, P., Zhang, G. and Zhang, H. (2018), "Application of Least-Squares Support Vector Machines for Quantitative Evaluation of Known Contaminant in Water Distribution System using Online Water Quality Parameters", *Sensors*, Vol. 18, No. 4, pp. 1 - 19.

Zadeh, L. A. (1965), "Information and Control", *Fuzzy Sets*, Vol. 8, No. 3, pp. 338 - 353.

Zadeh, L. A. (1978), "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, Vol. 1, No. 1, pp. 3 - 28.

Authors

A. Buabeng is a Lecturer at the Mathematical Sciences Department of the University of Mines and Technology (UMaT), Tarkwa, Ghana. He holds an MPhil in Mathematics (Statistics). He is currently pursuing his PhD in Mathematics. He is a member of Ghana Mathematics Society, Ghana Statistical Association, and the International Association of Engineers (IAENG). His research interests are in the application of Machine Learning, to Predictive Maintenance, Data Mining and Multivariate Prediction for Quality Control.

A. Simons is a Professor of Mechanical Engineering Department at the University of Mines and Technology, Tarkwa, Ghana. He holds the degrees of MSc from the Belarusian-Russian University, Mogilev, Belarus, PhD from St. Petersburg State Mining Institute (Technical University), St. Petersburg, Russia and NDT Level II from Trinity NDT College Bangalore, India. He is a member of America Society of Mechanical Engineers. and Ghana Institution of Engineering. His research and consultancy works cover Heat Transfer, Fuels and Internal Combusting Engines, Machine Design, Maintenance

Engineering, Accident Vehicle Assessment, Factory Technical Audit and Non-Destructive Testing (NDT).

N. K. Frempong is a Senior Lecturer at the Department of Statistics and Actuarial Science of the Kwame Nkrumah University of Science and Technology (KNUST). He holds two Master of Science degrees in Applied Statistics and Biostatistics from the University of Hasselt (Belgium). He obtained his PhD in Mathematical Statistics from KNUST. He is passionate about research in Public Health, Applied Statistics and Data Science.

Y. Y. Ziggah is a Lecturer at the Geomatic Engineering Department of the University of Mines and Technology (UMaT). He holds a BSc in Geomatic Engineering from Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. He obtained his Master of Engineering and Doctor of Philosophy degrees in Geodesy and Survey Engineering from China University of Geosciences (Wuhan), P. R. China. His research interests include artificial intelligent application in engineering, geodetic coordinate transformation, gravity field modelling, height systems and geodetic deformation modelling.