# Hybrid Intelligent Predictive Maintenance Model for Multiclass Fault Classification

**Albert Buabeng** ( abuabeng@umat.edu.gh )
University of Mines and Technology Faculty of Engineering https://orcid.org/0000-0001-5680-5615
**Anthony Simons**
University of Mines and Technology Faculty of Engineering
**Nana Kena Frimpong**
Kwame Nkrumah University of Science and Technology College of Science
**Yao Yevenyo Ziggah**
University of Mines and Technology Faculty of Mineral Resources Technology

**Research Article**

**Keywords:** Condition monitoring, Machine learning, Signal decomposition, Dimensionality reduction, Fault classification

**Posted Date:** July 6th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-600110/v1

# Hybrid Intelligent Predictive Maintenance Model for Multiclass Fault Classification

Albert Buabeng[1], Anthony Simons[2], Nana Kena Frempong[3], Yao Yevenyo Ziggah[4]

[1]Faculty of Engineering, Mathematical Sciences Department, University of Mines and Technology, Tarkwa, Ghana

[2]Faculty of Engineering, Mechanical Engineering Department, University of Mines and Technology, Tarkwa, Ghana

[3]College of Sciences, Statistics and Actuarial Science Department, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

[4]Faculty of Mineral Resources Technology, Geomatic Engineering Department, University of Mines and Technology, Tarkwa, Ghana

Corresponding Author:

Albert Buabeng

Faculty of Engineering

Mathematical Sciences Department

University of Mines and Technology

Tarkwa, Ghana

Phone: +233 545209247

ORCID Number: 0000-0001-5680-5615

Email: abuabeng@umat.edu.gh

Third Author ORCID Number: 0000-0002-7138-3526

Fourth Author ORCID Number: 0000-0002-9940-1845

**Abstract**

Data recorded from monitoring the health condition of industrial equipment are often high-dimensional, nonlinear, nonstationary and characterised by high levels of uncertainty. These factors limit the efficiency of machine learning techniques to produce desirable results when developing effective fault classification frameworks. This paper sought to propose a hybrid artificial intelligent predictive maintenance model based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), Principal Component Analysis (PCA) and Least Squares Support Vector Machine (LSSVM) optimised by the combination of Coupled Simulated Annealing and Nelder-Mead Simplex optimisation algorithms (ICEEMDAN-PCA-LSSVM). Here, ICEEMDAN was first employed as a denoising technique to decompose signals into series of Intrinsic Mode Functions (IMFs) of which only relevant IMFs containing the relevant fault features were retained for signal reconstruction. PCA was then employed as a dimension reduction technique through which the resulting set of uncorrelated features extracted served as input for LSSVM for classifying various fault types. The proposed technique is compared with three established methods (Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Artificial Neural Network (ANN)) with multiclass classification capabilities. The various techniques were tested on an experimental UCI machine learning benchmark data obtained from multi-sensors of a hydraulic test rig. The results from the analysis revealed that the proposed ICEEMDAN-PCA-LSSVM technique is versatile and outperformed all the compared classifiers in terms of accuracy, error rate and other evaluation metrics considered. The proposed hybrid technique drastically reduced the redundancies and the dimension of features, allowing for the efficient consideration of relevant features for the enhancement of classification accuracy and convergence speed.

**Keywords:** Condition monitoring, Machine learning, Signal decomposition, Dimensionality reduction, Fault classification

## 1 Introduction

As modern industrial companies strive to meet their operational targets in order to remain successful in profit maximisation, they are pressured to make use of various integrated and complex engineering machinery. Generally, working under extreme and challenging conditions, these industrial equipment are subjected to progressive deterioration, leading to a significant increase in the possibility of related component failure (Helwig et al. 2015a; Egusquiza et al. 2018). This impacts their availability and reliability to minimise operational downtime and maintenance related cost (Sheng et al. 2011). Due to these reasons, monitoring the conditions of these complicated systems as a requirement for predictive-based maintenance has gained increasing importance over the years since it determines the required maintenance action based on equipment's health status. This ensures the availability and reliability of industrial equipment and offers a significant improvement in their health condition, thus, ultimately increasing asset utilisation and reducing maintenance cost.

However, with the rising demands and increasing complexity of industrial systems, the number of installed sensors and their sampling rate are constantly growing (Schneider et al. 2018). As a result, the processing of high-dimensional data (signals) from multiple sensors for predictive-based maintenance is at risk of suffering from scalability, classification performance and the curse of dimensionality (Houle et al. 2010; Keogh and Mueen 2011; Har-Peled et al. 2012; Herrmann et al. 2012; Bach 2017). The complexity is further increased with the frequent operation of these machine components in extreme and dynamic environments, where their mode of operation is often nonlinear due to the effects of varying environmental and operating conditions such as pressure, volume flow, temperature, drift, noise and engineering variance. This contributes to the deterioration process of industrial machines or related components (Javed 2014; Helwig et al. 2015b). Consequently, the data recorded from monitoring the condition of such industrial systems are often nonlinear, nonstationary and masked with noise as output signals which are characterised by high levels of uncertainty and unpredictability (Wen 2011; Randall and Antoni 2011). Hence, the recorded data present integration and tractability challenges when developing an effective predictive framework. Therefore, there is a need to develop a robust and reliable predictive maintenance model that can tolerate uncertainty and efficiency during diverse operational conditions (Zhang and Randall 2009; Sarkar 2015; Wang 2017).

In literature, Machine Learning (ML) algorithms are among the most powerful and frequently used techniques in developing intelligent predictive maintenance frameworks in various applications (Çınar et al. 2020a). This is due to the enormous potential of ML algorithms to process multivariate and high-dimensional dataset (generated in industries by industrial equipment and machinery) through the extraction of hidden patterns, classification, prediction or visual representation (Helwig et al. 2015a; Raptodimos and Lazakis 2018; Kaur and Kaur 2020). However, a sizable number of these ML algorithms are very task-specific and thus are incapable of being implemented in other specialised tasks. Hence, their performance varies when implemented independently. Also, these ML algorithms are further constrained in producing the desired results when exposed to nonlinear and nonstationary high-dimensional datasets characterised by high levels of uncertainties (Cho et al. 2018). For these reasons, researchers in the field of predictive maintenance have directed their focus into building hybrid frameworks by leveraging on the strength and weakness of the multiple ML algorithms which unquestionably are improvements of existing techniques (Zhang et al. 2014; Chakraborty 2017; Di et al. 2019). Thus, a hybrid approach for analysing high-dimensional data has gained significant attention in the development of predictive maintenance frameworks, but little has been reported in literature. Hence, this study proposes a hybrid synergistic framework for classifying fault conditions based on the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), Principal Component Analysis (PCA) and the Least Squares Support Vector Machine (LSSVM) optimised by Coupled Simulated Annealing (CSA) and Nelder-Mead Simplex (NMS) optimisation algorithms.

The motivation for creating the hybrid is that, for ICEEMDAN, it has the ability to decompose nonlinear and nonstationary signals arising from complex systems into a series of Intrinsic Mode Functions (IMFs), where each resulting IMF represents the respective local transient features (Colominas et al. 2014). Besides, when compared to existing filtering techniques for extracting transient features from nonlinear and nonstationary signals such as the Wigner-Ville Distribution (WVD) (Wigner 1932), Short-Time Fourier Transform (STFT) (Peppin 1994; Newland 2005), Wavelet Transform (WT) (Daubechies 1989) and Empirical Mode

Decomposition (EMD) (Huang et al. 1998), the ICEEMDAN drastically reduces the contamination of noise in a signal and the issue of mode mixing. In so doing, the inability to separate different frequencies into separate IMF's has been greatly resolved. However, despite the good performance of ICEEMDAN, not much attention has been given regarding its application in the field of engineering (Zhang et al. 2018). Hence, this study introduces ICEEMDAN to denoise the signals and reduce the level of uncertainties that might have been introduced into our dataset due to the effects of varying environmental and operating conditions. In the case of PCA, it is a widely-known fact that processing of high-dimensional data which is typical of modern industries (i.e. generally, lots of measured processes, often in high and varying sampling rates are collected in order to detect and control the process) further complicates the model development phase due to numerous features. Hence, building an artificial intelligent fault classifier requires the utilisation of an efficient feature extraction and selection technique as they play a vital role in detecting relevant features in the classification space (Cho and Hoang 2017). As such, PCA is employed to extract the relevant features that contain the most characteristic fault information for each measured process. The PCA will further reduce the redundant influence and as well generate a set of linearly uncorrelated features to serve as inputs to the artificial intelligent classifier, LSSVM. The LSSVM technique was adopted as the intelligent classifier as it converts the inequality constraints from the classical SVM into linear equations for enhancing the computational capability and design over standard SVMs and Artificial Neural Networks (ANNs) (Suykens and Vandewalle 1999; Suykens et al. 2002). However, the classification accuracy of LSSVM is highly influenced by the regularisation and kernel parameters. Thus, the selection of an optimal regularisation and kernel parameters play a significant role in classification performance by distributing samples in a given search space (Cho and Hoang 2017; Mosavi and Edalatifar 2018). In literature, the use of optimisation algorithms such the Particle Swarm Optimisation (PSO) (Kennedy and Eberhart 1995), Genetic Algorithm (GA) (Whitley 1994), Ant Colony Optimization (ACO) (Dorigo et al. 2006), Differential Evolution (DE) (Ilonen et al. 2003), Artificial Bee Colony (ABC) (Karaboga et al. 2014), Gravitational Search Algorithm (GSA) (Rashedi et al. 2009) and others for estimating such parameters has been greatly explored. However, considering the theory of No Free Lunch [58] which implies that no single optimiser can boast of being superior to the others for all optimisation tasks, and as such meta-heuristics are task-specific, inspires the selection of CSA-NMS hybrid optimisation algorithms as the choice for estimating the optimal parameters required for training the LSSVM classifier in this study. The hybrid CSA-NMS is a computationally fast global optimiser with high local optima avoidance capability whiles the NMS is utilised for fine-tuning the regularisation and kernel parameters. Hence, the CSA-NMS improves the convergence speed (i.e. considerable reduction in time complexity) and classification accuracy for fault detection.

In continuance of that, the contributions of this study are to:

(a) proposes a hybrid synergistic ML predictive maintenance framework of ICEEMDAN-PCA-LSSVM for fault classification capable of effectively handling high levels of uncertainty from complex, nonlinear and diverse operational conditions of machinery; and
(b) evaluate and compare the performance of the proposed hybrid approach with its variants as well as some well-established ML classifiers.

The efficiency of the proposed hybrid ICEEMDAN-PCA-LSSVM technique is verified by applying to the University of California, Irvine (UCI) machine learning benchmark data obtained from multi-sensors of a hydraulic system (Helwig et al. 2015a). The proposed hybrid ICEEMDAN-PCA-LSSVM technique reduces the redundancies and the dimension of features, allowing for the efficient consideration of relevant features for the enhancement of classification accuracy and convergence speed. Moreover, the proposed hybrid technique advances the field of predictive maintenance by improving the general performance of feature extraction and fault diagnosis framework for various complex and nonlinear operational environment systems.

The remaining sections of the paper are organised as follows. Section 2 discusses some related works in the field of predictive maintenance. Section 3 describes the complex hydraulic system UCI benchmark dataset. Section 4 introduces the proposed hybrid ICEEMDAN-PCA-LSSVM technique and briefly presents the ICEEMDAN, PCA, optimised LSSVM techniques. Section 5 describes the application of the proposed

technique to multiclass fault classification of the complex hydraulic system. Concluding remarks are presented in Section 6.

## 2 Related works

As discussed in the introductory section, ML algorithms have extensively been used in solving numerous and diverse industrial problems such as predictive maintenance of manufacturing systems, machines and related components. A recent review of articles from academic databases (Scopus, ScienceDirect, Institute of Electronic and Electrical Engineers (IEEE) and Google Scholar) conducted by Çınar et al. (2020) on the application of ML algorithms in predictive maintenance within 2010 to 2020 yielded 788 research works as at July 30, 2020. Upon further limiting the search scope to only engineering, energy and material science (excluding reviews and conference reviews) resulted in 217 research works (Çınar et al. 2020b). However, among these 217 research works, most were published from the year 2015 to 2020, suggesting that the implementation of ML algorithms in the field of predictive maintenance is a relatively new approach with growing interesting in the world of science.

To collect related works pertaining to this study, specific keywords such as predictive maintenance, machine learning, fault classification and hydraulic system were used to search the various academic and scholarly databases. The search period was specified from 2015 to 2021 in order to produce the ML techniques that have used in predictive maintenance with a key interest in hydraulic systems, and how they have evolved over the years. The search displayed almost 100 research works, all focusing on various engineering setups utilising hydraulic components (Gomes et al. 2016; Xu et al. 2017a; Hao et al. 2020; Pugin 2020). Aside from the general maintenance of the hydraulic system, other research works focus on specific key components such as the valve (Vianna et al. 2015; Karanović et al. 2019; Lei et al. 2019), pump (Wang et al. 2016; Xu et al. 2017b; Casoli et al. 2019), accumulator (Niu et al. 2016; Pfeffer et al. 2016; Leon-Quiroga et al. 2020), cooler (Hathaway et al. 2018).

For a better comprehension of this study, the scope of the related works is further limited to prior works conducted using the same hydraulic system dataset (obtained from the UCI machine learning repository) employed. This will allow the tracking of progress that has been made regarding the usage of ML algorithms in monitoring the conditions of the considered hydraulic system data since its inception in 2015. Table 1 summarises the ML algorithms that have been utilised in developing predictive maintenance methods based on the UCI machine learning repository hydraulic system dataset under consideration. As seen, since the publication of the considered hydraulic dataset by Helwig et al. (2015), several works that have been conducted attempt to proposed predictive maintenance techniques that improve upon the efficiency and accuracy in classifying the four major components (accumulator, cooler, pump and valve conditions). Although improvements have been realised over the years, there still exist some aspects of the dataset that is yet to be explored. Hence, this study introduces the ICEEMDAN denoising technique as the first phase of pre-processing for dealing with high levels of uncertainties introduced into the dataset as a result of the nonlinear and dynamic operational conditions of engineering systems. Comparing the prior works, it was found that none of the existing studies addressed the issue of uncertainty (noise) before classifying the degradation states of the monitored conditions. Therefore, performing data denoising as a pre-processing step is a contribution to knowledge in predictive maintenance.

## 3 Condition monitoring of hydraulic system dataset

Time series data recorded from monitoring the condition of a hydraulic system is obtained from the UCI ML repository via http://archive.ics.uci.edu/ml/datasets/Condition+monitoring+of+hydraulic+systems. The essential details of the dataset are briefly discussed in this section, with detailed discussion found at the source (Helwig et al. 2015a). The dataset comprised of raw sensor readings with different sampling rates corresponding to different number of attributes (60 – 6000) per sensor. In all, the data consisting of 17 process sensors were made up of 2205 instances and 43680 attributes. The dataset also contains fault scenarios depicting the variations in fault condition (health status) of major components such as hydraulic accumulator, cooler, internal pump leakage and valve. The details of the hydraulic system dataset are shown in Table 2.

## 4 Proposed hybrid technique (ICEEMDAN-PCA -LSSVM)

Fig. 1 shows the schematic of the proposed synergistic hybrid technique from the data pre-processing stage through to the classification stage. First, ICEEMDAN is employed as a denoising technique to decompose the nonlinear and nonstationary signals into a series of IMFs of which only relevant IMFs containing the fault features are retained for the reconstruction of the signal. The selection of relevant IMFs is based on the stringent threshold for discriminating between relevant and spurious IMFs proposed by Ayenu-Prah and Attoh-Okine (2010). Statistical time-domain features that represent the shape and distribution of the signals are then extracted. PCA is then used to generate a set of uncorrelated features from which relevant features are selected to serve as input parameters for the LSSVM optimised by CSA-NMS for classifying fault types. The details of the various stages shown in Fig. 1 and the methods used are discussed in the subsequent sections.

### 4.1 Data pre-processing

Most machine components are often operated in extreme and dynamic environments where its mode of operation and the data recorded from monitoring the condition of such machinery are often nonlinear, nonstationary and masked with noise as output signals which are characterised by high levels of uncertainty and unpredictability (Wen 2011; Randall and Antoni 2011). These limitations present integration and tractability challenges when developing effective classification and predictive models. Thus, in order to develop a robust and reliable predictive maintenance model from noisy signals, a reliable technique for pre-processing (denoising) and extracting relevant features is required.

In literature, among the frequently used techniques for pre-processing (denoising) and extracting relevant features from such signals (data sets) is to decompose the signals into transient features. Over the years, techniques such as Wigner-Ville Distribution (WVD) (Wigner 1932), Short-Time Fourier Transform (STFT) (Peppin 1994; Newland 2005), Wavelet Transform (WT) (Daubechies 1989) and Empirical Mode Decomposition (EMD) (Huang et al. 1998) have been employed for extracting transient features from nonlinear and nonstationary signals. However, studies have shown that these methods have some limitations. For example, the WVD is useful in extracting transient features in time-frequency signals but cannot account for the local transient features of the signals at a given time and presents cross-terms when signal with many frequency components is being analysed. Moreover, WVD spreads noise and may result in negative amplitude values which are irrelevant leading to further complications (Cohen 1989; Marwala 2012). As opposed to the WVD, the STFT is useful in extracting the localised transient feature by transforming a small-time window into a frequency domain. Conversely, STFT is restricted as any increase in the time resolution negatively affects the frequency resolution, and vice versa (Hlawatsch and Boudreaux-Bartels 1992). As a remedy, the WT was proposed as an efficient alternative for dealing with fixed time-frequency resolution problems and transient signals in general (Galli et al. 1996). Nonetheless, a low resolution is attained at higher frequencies as the frequency from WT is logarithmically scaled (Barschdorff and Femmer 1995).

To overcome the deficiencies of these techniques, a nonparametric feature extraction method, the EMD was proposed to decompose nonlinear and nonstationary signals arising from complex systems into a series of Intrinsic Mode Functions (IMFs), where each resulting IMF represents the respective local transient features (Huang et al. 1998). On the contrary, the issue of mode mixing, that is, the inability to separate different frequencies into separate IMF's is a major drawback of the EMD (Haddar 2018; Mahgoun et al. 2018; Cheng et al. 2019). To eliminate the mode mixing problem, a noise-assisted version, the Ensemble EMD (EEMD) that involves adding gaussian white noise to the original signals was proposed (Wu and Huang 2009). However, the resulting reconstructed signals led to the formation of extra modes which requires extra iteration effort to decompose the given signal. To optimise the decomposition of signals whiles reducing reconstruction error, the Complete Ensemble EMD with Adaptive Noise (CEEMDAN) (Torres et al. 2011) has been proposed. In this study, the recently enhanced formulation of CEEMDAN known as the Improved Complete Ensemble EMD with Adaptive Noise (ICEEMDAN) that has the ability to reduce the contamination of noise in a signal (Colominas et al. 2014) is adopted. ICEEMDAN is proposed as a pre-processing technique to extract the required fault features whiles reducing the noise and high levels of uncertainty.

### 4.1.1 Improved Complete Ensemble EMD with Adaptive Noise (ICEEMDAN)

The ICEEMDAN is an adaptive method for decomposing nonlinear and nonstationary signals (both time and frequency domains) into a series of Intrinsic Mode Functions (IMFs), where each resulting IMF represents the respective local transient features (Huang et al. 1998). The ICEEMDAN technique, has great improvement on the Ensemble EMD (EEMD) and Complete Ensemble EMD (CEEMD) methods, as was stated in Section 4.1. The ICEEMDAN is illustrated by the addition of a Gaussian noise series to the actual signal as shown in Equation (1).

$$x^i(t) = x(t) + \beta_0 E_1\left(w^i(t)\right) \tag{1}$$

where $x(t)$ is the signal, $\beta_{i-1} = \varepsilon_{i-1}\text{std}(x)/\text{std}(E_i\left(w^i(t)\right)$ is the ratio coefficient, $E_k(\cdot)$ is the operator for producing the $k^{th}$ IMF generated by EMD, $w^i(t)$ is a realisation of unit variance white Gaussian noise with a zero mean.

Using Equation (1), EMD is used to gain the first residual $\left(r_1\right)$ of the signal by computing the local mean of $I$ realisations as shown in Equation (2).

$$r_1 = \left\langle M\left(x^i(t)\right)\right\rangle \tag{2}$$

where $M(\cdot)$ is the operator for estimating the local mean.

For the first IMF, $k = 1$, is estimated using Equation (3).

$$IMF_1 = x(t) - r_1 \tag{3}$$

The second residual $\left(r_2\right)$ is estimated as the average of the local means of the realisations as shown in Equation (4).

$$r_2 = \left\langle M\left(r_1 + \beta_1 E_2\left(w^i(t)\right)\right)\right\rangle \tag{4}$$

Then, the second IMF at $k = 2$ is estimated using Equation (5).

$$IMF_2 = r_1 - r_2 \tag{5}$$

The $k^{th}$ residual and $k^{th}$ IMF are estimated using Equations (6) and (7), respectively.

$$r_k = \left\langle M\left(r_{k-1} + \beta_{k-1} E_k\left(w^i(t)\right)\right)\right\rangle \tag{6}$$

$$IMF_k = r_{k-1} - r_k \tag{7}$$

The process from Equations (6) and (7) is repeated for the next $k$.

### 4.1.2 Signal reconstruction

In the application of the ICEEMDAN on the data, it is of paramount importance to select the relevant IMFs which contains as much fault information as possible from the series of IMFs generated by ICEEMDAN whiles neglecting the spurious IMFs. As such, an effective, quick and repeatable scientific framework is needed to discriminate between relevant and spurious IMFs for the signal reconstruction to aid improve the performance of the fault diagnosis system. In this paper, a stringent threshold for discriminating between relevant and spurious IMFs in the presence of high levels of noise, proposed by Ayenu-Prah and Attoh-Okine (2010) is employed as the criterion for selecting the relevant IMFs. The stringent threshold $(\lambda)$ is expressed as a function of the correlation coefficient $(\rho)$ between the observed signal $x(t)$, and each $IMF_i$, as shown in Equation (8).

$$\lambda = \frac{\max\left(\rho_i\right)}{10 \times \max\left(\rho_i\right) - 3}, \qquad i = 1, 2, \text{K}, k \tag{8}$$

where $\rho_i$ is estimated using Equation (9).

$$\rho_i = \frac{\text{cov}\left(x(t), IMF_i\right)}{\sqrt{\text{cov}\left(x(t), x(t)\right) \bullet \text{cov}\left(IMF_i, IMF_i\right)}} \tag{9}$$

Using Equation (8), an IMF is regarded as relevant if its corresponding $\rho_i \geq \lambda$. After identifying all the relevant IMFs, they are then summed to form the reconstructed signal for subsequent analysis.

It must be noted that the standard ICEEMDAN algorithm proposed by Colominas et al. (2014) whose source code is found at http://bioingenieria.edu.ar/grupos/ldnlys/metorres/re_inter.htm#Codigos, allows for the decomposition of only one signal or feature or univariate series into IMF's at a time. Hence, the implementation of ICEEMDAN using the above source code becomes practically impossible with high-dimensional datasets which is the domain of this study. To overcome this drawback, the standard ICEEMDAN algorithm is modified by integrating the stringent threshold (Equation (8)) as a criterion for discriminating between relevant and spurious IMFs, and finally converting to an iterative algorithm using the *for loop* function such that the decomposition followed reconstruction will be done for all features. This modification extended the capabilities of the standard ICEEMDAN source code by automating the decomposition and reconstruction of multiple signals all in a single execution of the algorithm as shown in Fig. 2.

### 4.2 Feature extraction and selection

After denoising the 17 process sensors using the ICEEMDAN technique discussed in the previous section, the next issue that needs to be addressed is the dimension of features (43680), which is relatively high. Consequently, conventional ML techniques will suffer from tractability, scalability, high time complexities and more importantly classification performance issues (Houle et al. 2010; Keogh and Mueen 2011; Har-Peled et al. 2012; Herrmann et al. 2012; Bach 2017). Also, feeding the denoised 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 represents the characteristic properties of the hydraulic dataset from the 43680 denoised features.

#### 4.2.1 Extraction of statistical features

In prior works such as Helwig et al. (2015a, b), the hydraulic system dataset was partitioned into various time intervals. In this study, as a means of ensuring uniformity within all cycles, the various time intervals were allocated after averaging each feature per sensor. Also, similar to prior research works where the data for each cycle per sensor were all partitioned into various segments, the partitioning of the cycles in this study varied with varying characteristic time intervals of the cycles. Fig. 3(a), for instance, shows 13 characteristic time intervals after averaging each feature in the PS1 sensor data. Similarly, PS5 was partitioned into 19 characteristic time intervals as shown in Fig. 3(b). The column labelled "Time Interval" in Table 2 shows the various number of characteristic time intervals each sensor was segmented into. Different statistical time-domain features such as mean, median, variance, standard deviation, skewness, kurtosis and maximum peak value in each segment were then extracted. This resulted in a pool of 1806 features (approximately 2418.61% reduction). However, building an intelligent classifier with all the 1806 features which is still high in terms of dimension will significantly increase the complexity of the framework. Therefore, before an artificial intelligent fault classifier can be applied to the resulting pool of 1806 features, it is imperative for a second stage involving dimensionality reduction to be carried out. Hence, this study employs the PCA, an effective dimensionality reduction technique. The PCA is capable of extracting relevant features containing the most characteristic fault information from the resulting pool of 1806 time-domain features extracted from the reconstructed signals.

#### 4.2.2 Principal Component Analysis (PCA)

PCA is one of the standard unsupervised techniques often used to effectively transform large sets of features into few independent and uncorrelated features whiles retaining as maximum variability as possible. The algorithm is summarised as follows.

Suppose $X = (X_1, X_2, \mathrm{K}\ , X_p)^T$ is a random vector of features of $p$ dimension with population covariance matrix as shown in Equation (10).

328
$$\sum_X = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \mathrm{L} & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \mathrm{L} & \sigma_{2p} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ \sigma_{p1} & \sigma_{p2} & \mathrm{L} & \sigma_p^2 \end{pmatrix} \tag{10}$$

329 Consider the linear combination expressed in Equation (11).

330
$$\begin{aligned} Y_1 &= e_{11}X_1 + e_{12}X_2 + \mathrm{L} + e_{1p}X_p \\ Y_2 &= e_{21}X_1 + e_{22}X_2 + \mathrm{L} + e_{2p}X_p \\ \mathrm{M} \quad & \mathrm{M} \quad \mathrm{M} \quad \mathrm{M} \quad \mathrm{M} \\ Y_p &= e_{p1}X_1 + e_{p2}X_2 + \mathrm{L} + e_{pp}X_p \end{aligned} \tag{11}$$

331 Each linear combination in Equation (11) can be thought of as linear regression, predicting $Y_i$ using
332 $X_1, X_2, \mathrm{K}, X_p$ where $e_{i1}, e_{i2}, \mathrm{K}, e_{ip}$ are the regression coefficients.

333 $Y_1$, the first Principal Components (PC1) is the linear combination of $X_i$ such that, $Y_1$ accounts for the
334 maximum variation in the data as shown in Equation (12).

335
$$\mathrm{var}(Y_1) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{1l}\sigma_{kl} \tag{12}$$

336 As such, $e_{11}, e_{12}, \mathrm{K}, e_{1p}$ are defined such that maximum variance is achieved, subject to the constraint shown
337 in Equation (13).

338
$$\sum_{j=1}^{p} e_{1j}^2 = 1 \tag{13}$$

339 $Y_2$, the second Principal Component (PC2) is the linear combination $X_i$ such that, $Y_2$ accounts for as much
340 variance remaining after PC1, as shown in Equation (14).

341
$$\mathrm{var}(Y_2) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{2k}e_{2l}\sigma_{kl} \tag{14}$$

342 As such, $e_{21}, e_{22}, \mathrm{K}, e_{2p}$ are selected to maximise the variance of $Y_2$, subject to the constraint shown in
343 Equation (15).

344
$$\sum_{j=1}^{p} e_{2j}^2 = 1 \tag{15}$$

345 For $Y_i$, the $i^{th}$ Principal Component, $(e_{i1}, e_{i2}, \mathrm{K}, e_{ip})$ selected to maximise the variance of $Y_i$ (Equation
346 (16), subject to the constraint shown in Equation (17).

347
$$\mathrm{var}(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} \tag{16}$$

348
$$\sum_{j=1}^{p} e_{ij}^2 = 1 \tag{17}$$

349 All PC's generated are uncorrelated, that is, $\mathrm{cov}\left(Y_{i-1}, Y_i\right) = 0$ and will serve as the inputs data for the artificial
350 intelligent classifier.

### 4.3 Classification and parameter optimisation

#### 4.3.1 Least Squares Support Vector Machine (LSSVM)

The LSSVM technique as proposed by Suykens and Vandewalle (Suykens and Vandewalle 1999) is an extended least squares version of the standard SVM classifier. LSSVM imposes equality constraints by utilising a set of linear equations formulated from all training data during the training process. Thus, enhancing LSSVMs computational capability and generalisation over standard SVMs whiles reducing computational complexity.

Suppose $\{x_i, y_i\}$ is a random vector of training features, for $i = 1,2,\mathrm{K}\,,n$ instances, where input features with its corresponding target are $x_i \in ¡^{n}$ and $y_i \in ¡^{n}$, respectively. The LSSVM formulated is expressed as an optimisation problem with the objective function shown in Equation (18) subject to the equality constraint shown in Equation (19).

$$\min_{\omega,b,\xi} J(\omega,b,\xi) = \frac{1}{2}\omega^T\omega + \gamma\frac{1}{2}\sum_{i=1}^{n}\xi_i^2 \tag{18}$$

$$y_i = \omega^T\phi(x_i) + b + \xi_i^2, \qquad i = 1,2,\mathrm{K}\,,n \tag{19}$$

where $J(\omega,b,\xi)$ is the objective function with weight vector dimensional space $\omega \in ¡^{n}$, error variables $\xi_i \in ¡^{n}$ and bias term $b$. $\gamma$ is an influential positive penalty factor of the trade-off between the margin and training error, $\phi(\cdot)$ is a function which maps the input space into a high-dimensional space.

The LSSVM model is expressed as shown in Equation (20).

$$f(x) = \omega^T\phi(x) + b \tag{20}$$

Using Equation (18), a Lagrangian $L$, is then defined (Equation (21)).

$$L(\omega,b,\xi;\alpha) = J(\omega,b,\xi) - \sum_{i=1}^{n}\alpha_i\left(\xi_i - y_i + \omega^T\phi(x_i) + b\right) \tag{21}$$

where $\alpha_i \in ¡^{n}$ is the Langrage multiplier, which follows the Karush-Kuhn-Tucker optimality conditions as shown in Equation (22).

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \displaystyle\sum_{i=1}^{n}\alpha_i\phi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \displaystyle\sum_{i=1}^{n}\alpha_i = 0 \\ \dfrac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma\xi_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow \xi_i - y_i + \omega^T\phi(x_i) + b = 0 \end{cases} \tag{22}$$

A solution (Equation (23)) is obtained after eliminating $\omega$ and $\xi_i$.

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & K + \gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{23}$$

where $1_v = [1,1,\mathrm{K}\,,1]^T$, $\alpha = [\alpha_1,\alpha_2,\mathrm{K}\,,\alpha_n]^T$ and $y = [y_1, y_2,\mathrm{K}\,, y_m]^T$.

According to the Mercer's condition, the kernel function $K(x_i, x_j)$, can be chosen such that

$$K\left(x_i, x_j\right) = \phi\left(x_i\right)^T \phi\left(x_j\right), \;\; i, j = 1,2,\mathrm{K}\,,n \tag{24}$$

This results in the final LSSVM model (Equation (25)) used for estimating the function.

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b \tag{25}$$

where $\alpha, b$ are solutions to Equation (23).

However, it is a widely known fact that the classification accuracy of LSSVM is highly influenced by the optimal selection of the kernel function and regularisation parameter. In this paper, the frequently used kernel function, the Radial Basis Function (RBF) was selected due to its excellent general performance, wider convergence domain, high-resolution power and requires fewer parameters (Keerthi and Lin 2003; Du et al. 2016; Wang et al. 2018). The RBF kernel function is expressed in Equation (26).

$$RBF = K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{26}$$

For the selection of the optimal bandwidth $(\sigma)$ and regularisation $(\gamma)$ parameters for an effective classification by LSSVM, a hybrid of Coupled Simulated Annealing (CSA) and Nelder-Mead Simplex (NMS) optimisation algorithms were adopted to determine the starting points and fine-tune the parameters, respectively.

#### 4.3.2 Coupled Simulated Annealing (CSA)

CSA is an accelerated global optimisation technique characterised by a set of parallel simulated annealing processes coupled with their acceptance probability function (Xavier-de-Souza et al. 2009). The introduction of a coupling term in CSA allows for the creation of cooperative behaviour through the exchange of information which help steer the overall optimisation process toward the global optimum. This minimises the sensitivity of the technique to initialisation parameters, thus resulting in a much efficient optimisation.

In CSA, the acceptance probability function $A_\theta\left(\rho, x_i \to y_i\right)$ for every $x_i \in \Theta$, $y_i \in \Omega$ with coupling term $\rho$ is as shown in Equation (27).

$$A_\theta\left(\rho, x_i \to y_i\right) = \frac{\exp\left(\dfrac{-E\left(y_i\right)}{T_k^{ac}}\right)}{\exp\left(\dfrac{-E\left(y_i\right)}{T_k^{uc}}\right) + \rho} \tag{27}$$

where $x_i$ and $y_i$ are $i^{th}$ the current state and its corresponding probing state respectively, for $i = 1,2,\mathrm{K}\,,m$, with $m$ being the number of entries in $\Theta$. The $\Theta = \left\{x_i\right\}_{i=1}^{m}$ denotes the set of current states. The coupling term $\rho$, is a function of the energy of the entries in $\Theta$ as shown in Equation (28)

$$\rho = \sum_{x \in \{x_1, x_1, \mathrm{K}\,, m\}} \exp\left(\frac{E\left(x\right)}{T_k^{ac}}\right) \tag{28}$$

Hence, CSA considers many current states $x_i = \left\{x_1, x_2, \mathrm{K}\,, x_m\right\}$ in the set of current states $\Theta$ and accepts a probing state $y_i = \left\{y_1, y_2, \mathrm{K}\,, y_m\right\}$ which is not based only on its corresponding $x_i$ but also on the coupling term $\rho$. In this paper, CSA is used to determine the starting values of the optimal bandwidth $(\sigma)$ and the regularisation $(\gamma)$ parameters of the LSSVM approach which are then passed on to the NMS technique for fine-tuning.

### 4.3.3 Nelder-Mead Simplex (NMS) Algorithm

The NMS algorithm is one of the popular direct search algorithms for optimising multidimensional unconstrained problems (Nelder and Mead 1965). Unlike other simplex methods, the NMS algorithm is known to be an improvement as it allows the simplex to vary not only in size but as well in shape (Baudin 2009). The algorithm preserves a simplex (i.e. a geometric figure based on $n$ parameters defined by $n+1$ vertices in $n+1$ initial experiments) which are approximations of the optimal point. The vertices are sorted based on increasing function values from the objective function. The algorithm then tries to replace the worst vertex with a new point which is dependent on the worst point and the centre of the best vertices. The expansions and contraction operations (varying in size and shape) of the algorithm allows for convergence speed which significantly reduces the time complexity.

Suppose Equation (29) is an unconstrained optimisation problem

$$\min f(x) \tag{29}$$

where $f$ is the objective function with $x \in ¡^{n}$ and $n$ as the number of optimisation parameters. The NMS algorithm is based on the iterative update of a simplex $x_i = \{x_1, x_2, \mathrm{K}, x_{n+1}\}$ which is an $n \times (n+1)$ matrix with each column representing a simplex vertex $(v_i)$. The algorithm utilises four major coefficient parameters: the reflection $(x_r > 0)$, expansion $(x_e > 1)$, contraction $(0 < x_c < 1)$ and shrinkage $(0 < \sigma < 1)$ . The NMS algorithm is presented as follows:

**Step 1:** Build $n+1$ vertices of $x_i$ such that $f(x_1) \le f(x_2) \le \mathrm{K} \le f(x_{n+1})$ .

**Step 2:** Use Equation (30) to compute the reflection point $x_r$

$$x_r = \overline{m} + \alpha(\overline{m} - x_{n+1}) \tag{30}$$

where $\overline{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the centroid of the $n$ best vertices except $x_{n+1}$. If $f(x_1) \le f(x_r) < f(x_n)$, the iteration is terminated by accepting $x_r$ .

**Step 3:** If $f(x_r) < f(x_1)$, the *expansion* point $x_e$ is estimated using Equation (31)

$$x_e = \overline{m} + \beta(x_r - \overline{m}) \tag{31}$$

If $f(x_e) \le f(x_r)$, $x_e$ is accepted, else $x_r$ is accepted whiles $x_e$ is discarded and iteration is terminated.

**Step 4:** If $f(x_r) \ge f(x_n)$, estimate the contraction point between $\overline{m}$ and the best among $x_r$ and $x_{n+1}$

i. If $f(x_r) < f(x_{n+1})$, an outside contraction point $x_{cout}$ is estimated using Equation (32)

$$x_{cout} = \overline{m} + \gamma(x_r - \overline{m}) \tag{32}$$

If $f(x_{cout}) \le f(x_r)$ , $x_{cout}$ is accepted and iteration is terminated, else skip to Step 5.

ii. Elseif $f(x_r) \ge f(x_{n+1})$, an inside contraction point $x_{cin}$ is estimated using Equation (33)

$$x_{cin} = \overline{m} + \gamma(x_r - \overline{m}) \tag{33}$$

If $f(x_{cin}) \le f(x_r)$, $x_{cin}$ is accepted and iteration is terminated, else skip to step 5.

**Step 5:** $f$ is then evaluated at $n$ points $v_i = x_1 + \sigma\left(x_i - x_1\right)$, for every $i = 2,\mathrm{K}\ ,n+1$. For the next iteration, the vertices $v_i$ are assigned to $x_i$.

The standard values for the four major coefficient parameters: $x_r = 1$, $x_e = 2$, $x_r = \frac{1}{2}$ and $\sigma = \frac{1}{2}$.

The optimal bandwidth $(\sigma)$ and regularisation $(\gamma)$ parameters proposed and refined by the combination of CSA and NMS algorithms are then passed on to LSSVM to facilitate the classification.

### 4.4 Classification performance evaluation

From the statistical viewpoint, deducing a comprehensive evaluation of multiclass classification models based on a single performance index is not easy and enough. For these reasons, eight evaluation metrics namely accuracy, error rate, precision, recall (sensitivity), specificity, F score, Mathews correlation coefficient and geometric mean were used in this study for the purpose of reliability and also to overcome the above drawback.

#### 4.4.1 Accuracy

Accuracy is one of the most widely used evaluation metrics for assessing the performance of classification algorithms (Paul and Maji 2010; Qasem and Nour 2015; Alsalem et al. 2018). Classification accuracy is expressed in Equation (34) as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{34}$$

where $TP$ is the number of correct classification counts when there is a fault condition, $TN$ is the number of correct classification counts when there is no fault condition, $FP$ is the number of misclassification counts when there is a fault condition and $FN$ is the number of misclassification counts when there is no fault condition for a specific classification model.

#### 4.4.2 Error rate

This metric is one of the main indicators for evaluating classification performance by measuring the errors (misclassifications) incurred by a classifier (Chiang and Ho 2008; De Paz et al. 2013; Mohapatra and Chakravarty 2015; Alsalem et al. 2018). The error rate is expressed as shown in Equation (35).

$$Error\ Rate = \frac{FP + FN}{TP + FP + TN + FN} \tag{35}$$

#### 4.4.3 Precision

The precision of a classifier measures the exactness of classification after prediction (Agaian et al. 2014; Mohapatra and Chakravarty 2015; Alsalem et al. 2018). The evaluation metric is expressed as a ratio of true positives $(TP)$ to the sum of true positives $(TP)$ and false positives $(FP)$ as shown in Equation (36).

$$Precision = \frac{TP}{TP + FP} \tag{36}$$

#### 4.4.4 Recall (Sensitivity)

Recall, also known as sensitivity, is a measure of a classifier's capacity to determine positive instances. It measures the fraction of positive instances that are correctly classified (Agaian et al. 2014; Tai et al. 2011; Alsalem et al. 2018) and is expressed as a ratio of true positives $(TP)$ to the sum of true positives $(TP)$ and false negatives $(FN)$ as shown in Equation (37).

$$Recall\ /\ Sensitivity = \frac{TP}{TP + FN} \tag{37}$$

#### 4.4.5 Specificity

Specificity measures the fraction of negative instances that are correctly classified (Agaian et al. 2014; Laosai and Chamnongthai 2014). That is, the metric denotes the test's ability to identify negative result as expressed in Equation (38).

$$Specificity = \frac{TN}{TN + FP} \tag{38}$$

#### 4.4.6 F score

The F Score describes the overall performance of a classification model as the harmonic mean of precision and recall (Agaian et al. 2014; Mohapatra and Chakravarty 2015; Alsalem et al. 2018). The metric ranges from zero to one, with high values indicating high classification performance and vice versa. F Score is given in Equation (39).

$$FScore = \frac{2(Precision * Recall)}{Precision + Recall} \tag{39}$$

#### 4.4.7 Matthews Correlation Coefficient (MCC)

MCC (Equation (40)) measures the relationship between the observed and the predicted classification and is generally regarded as a balanced metric for evaluating the performance of classifiers even with varying class sizes (Boughorbel et al. 2017; Chicco 2017). MCC, as compared to other classification evaluation metrics is known to be more informative as it considers the balance ratios of the four confusion matrix categories. An MCC coefficient ranges from -1 to +1, with +1 suggesting a perfect classification while -1 implies a total misclassification.

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{40}$$

#### 4.4.8 Geometric Mean (GM)

GM (Equation (41)) is an aggregate of both sensitivity and specificity evaluation metrics. The metric seeks to maximise the rate of true positives and negatives instances whiles maintaining a balance between both rates (Hossin and Sulaiman 2015; Kuncheva et al. 2018). Thus, making the metric suitable for imbalanced datasets.

$$GM = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}} \tag{41}$$

## 5 Results and discussion

### 5.1 Data pre-processing

#### 5.1.1 Denoising

The hydraulic system dataset was first subjected to the modified ICEEMDAN technique (integrated with the stringent threshold (Equation (8)) as a criterion for discriminating between relevant and spurious IMFs) as discussed in Section 4.1.2. The modification was automated such that, the denoising (decomposition into IMFs with ICEEMDAN and reconstruction with Equation (8)) was repeated for all the sensors considered in this study as manual involvement was highly impractical. The modified ICEEMDAN technique successfully eliminated most redundancies masked as noise in the sensor data.

#### 5.1.2 Feature extraction and selection

Here, two different data scenarios were considered when applying the PCA. First, the PCA was applied to the resulting pool of 1806 features which have been denoised using the ICEEMDAN technique. Secondly, the PCA was applied directly to the original extracted 1806 features without denoising. The motive is to ascertain the extent to which the denoised data could improve the classifiers prediction capability. Fig. 4(a) shows the scree plot of the eigenvalues and the proportion of variance explained regarding the undenoised

1806 possible components whiles Fig. 4(b) shows the PCA results for the denoised 1806 features. Using the Kaiser's criterion for retaining PCs with eigenvalues greater than 1.0, 161 PCs were retained for the undenoised data representing 91.62% of the variance in the 1806 features. With regards to the denoised data using the ICEEMDAN, the number of selected PCs was substantially reduced by half (82) when compared to the undenoised data, representing 96.04% of the variance in the 1806 features as shown in Table 3. The results from both approaches (with and without the use of ICEEMDAN) clearly indicate the relevance of the ICEEMDAN technique in removing substantial levels of noise in the original hydraulic sensor data before feature extraction and selection. Thus, ensuring the extraction of relevant features that contains the most characteristic fault information projected onto a minimal number of uncorrelated PCs whiles maintaining a high proportion of variance (96.04%) during the dimensionality reduction phase.

### 5.2 Multiclass classification

Multiclass classification refers to classifying input features to one of a predefined set based on an optimal subset of features selected during the feature selection stage. In this study, the multiclass classification model result of the optimised LSSVM was compared with three benchmark machine learning techniques that have been used extensively in literature and have all demonstrated promising multiclass classification capabilities. The methods include Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Artificial Neural Network (ANN), specifically, the Generalised Regression Neural Network (GRNN). If the performance of the tested classification model is deemed satisfactory, the classification model can be used to classify the unknown future health status of the hydraulic component. Fig. 5 shows the illustration of the procedure used in building the multiclass classification models. The classification experiment was simulated on MATLAB (2018a version) using an ASUS computer with a 2.3 GHz processor computer with 4 GB memory.

In order to ascertain the reliability of the multiclass classification models, the selected features from PCA are randomly partitioned into two sets; 70% (training set) to be used in training the multiclass classification models and the 30% (testing set) is used to validate the optimum trained models. However, the random nature of the hybrid CSA-NMS optimisation algorithms in providing the optimal bandwidth and the regularisation parameters of the LSSVM will yield slightly different output after each run. As a result, the LSSVM is implemented using the one-vs-one with RBF kernel, and is trained 10 times based on randomly partitioned data into training and test sets. The final classification output is presented as an average, standard deviation (STD), best and worst of all the outputs from the 10 runs as shown in Table 4. Notably, the proposed hybrid technique achieved average test classification accuracies greater the 99.40% for all four monitored conditions. Also found in Table 4 is the training accuracy, its deviation (10-fold cross-validation) and CPU time regarding the training of LSSVM optimised by CSA-NMS for the four hydraulic components.

#### 5.2.1 Classification rate based on optimised LSSVM compared to other investigated classifiers

Table 5 shows the average multiclass classification test results for hydraulic accumulator condition based on pre-processed ICEEMDAN-PCA and PCA input features for the optimised LSSVM, LDA, SVM and ANN. Although all the classifiers showed satisfactory test performance, the proposed ICEEMDAN-PCA-LSSVM outperformed all the other methods. This is evident from the various evaluation metrics employed (Table 5). The performance (99.44%) by the proposed technique in classifying the accumulator condition is remarkably significant when compared to the results reported in prior work after classification based LDA (54.0%), ANN (50.4%), SVM-linear (51.6%) and SVM-RBF (65.7%) (Helwig et al. 2015a). Also, another prior work conducted by Chawathe (2019) reported similar classification results ranging from >35% to <100% based on seven different classifiers/ensemble. In prior works, the classification of accumulator condition is reported as the most difficult among the four monitored components (Helwig et al. 2015a; Chawathe 2019). Hence, the proposed hybrid model of ICEEMDAN-PCA-LSSVM demonstrates superiority of producing a more accurate classification of the hydraulic accumulator condition.

Also seen in Table 5 is a comparative analysis of the proposed ICEEMDAN-PCA-LSSVM with its variant, undenoised PCA-LSSVM. The comparative analysis is performed to assess the contribution of the ICEEMDAN to the proposed hybrid technique. As seen, the proposed ICEEMDAN-PCA-LSSVM model accurately classified 99.44% of the test instances with an error of 0.56% as compared to the 94.09% accuracy with 5.91% misclassification rate (error) in the case of not using ICEEMDAN to pre-process the data (i.e.

PCA-LSSVM). This suggests that pre-processing (denoising) the original signals with ICEEMDAN improved the proposed hybrid technique by 5.35%. In terms of precision, the proposed ICEEMDAN-PCA-LSSVM was capable of classifying 99.56% of the positive instances correctly out of the total test instances whiles the PCA-LSSVM technique could only classify 93.03% of positive instances. Also, the highest sensitivity and specificity values of 0.9941 and 0.9980 imply the proposed models' (ICEEMDAN-PCA-LSSVM) ability to determine 99.41% of all positive instances and 99.80% of all negative instances, respectively. The sensitivity and specificity values of the PCA-LSSVM technique, on the other hand, could classify only 93.33% and 98.10% of positive and negative instances respectively. This suggests that the proposed ICEEMDAN-PCA-LSSVM model has high potency for differentiating positive instances from negative instances. Furthermore, the ICEEMDAN-PCA-LSSVM model demonstrated the highest f-score of 0.9948 as compared to the 0.9317 of the PCA-LSSVM technique regarding overall performance of classification. A correlation coefficient of 0.9929 and 0.9126 for ICEEMDAN-PCA-LSSVM and PCA-PSSVM implies a positive relationship between the classified and the observed test instances for both techniques. However, it was stronger for the proposed ICEEMDAN-PCA-LSSVM model. All these evaluation metrics along with a high geometric mean value of 0.9960 collectively suggest nearly perfect classification results for accumulator condition is achieved using the proposed hybrid model of ICEEMDAN-PCA-LSSVM.

In the case of the cooler condition, similar average multiclass classification test results as in the accumulator condition were obtained (Table 6). However, the classification results obtained by the proposed ICEEMDAN-PCA-LSSVM, ICEEMDAN-PCA-ANN and PCA-LSSVM models were quite similar. That is, these models achieved the same highest level of accuracy (99.83%) with a corresponding lowest misclassification rate (error) of 0.17%. Similar results were obtained for the remaining evaluation metrics considered. The interpretation here is that the proposed ICEEMDAN-PCA-LSSVM model can adequately be used in classifying cooler conditions.

Table 7 shows the average multiclass classification test results for internal pump leakage condition. As observed in Table 7, the ICEEMDAN based classification with LSSVM and SVM models showed better accuracy and generalisability. This is evident from the perfect test classification results (100%) obtained by both the proposed hybrid model (ICEEMDAN-PCA-LSSVM) and the ICEEMDAN-PCA-SVM model. These 100% test classification results obtained by the proposed technique is seen to be a significant improvement to prior works of classification based LDA (73.6%), ANN (80.0%), SVM-linear (72.4%) and SVM-RBF (64.2%) (Helwig et al. 2015a). Also, similar classification results ranging from >55% to <100% based on seven different classifiers/ensemble was obtained by Chawathe (2019). Deductions from prior works indicate the classification of the pump condition to be the next most difficult after the accumulator condition (Helwig et al. 2015a; Chawathe 2019). This suggests that the proposed hybrid model of ICEEMDAN-PCA-LSSVM and ICEEMDAN-PCA-SVM are optimal for the accurate classification of internal pump leakage condition.

With regards to the average multiclass classification of the valve conditions (Table 8), the proposed ICEEMDAN-PCA-LSSVM model showed better classification results than the alternative models. This is manifested in the 99.84% accuracy with a 0.16% misclassification (error) rate obtained. The same trend of results was noticed for the remaining evaluation metrics considered (Table 8). Here, the proposed ICEEMDAN-PCA-LSSVM model is considered to be suitable in classifying the valve condition of the hydraulic system.

The overall analyses of all the multiclass classification results of the various fault conditions (accumulator, cooler, pump leakage and valve) have shown the superiority of the ICEEMDAN-PCA-LSSVM proposed. This is because the ICEEMDAN-PCA-LSSVM is able to classify all the fault conditions better than the other methods. Hence, the method possesses the added advantage of being versatile in multiclass classification tasks as observed in this study for the various monitored conditions. Methods such as ICEEMDAN-PCA-ANN and PCA-LSSVM which achieved relatively similar classification results could only do so in the case of cooler conditions (Table 6) while ICEEMDAN-PCA-SVM only classified the internal pump leakage correctly (see Table 7). These suggest that the proposed ICEEMDAN-PCA-LSSVM model is highly potent in classifying a wide range of monitored conditions irrespective of the dynamic operational condition of the

machine component. This will aid machine operators to spot significant changes in machine components which is an indication of fault development. Thus, increasing the availability and reliability of machines, as well as reducing maintenance-related cost by presenting operators with an accurate and informed decision as to when to schedule maintenance.

Furthermore, this study has practically demonstrated the relevance of denoising the original signals by the ICEEMDAN technique to enhance the overall classification results. This is evident in the results presented in Tables 5 - 8 where two scenarios of denoising the originally recorded data using ICEEMDAN and using the originally recorded data without denoising in the modelling process. The strength of the proposed modelling approach (ICEEMDAN-PCA-LSSVM) can additionally be viewed from Figs. 6 and 7. Here, only the PCA-LSSVM was chosen to be compared with the proposed ICEEMDAN-PCA-LSSVM model because it was the best of the models based on undenoised features. Considering the relevance of denoising the signals with ICEEMDAN hybridised with PCA and LSSVM for improving the classification of diverse monitored conditions, it may be limited when experimented on signals with mutations and similar dominant frequencies.

## 6 Conclusions

In this study, a hybrid approach based on the ICEEMDAN, PCA and LSSVM optimised by CSA-NMS for multiclass fault classification has been developed and evaluated. In ascertaining the superiority of the proposed hybrid approach (ICEEMDAN-PCA-LSSVM), a comparative analysis with three well-established benchmark classifiers (i.e. LDA, SVM and ANN) was carried out using benchmark data obtained from multi-sensors of a hydraulic test rig. From the analysis of the results, it was found that the proposed hybrid ICEEMDAN-PCA–LSSVM model was versatile and had the best results across all the performance indicators leading to better diagnostic of various fault conditions. The proposed hybrid approach also had the tendency of reducing the redundancies and the dimension of features thereby allowing for the efficient consideration of relevant features for the enhancement of classification accuracy and convergence speed. Hence, the proposed hybrid technique possesses the added advantage of being versatile in classifying various monitored conditions. The compared classifiers achieved relatively similar classification performance; however, they could only do so in the case of the cooler and the internal pump leakage. These suggest that the proposed hybrid is highly potent in classifying a wide range of monitored conditions irrespective of the dynamic operational conditions. Although the proposed hybrid technique generally improves the classification of diverse monitored conditions, it may be limited when experimented on signals with mutations and similar dominant frequencies. Also, the feature extraction technique as well as those utilised in prior research works may be constrained when dealing with extremely larger datasets due to the level of manual involvement by the user. Hence, future works should explore the following: the usage of filters capable of addressing signals with mutation and similar dominant frequencies, deep learning approaches to feature extraction. This will ultimately improve the optimality of the pre-processed signals for various fault classification tasks in the field of predictive maintenance.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## References


Agaian S, Madhukar M, Chronopoulos AT (2014) Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Systems journal 8:995–1004

Alsalem MA, Zaidan AA, Zaidan BB, et al (2018) Systematic review of an automated multiclass detection and classification system for acute Leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects. Journal of medical systems 42:204

Ayenu-Prah A, Attoh-Okine N (2010) A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition. Advances in Adaptive Data Analysis 2:1–24

Bach F (2017) Breaking the curse of dimensionality with convex neural networks. Journal of Machine Learning Research 18:1–53

Barschdorff D, Femmer U (1995) Signal processing and pattern recognition methods for biomedical sound analysis. In: 2nd International Symposium Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques. p 279

Baudin M (2009) Nelder mead user's manual. Digiteo

Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PloS one 12:e0177678

Casoli P, Pastori M, Scolari F, Rundo M (2019) A vibration signal-based method for fault identification and classification in hydraulic axial piston pumps. Energies 12:953

Chakraborty T (2017) EC3: Combining Clustering and Classification for Ensemble Learning. In: 2017 IEEE International Conference on Data Mining (ICDM). pp 781–786

Chawathe SS (2019) Condition monitoring of hydraulic systems by classifying sensor data streams. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, pp 0898–0904

Cheng Y, Wang Z, Chen B, et al (2019) An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis. ISA transactions

Chiang J-H, Ho S-H (2008) A combination of rough-based feature selection and RBF neural network for classification using gene expression data. IEEE transactions on nanobioscience 7:91–99

Chicco D (2017) Ten quick tips for machine learning in computational biology. BioData mining 10:35

Cho M-Y, Hoang TT (2017) Feature selection and parameters optimization of svm using particle swarm optimization for fault classification in power distribution systems. Computational intelligence and neuroscience 2017:

Cho S, May G, Tourkogiorgis I, et al (2018) A Hybrid Machine Learning Approach for Predictive Maintenance in Smart Factories of the Future. In: IFIP International Conference on Advances in Production Management Systems. Springer, pp 311–317

Çınar ZM, Abdussalam Nuhu A, Zeeshan Q, et al (2020a) Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability 12:8211

Çınar ZM, Abdussalam Nuhu A, Zeeshan Q, et al (2020b) Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. Sustainability 12:8211. https://doi.org/10.3390/su12198211

Cohen L (1989) Time-frequency distributions-a review. Proceedings of the IEEE 77:941–981

Colominas MA, Schlotthauer G, Torres ME (2014) Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomedical Signal Processing and Control 14:19–29

Daubechies I (1989) Orthonormal bases of wavelets with finite support—connection with discrete filters. In: Wavelets. Springer, pp 38–66

De Paz JF, Bajo J, López VF, Corchado JM (2013) Biomedic Organizations: An intelligent dynamic architecture for KDD. Information Sciences 224:49–61

Di Z, Kang Q, Peng D, Zhou M (2019) Density Peak-based Pre-clustering Support Vector Machine for Multi-class Imbalanced Classification. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). pp 27–32

Dorigo M, Birattari M, Stutzle T (2006) Ant colony optimization. IEEE Computational Intelligence Magazine 1:28–39. https://doi.org/10.1109/MCI.2006.329691

Du D, Jia X, Hao C (2016) A New Least Squares Support Vector Machines Ensemble Model for Aero Engine Performance Parameter Chaotic Prediction. Mathematical Problems in Engineering 2016:

Egusquiza M, Egusquiza E, Valero C, et al (2018) Advanced condition monitoring of Pelton turbines. Measurement 119:46–55

Galli AW, Heydt GT, Ribeiro PF (1996) Exploring the power of wavelet analysis. IEEE Computer Applications in Power 9:37–41

Gomes JP, Rodrigues LR, Leao BP, et al (2016) Using degradation messages to predict hydraulic system failures in a commercial aircraft. IEEE Transactions on Automation Science and Engineering 15:214–224

718 Haddar M (2018) L-Kurtosis and Improved Complete Ensemble EMD in Early Fault Detection Under
719 Variable Load and Speed. In: Advances in Acoustics and Vibration II: Proceedings of the Second
720 International Conference on Acoustics and Vibration (ICAV2018), March 19-21, 2018, Hammamet,
721 Tunisia. Springer, p 3

722 Hao W, Dong Y, Cai Y (2020) Design of Safety Monitoring System for Stage Machinery System. In: 2020
723 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). pp
724 320–324

725 Har-Peled S, Indyk P, Motwani R (2012) Approximate nearest neighbor: Towards removing the curse of
726 dimensionality. Theory of computing 8:321–350

727 Hathaway BJ, Garde K, Mantell SC, Davidson JH (2018) Design and characterization of an additive
728 manufactured hydraulic oil cooler. International Journal of Heat and Mass Transfer 117:188–200

729 Helwig N, Pignanelli E, Schütze A (2015a) Condition monitoring of a complex hydraulic system using
730 multivariate statistics. In: Instrumentation and Measurement Technology Conference (I2MTC),
731 2015 IEEE International. IEEE, pp 210–215

732 Helwig N, Pignanelli E, Schütze A (2015b) Detecting and Compensating Sensor Faults in a Hydraulic
733 Condition Monitoring System. In: In prpceeding Sensor. SENSOR, pp 641–646

734 Herrmann FJ, Friedlander MP, Yilmaz O (2012) Fighting the curse of dimensionality: Compressive sensing
735 in exploration seismology. IEEE Signal Process Mag 29:88–100

736 Hlawatsch F, Boudreaux-Bartels GF (1992) Linear and quadratic time-frequency signal representations. IEEE
737 signal processing magazine 9:21–67

738 Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations.
739 International Journal of Data Mining & Knowledge Management Process 5:1

740 Houle ME, Kriegel H-P, Kröger P, et al (2010) Can shared-neighbor distances defeat the curse of
741 dimensionality? In: International Conference on Scientific and Statistical Database Management.
742 Springer, pp 482–500

743 Huang NE, Shen Z, Long SR, et al (1998) The empirical mode decomposition and the Hilbert spectrum for
744 nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series
745 A: Mathematical, Physical and Engineering Sciences 454:903–995

746 Ilonen J, Kamarainen J-K, Lampinen J (2003) Differential Evolution Training Algorithm for Feed-Forward
747 Neural Networks. Neural Processing Letters 17:93–105. https://doi.org/10.1023/A:1022995128597

748 Javed K (2014) A robust & reliable Data-driven prognostics approach based on extreme learning machine
749 and fuzzy clustering. PhD Thesis

750 Karaboga D, Gorkemli B, Ozturk C, Karaboga N (2014) A comprehensive survey: artificial bee colony (ABC)
751 algorithm and applications. Artif Intell Rev 42:21–57. https://doi.org/10.1007/s10462-012-9328-0

752 Karanović V, Jocanović M, Baloš S, et al (2019) Impact of contaminated fluid on the working performances
753 of hydraulic directional control valves. Journal of Mechanical Engineering

754 Kaur H, Kaur M (2020) Fault Classification in a Transmission Line Using Levenberg–Marquardt Algorithm
755 Based Artificial Neural Network. In: Data Communication and Networks. Springer, pp 119–135

756 Keerthi SS, Lin C-J (2003) Asymptotic behaviors of support vector machines with Gaussian kernel. Neural
757 computation 15:1667–1689

758 Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95 - International
759 Conference on Neural Networks. pp 1942–1948 vol.4

760 Keogh E, Mueen A (2011) Curse of dimensionality. In: Encyclopedia of machine learning. Springer, pp 257–
761 258

762 König C, Helmi AM (2020) Sensitivity Analysis of Sensors in a Hydraulic Condition Monitoring System
763 Using CNN Models. Sensors 20:3307. https://doi.org/10.3390/s20113307

764 Kuncheva LI, Arnaiz-González Á, Díez-Pastor J-F, Gunn IAD (2018) Instance Selection Improves Geometric
765 Mean Accuracy: A Study on Imbalanced Data Classification. arXiv:180407155 [cs, stat]

Laosai J, Chamnongthai K (2014) Acute leukemia classification by using SVM and K-Means clustering. 2014 International Electrical Engineering Congress (iEECON) 1–4. https://doi.org/10.1109/ieecon.2014.6925840

Lei Y, Jiang W, Jiang A, et al (2019) Fault diagnosis method for hydraulic directional valves integrating PCA and XGBoost. Processes 7:589

Leon-Quiroga J, Newell B, Krishnamurthy M, et al (2020) Energy Efficiency Comparison of Hydraulic Accumulators and Ultracapacitors. Energies 13:1632

Mahgoun H, Chaari F, Felkaoui A, Haddar M (2018) L-Kurtosis and Improved Complete Ensemble EMD in Early Fault Detection Under Variable Load and Speed. In: International Conference on Acoustics and Vibration. Springer, pp 3–15

Marwala T (2012) Condition monitoring using computational intelligence methods: applications in mechanical and electrical systems. Springer Science & Business Media

Mohapatra P, Chakravarty S (2015) Modified PSO based feature selection for Microarray data classification. In: 2015 IEEE Power, Communication and Information Technology Conference (PCITC). IEEE, pp 703–709

Mosavi A, Edalatifar M (2018) A Hybrid Neuro-Fuzzy Algorithm for Prediction of Reference Evapotranspiration. In: International Conference on Global Research and Education. Springer, pp 235–243

Nelder JA, Mead R (1965) A simplex method for function minimization. The computer journal 7:308–313

Newland DE (2005) An Introduction to Random Vibrations, Spectral & Wavelet Analysis: Third Edition, Third edition. Dover Publications, Mineola, N.Y

Niu G, Shang F, Krishnamurthy M, Garcia JM (2016) Evaluation and selection of accumulator size in electric-hydraulic hybrid (EH2) powertrain. In: 2016 IEEE Transportation Electrification Conference and Expo (ITEC). IEEE, pp 1–6

Paul S, Maji P (2010) Rough set based gene selection algorithm for microarray sample classification. In: 2010 International Conference on Methods and Models in Computer Science (ICM2CS-2010). IEEE, pp 7–13

Peppin RJ (1994) An Introduction to Random Vibrations, Spectral and Wavelet Analysis. Hindawi

Pfeffer A, Glück T, Kemmetmüller W, Kugi A (2016) Mathematical modelling of a hydraulic accumulator for hydraulic hybrid drives. Mathematical and Computer Modelling of Dynamical Systems 22:397–411

Pugin KG (2020) Improving the reliability of hydraulic systems of technological machines. IOP Conf Ser: Mater Sci Eng 971:052042. https://doi.org/10.1088/1757-899X/971/5/052042

Qasem M, Nour M (2015) Improving Accuracy for Classifying Selected Medical Datasets with Weighted Nearest Neighbors and Fuzzy Nearest Neighbors Algorithms. In: 2015 International Conference on Cloud Computing (ICCC). IEEE, pp 1–9

Quatrini E, Costantino F, Pocci C, Tronci M (2020) Predictive model for the degradation state of a hydraulic system with dimensionality reduction. Procedia Manufacturing 42:516–523

Randall RB, Antoni J (2011) Rolling element bearing diagnostics—A tutorial. Mechanical systems and signal processing 25:485–520

Raptodimos Y, Lazakis I (2018) Using artificial neural network-self-organising map for data clustering of marine engine condition monitoring applications. Ships and Offshore Structures 13:649–656

Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Information sciences 179:2232–2248

Sarkar C (2015) Improving Predictive Modeling in High Dimensional, Heterogeneous and Sparse Health Care Data

Schneider T, Helwig N, Schütze A (2018) Automatic feature extraction and selection for condition monitoring and related datasets. In: 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). IEEE, pp 1–6

Schneider T, Helwig N, Schütze A (2017) Automatic feature extraction and selection for classification of cyclical time series data. tm-Technisches Messen 84:198–206

Sheng C, Li Z, Qin L, et al (2011) Recent progress on mechanical condition monitoring and fault diagnosis. Procedia Engineering, 15, 142-146. 15:142–146

Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural processing letters 9:293–300

Suykens JAK, Gestel TV, De Brabanter J, et al (2002) Least squares support vector machines. World Scientific

Tai W-L, Hu R-M, Hsiao HC, et al (2011) Blood cell image classification based on hierarchical SVM. In: 2011 IEEE International Symposium on Multimedia. IEEE, pp 129–136

Torres ME, Colominas MA, Schlotthauer G, Flandrin P (2011) A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 4144–4147

Vianna WOL, de Souza Ribeiro LG, Yoneyama T (2015) Electro hydraulic servovalve health monitoring using fading extended Kalman filter. In: 2015 IEEE Conference on Prognostics and Health Management (PHM). IEEE, pp 1–6

Wang K, Wen X, Hou D, et al (2018) Application of Least-Squares Support Vector Machines for Quantitative Evaluation of Known Contaminant in Water Distribution System Using Online Water Quality Parameters. Sensors 18:938

Wang L (2017) Heterogeneous Data and Big Data Analytics. Automatic Control and Information Sciences 3:8–15. https://doi.org/10.12691/acis-3-1-3

Wang X, Lin S, Wang S (2016) Remaining useful life prediction model based on contaminant sensitivity for aviation hydraulic piston pump. In: 2016 IEEE International Conference on Aircraft Utility Systems (AUS). IEEE, pp 266–272

Wen X (2011) A hybrid intelligent technique for induction motor condition monitoring. PhD Thesis, University of Portsmouth

Whitley D (1994) A genetic algorithm tutorial. Stat Comput 4:65–85. https://doi.org/10.1007/BF00175354

Wigner EP (1932) On the quantum correction for thermodynamic equilibrium. In: Part I: Physical Chemistry. Part II: Solid State Physics. Springer, pp 110–120

Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in adaptive data analysis 1:1–41

Xavier-de-Souza S, Suykens JA, Vandewalle J, Bollé D (2009) Coupled simulated annealing. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40:320–335

Xu G, Ma C, Gao Z, et al (2017a) Modeling and simulation of aero-hydraulic pump wear failure. In: 2017 Prognostics and System Health Management Conference (PHM-Harbin). IEEE, pp 1–7

Xu G, Ma C, Gao Z, et al (2017b) Modeling and simulation of aero-hydraulic pump wear failure. In: 2017 Prognostics and System Health Management Conference (PHM-Harbin). IEEE, pp 1–7

Zhang J, Guo Y, Shen Y, et al (2018) Improved CEEMDAN–wavelet transform de-noising method and its application in well logging noise reduction. Journal of Geophysics and Engineering 15:775–787

Zhang X-Y, Yang P, Zhang Y-M, et al (2014) Combination of classification and clustering results with label propagation. IEEE Signal Processing Letters 21:610–614

Zhang Y, Randall RB (2009) Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram. Mechanical Systems and Signal Processing 23:1509–1517

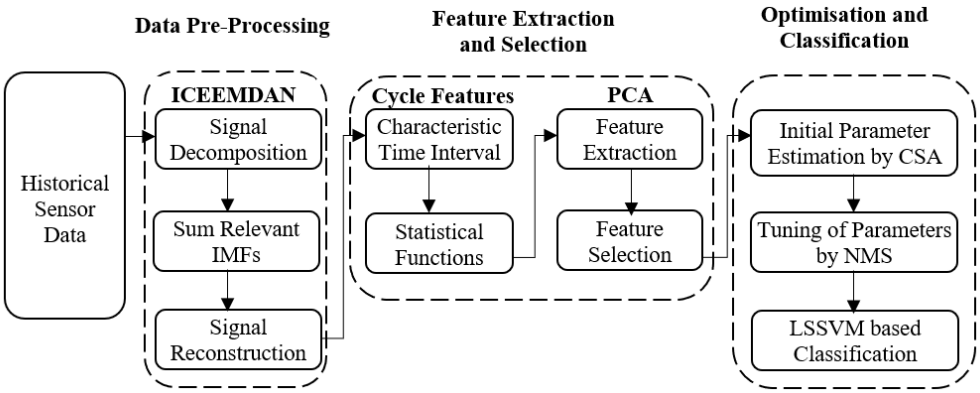
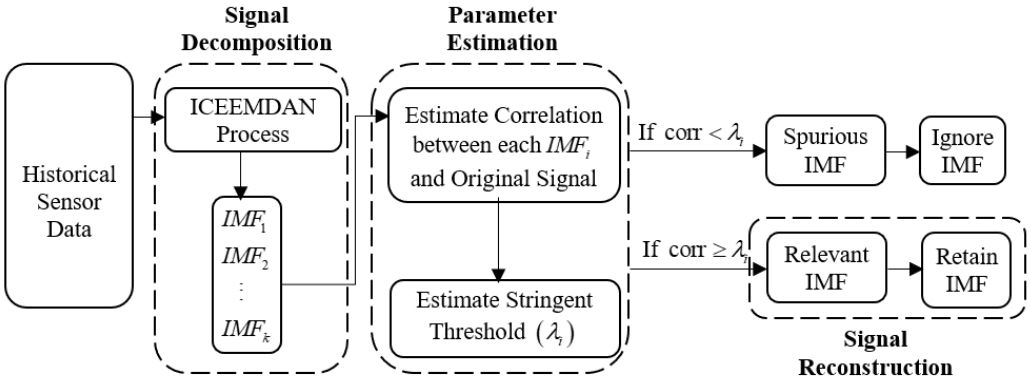
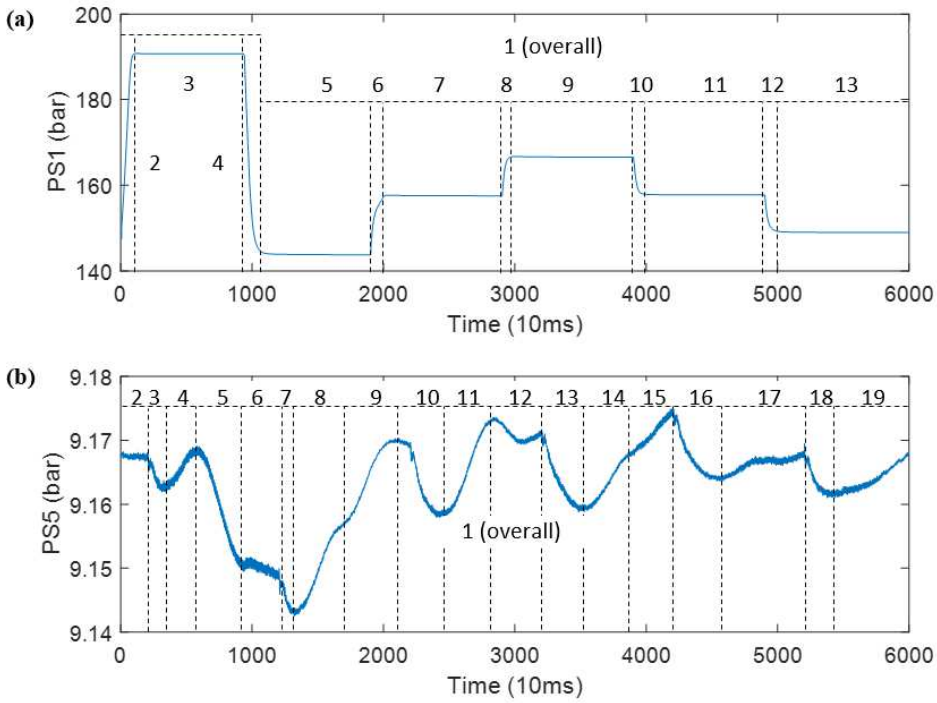



**Fig. 1** Systematic diagram of the proposed hybrid novel technique




861
862 **Fig. 2** Systematic diagram of data pre-processing




863

864 **Fig. 3** Partition of (a) PS1 and (b) PS5 sensor data into time-intervals

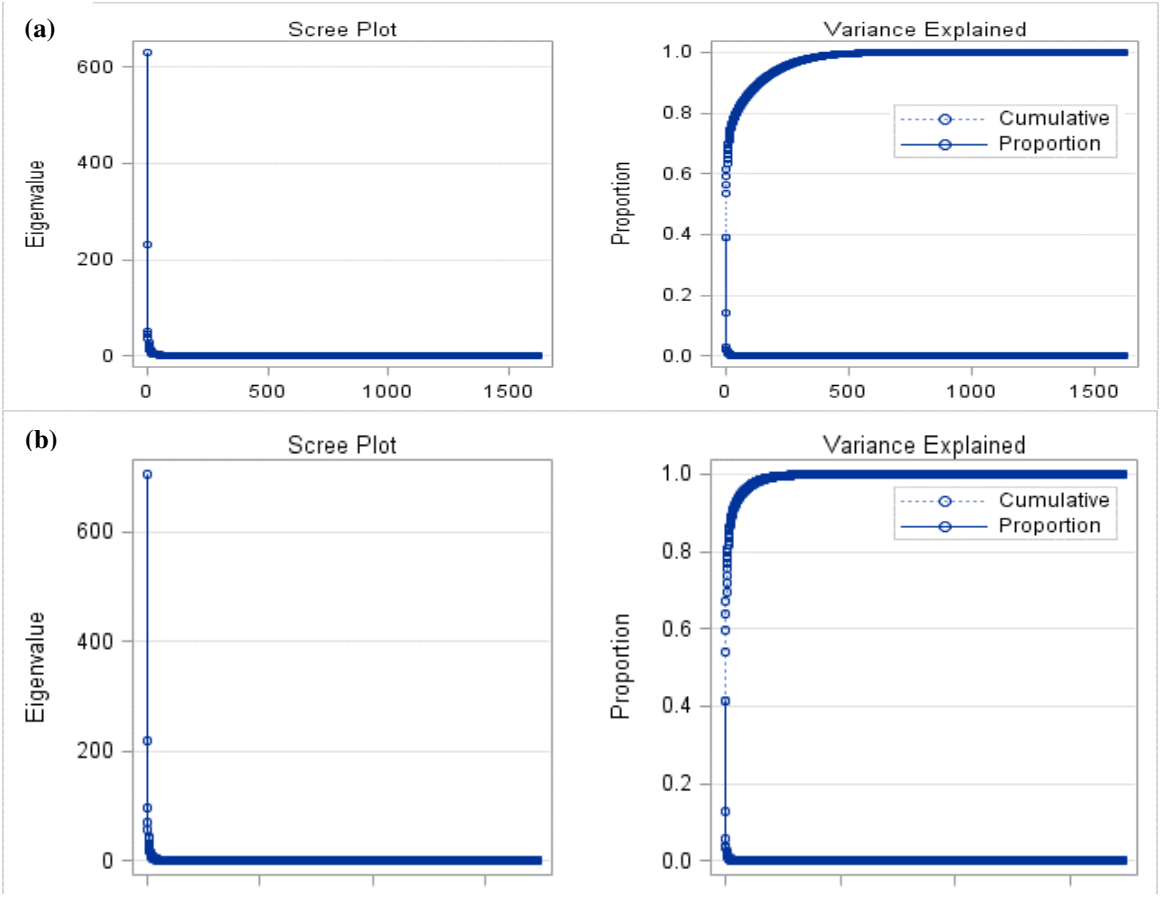



**Fig. 4** Scree plot and variance explained for (a) pre-processed without ICEEMDAN (b) pre-processed with ICEEMDAN

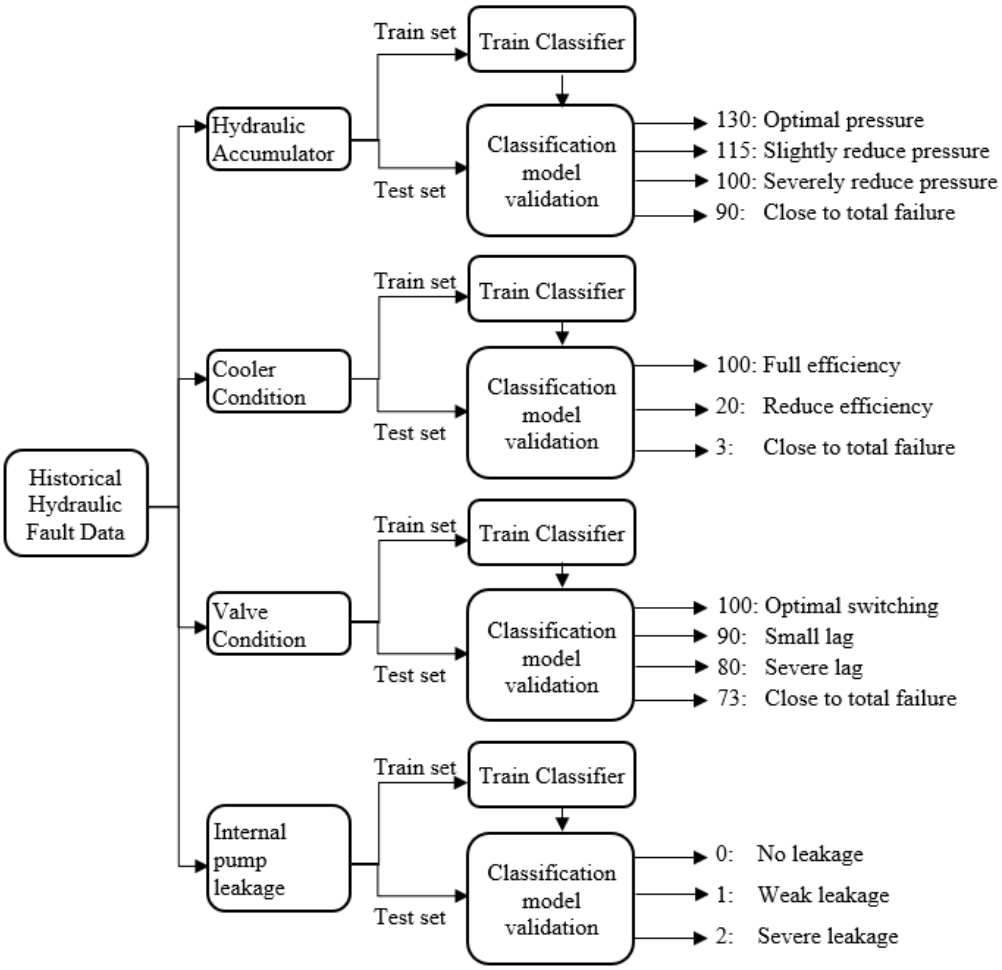



**Fig. 5** Multiclass classification procedure

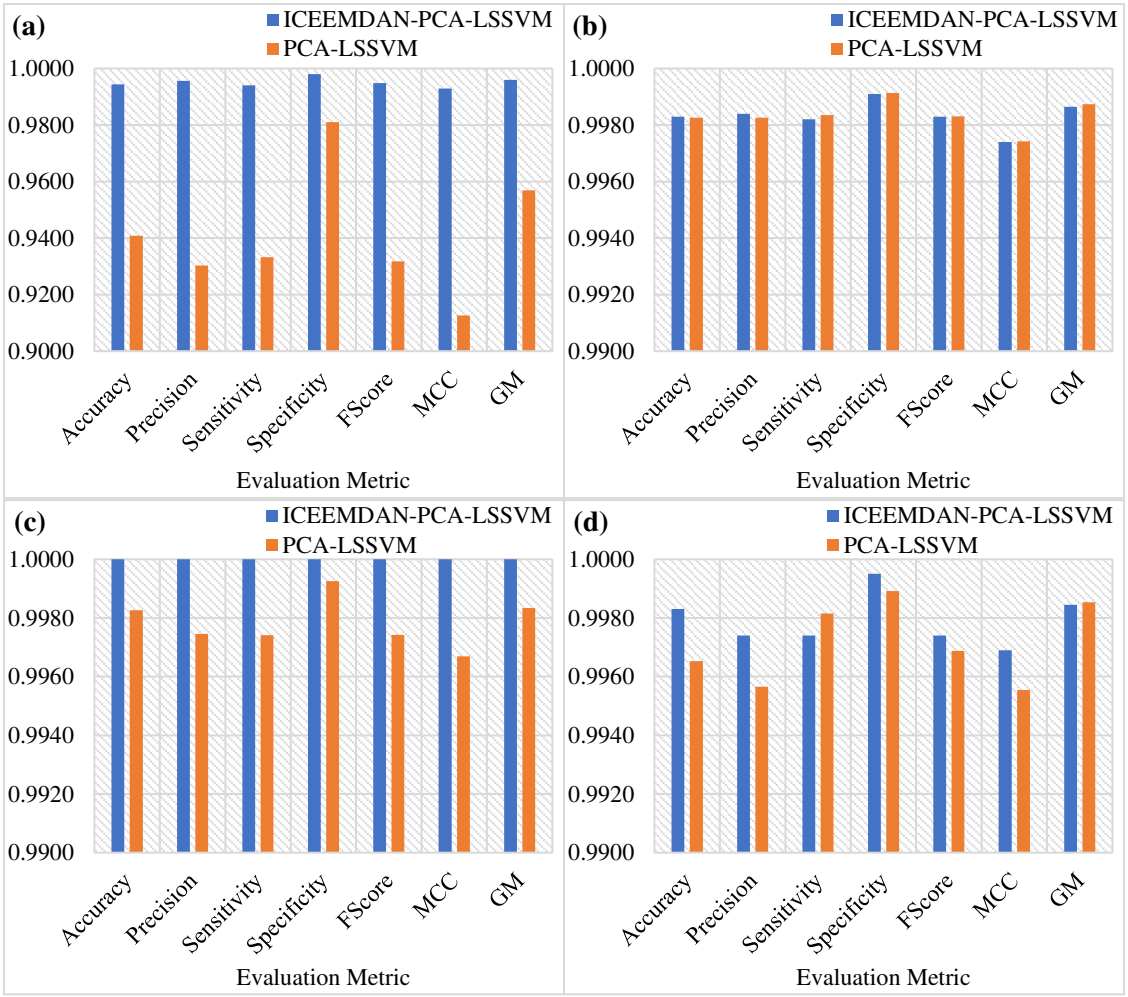



**Fig. 6** Classification performance of ICEEMDAN-PCA-LSSVM and PCA-LSSVM (without ICEEMDAN) for (a) accumulator, (b) cooler, (c) internal pump leakage, and (d) valve conditions

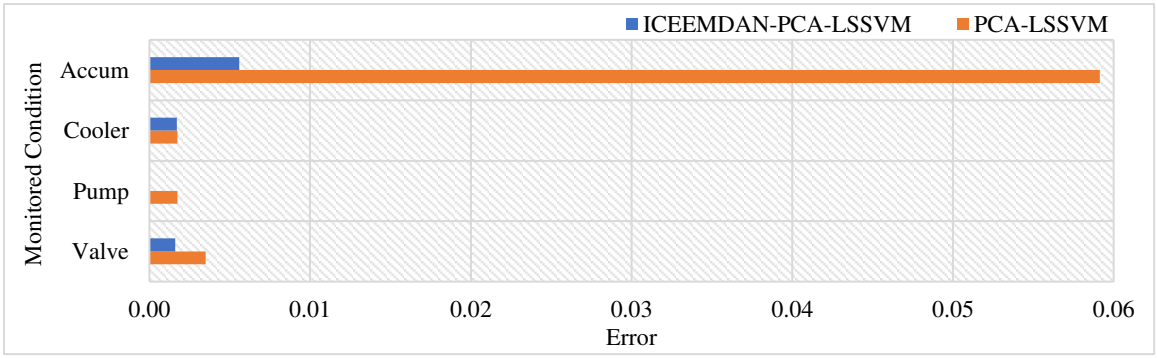



**Fig. 7** Misclassification rate (error) of ICEEMDAN-PCA-LSSVM and PCA-LSSVM (without ICEEMDAN)

**Table 1** Summary of related works

| Year | Objectives | ML Technique Used | | | Key Findings | Reference |
|---|---|---|---|---|---|---|
| | | Pre-processing | | Classifier | | |
| | | Extraction | Selection | | | |
| 2015 | • Use scatter-based ML with automated feature extraction and selection capability to develop an adaptive approach for fault classification | • Signal shape feature functions (shape slope of linear fit, position of maximum value)<br>• Distribution feature functions (median, variance, skewness, kurtosis) | • Pearson correlation analysis | • LDA<br>• ANN<br>• SVM (with RBF and linear kernel) | • developed an automated feature extraction and selection technique for relevant fault features<br>• developed a flexible condition monitoring<br>• system for the classification accumulator, cooler pump and valve conditions. | (Helwig et al. 2015a) |
| 2015 | • Extend prior work of Helwig et al. (2015) in fault diagnosis and the detection of sensor malfunction | • Signal shape feature functions (shape slope of linear fit, position of maximum value)<br>• Distribution feature functions (median, variance, skewness, kurtosis) | • Pearson correlation analysis<br>• LDA | • KNN | • The proposed approach is versatile and capable of detecting sensor malfunctions.<br>• Compensates the failure of up to 5 sensors with negligible information loss and reduction in the classification result | (Helwig et al. 2015b) |
| 2017 | • Automate the reduction in dimension by applying four complimentary feature extraction methods and three features selection algorithms | • ALA<br>• PCA<br>• BFC<br>• BDW | • Pearson correlation analysis<br>• RFESVM<br>• Univariate RELIEF | • LDA | • Reduced the manual effort in feature extraction.<br>• The classification of accumulator condition was improved whiles the others were comparatively similar to the results obtained by Helwig et al. (2015) | (Schneider et al. 2017) |
| 2019 | • Sought to improve the classification accuracy of prior work.<br>• Investigate a trade-off between the number of features and accuracy | • Distribution feature functions (mean, variance, skewness, kurtosis) | • None<br>• OneR<br>• J48<br>• Information gain | • ZeroR<br>• OneR<br>• JRip<br>• PART<br>• J48<br>• RF<br>• NB | • Higher accuracy is achieved with the proposed method than prior work.<br>• Higher accuracy is still achieved with a lesser number of features | (Chawathe 2019) |
| 2020 | • Develop a predictive model for the degradation of major components of the hydraulic system | • Signal shape feature functions (shape slope of linear fit, position of maximum value)<br>• Distribution feature functions (median, variance, skewness, kurtosis) | • Pearson correlation analysis<br>• LDA | • ANN<br>• SVM<br>• LR<br>• DF | • High accuracy is achieved when reduction in features allows for optimal combination of inputs that linearly separates the groups while minimising inter-group distance.<br>• Selecting features that are highly correlated with target outputs enhances performance | (Quatrini et al. 2020) |
| 2020 | • Predict different levels of degradation of major components of the hydraulic system | • Encodings of the convolution layers | • Encodings of the convolution layers | • CNN | • developed a deep learning-based condition monitoring system for the hydraulic system dataset without explicitly engineering the features | (König and Helmi 2020) |

*NB: Classifiers abbreviations are as follows: Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Radial Basis Function (RBF), Logistic Regression (LR), Decision Forest (DF), Random Forest (RF), Naïve Bayes (NB), Convolutional Neural Networks (CNN), K*

*Nearest Neighbour (KNN), Adaptive Linear Approximation (ALA), Principal Component Analysis (PCA), Best Fourier Coefficients (BFC), Best Daubechies Wavelets (BDW), Recursive Feature Elimination Support Vector Machines (RFESVM)*

885 **Table 2** Details of hydraulic system dataset

| Type | Physical Quantity | Sensor | Unit | Sampling Rate (Hz) | Class | Time Interval |
|---|---|---|---|---|---|---|
| Sensor | Pressure | PS1 | bar | 100 | Regression | 13 |
| | | PS2 | | | | 14 |
| | | PS3 | | | | 18 |
| | | PS4 | | | | 25 |
| | | PS5 | | | | 19 |
| | | PS6 | | | | 19 |
| | Motor Power | EPS1 | W | | | 13 |
| | Volume Flow | FS1 | l/min | 10 | | 18 |
| | | FS2 | | | | 18 |
| | Temperature | TS1 | °C | 1 | | 7 |
| | | TS2 | | | | 7 |
| | | TS3 | | | | 8 |
| | | TS4 | | | | 15 |
| | Vibration | VS1 | mm/s | | | 15 |
| | Cooling Efficiency | CE | % | | | 13 |
| | Cooling Power | CP | kW | | | 13 |
| | System Efficiency | SE | % | | | 23 |
| Hydraulic Component | Accumulator | - | bar | - | Classification | |
| | Cooler | | % | | | |
| | Internal Pump Leakage | | - | | | |
| | Valve | | % | | | |

886 Source: Modified after Helwig et al. (2015)

887 **Table 3** Summary of PCA results of denoised and undenoised data

| Pre-Processed without ICEEMDAN | | | | Pre-Processed with ICEEMDAN | | | |
|---|---|---|---|---|---|---|---|
| PC | Eigenvalue | Proportion | Cumulative | PC | Eigenvalue | Proportion | Cumulative |
| 1 | 630.4540 | 0.3906 | 0.3906 | 1 | 705.0536 | 0.4145 | 0.4145 |
| 2 | 232.4312 | 0.1440 | 0.5346 | 2 | 220.0692 | 0.1294 | 0.5439 |
| 3 | 50.9179 | 0.0315 | 0.5662 | 3 | 95.9180 | 0.0564 | 0.6003 |
| 4 | 44.5339 | 0.0276 | 0.5938 | 4 | 69.1454 | 0.0406 | 0.6409 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 160 | 1.0071 | 0.0006 | 0.9150 | 81 | 1.0160 | 0.0006 | 0.9598 |
| 161 | 1.0024 | 0.0006 | 0.9156 | 82 | 1.0067 | 0.0006 | 0.9604 |
| 162 | 0.9791 | 0.0006 | 0.9162 | 83 | 0.9963 | 0.0006 | 0.9609 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

888

889 **Table 4** Multiclass classification of fault conditions based on ICEEMDAN-PCA-LSSVM

| Hydraulic Component | Training | | Testing | | |
|---|---|---|---|---|---|
| | Accuracy ± STD | CPU Time (sec) | Accuracy ± STD | Best | Worst |
| Accumulator | 1.0000 ± 0.000 | 38.2344 | 0.9944 ± 0.0007 | 0.9948 | 0.9930 |
| Cooler | 1.0000 ± 0.000 | 9.1250 | 0.9983 ± 0.0006 | 1.0000 | 0.9948 |
| Pump | 1.0000 ± 0.000 | 32.4375 | 1.0000 ± 0.0000 | 1.0000 | 1.0000 |
| Valve | 1.0000 ± 0.000 | 22.1094 | 0.9984 ± 0.0005 | 1.0000 | 0.9983 |

890

**Table 5** Accumulator condition classification rate

| Metric | Pre-Processed with ICEEMDAN-PCA | | | | Pre-Processed PCA (without ICEEMDAN) | | | |
|---|---|---|---|---|---|---|---|---|
| | LSSVM | LDA | SVM | ANN | LSSVM | LDA | SVM | ANN |
| Accuracy | 0.9944 | 0.9896 | 0.9913 | 0.9791 | 0.9409 | 0.9270 | 0.8904 | 0.8591 |
| Error | 0.0056 | 0.0104 | 0.0087 | 0.0209 | 0.0591 | 0.0730 | 0.1096 | 0.1409 |
| Precision | 0.9956 | 0.9880 | 0.9913 | 0.9772 | 0.9303 | 0.9185 | 0.8762 | 0.8494 |
| Sensitivity | 0.9941 | 0.9871 | 0.9901 | 0.9778 | 0.9333 | 0.9239 | 0.8798 | 0.8400 |
| Specificity | 0.9980 | 0.9966 | 0.9971 | 0.9931 | 0.9810 | 0.9766 | 0.9644 | 0.9527 |
| F Score | 0.9948 | 0.9874 | 0.9906 | 0.9775 | 0.9317 | 0.9192 | 0.8772 | 0.8438 |
| MCC | 0.9929 | 0.9842 | 0.9878 | 0.9705 | 0.9126 | 0.8970 | 0.8420 | 0.7977 |
| GM | 0.9960 | 0.9918 | 0.9936 | 0.9854 | 0.9569 | 0.9499 | 0.9212 | 0.8945 |

893 **Table 6** Cooler condition classification rate

| Metric | Pre-Processed with ICEEMDAN-PCA | | | | Pre-Processed PCA (without ICEEMDAN) | | | |
|---|---|---|---|---|---|---|---|---|
| | LSSVM | LDA | SVM | ANN | LSSVM | LDA | SVM | ANN |
| Accuracy | 0.9983 | 0.9948 | 0.9948 | 0.9983 | 0.9983 | 0.9948 | 0.9965 | 0.9965 |
| Error | 0.0017 | 0.0052 | 0.0052 | 0.0017 | 0.0017 | 0.0052 | 0.0035 | 0.0035 |
| Precision | 0.9984 | 0.9949 | 0.9949 | 0.9983 | 0.9983 | 0.9950 | 0.9964 | 0.9966 |
| Sensitivity | 0.9982 | 0.9949 | 0.9949 | 0.9983 | 0.9983 | 0.9948 | 0.9966 | 0.9966 |
| Specificity | 0.9991 | 0.9974 | 0.9974 | 0.9991 | 0.9991 | 0.9973 | 0.9983 | 0.9982 |
| F Score | 0.9983 | 0.9949 | 0.9949 | 0.9983 | 0.9983 | 0.9949 | 0.9965 | 0.9966 |
| MCC | 0.9974 | 0.9922 | 0.9922 | 0.9974 | 0.9974 | 0.9922 | 0.9948 | 0.9948 |
| GM | 0.9986 | 0.9961 | 0.9961 | 0.9987 | 0.9987 | 0.9961 | 0.9974 | 0.9974 |

894

**Table 7** Internal pump leakage condition classification rate

| Metric | Pre-Processed with ICEEMDAN-PCA | | | | Pre-Processed PCA (without ICEEMDAN) | | | |
|---|---|---|---|---|---|---|---|---|
| | LSSVM | LDA | SVM | ANN | LSSVM | LDA | SVM | ANN |
| Accuracy | 1.0000 | 0.9861 | 1.0000 | 0.9704 | 0.9983 | 0.9948 | 0.9878 | 0.9165 |
| Error | 0.0000 | 0.0139 | 0.0000 | 0.0296 | 0.0017 | 0.0052 | 0.0122 | 0.0835 |
| Precision | 1.0000 | 0.9807 | 1.0000 | 0.9597 | 0.9975 | 0.9924 | 0.9825 | 0.8832 |
| Sensitivity | 1.0000 | 0.9793 | 1.0000 | 0.9623 | 0.9974 | 0.9953 | 0.9819 | 0.8841 |
| Specificity | 1.0000 | 0.9940 | 1.0000 | 0.9862 | 0.9993 | 0.9978 | 0.9948 | 0.9619 |
| F Score | 1.0000 | 0.9794 | 1.0000 | 0.9610 | 0.9974 | 0.9938 | 0.9820 | 0.8836 |
| MCC | 1.0000 | 0.9739 | 1.0000 | 0.9467 | 0.9967 | 0.9911 | 0.9769 | 0.8453 |
| GM | 1.0000 | 0.9866 | 1.0000 | 0.9742 | 0.9983 | 0.9965 | 0.9883 | 0.9222 |

**Table 8** Valve condition classification rate

| Metric | Pre-Processed with ICEEMDAN-PCA | | | | Pre-Processed PCA (without ICEEMDAN) | | | |
|---|---|---|---|---|---|---|---|---|
| | LSSVM | LDA | SVM | ANN | LSSVM | LDA | SVM | ANN |
| Accuracy | 0.9984 | 0.9983 | 0.9965 | 0.9548 | 0.9965 | 0.9948 | 0.9930 | 0.8783 |
| Error | 0.0016 | 0.0017 | 0.0035 | 0.0452 | 0.0035 | 0.0052 | 0.0070 | 0.1217 |
| Precision | 0.9976 | 0.9974 | 0.9948 | 0.9451 | 0.9957 | 0.9924 | 0.9921 | 0.8536 |
| Sensitivity | 0.9977 | 0.9974 | 0.9948 | 0.9469 | 0.9982 | 0.9953 | 0.9929 | 0.8649 |
| Specificity | 0.9995 | 0.9995 | 0.9990 | 0.9843 | 0.9989 | 0.9978 | 0.9976 | 0.9584 |
| F Score | 0.9976 | 0.9974 | 0.9948 | 0.9457 | 0.9969 | 0.9938 | 0.9925 | 0.8585 |
| MCC | 0.9972 | 0.9969 | 0.9938 | 0.9300 | 0.9955 | 0.9911 | 0.9899 | 0.8160 |
| GM | 0.9986 | 0.9985 | 0.9969 | 0.9654 | 0.9985 | 0.9965 | 0.9952 | 0.9105 |