IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 29, 2020, accepted December 8, 2020, date of publication December 11, 2020,
date of current version December 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044195

Multistage Centrifugal Pump Fault Diagnosis by Selecting Fault Characteristic Modes of Vibration and Using Pearson Linear Discriminant Analysis

ZAHOOR AHMAD, ALEXANDER E. PROSVIRIN, JAEYOUNG KIM, (Member, IEEE), AND JONG-MYON KIM, (Member, IEEE)
School of Electrical, Electronics, and Computer Engineering, University of Ulsan, Ulsan 44610, South Korea
Corresponding author: Jong-Myon Kim (jmkim07@ulsan.ac.kr)

This work was supported by the Ministry of Education and National Research Foundation of Korea through the "Leaders in INdustry-university Cooperation +" Project.

ABSTRACT This paper proposes a three-stage fault diagnosis strategy for multistage centrifugal pumps. First, the proposed method identifies and selects fault characteristic modes of vibration to overcome the substantial noise produced by other unrelated macro-structural vibrations. In the second stage, raw hybrid statistical features are extracted from the fault characteristic modes of vibration in time, frequency, and the time-frequency domain. These extracted features result in a high-dimensional feature space. However, in general, not all of the features are best to characterize the ongoing processes in a centrifugal pump, and some of the extracted features might be irrelevant or even redundant, which can affect the fault classification capabilities of the classification algorithm. In the third stage, a novel dimensionality reduction technique, called Pearson Linear Discriminant Analysis (PLDA), is introduced. PLDA assesses the helpfulness of the feature parameters. This technique selects highly interclass-correlated features and adds them to a helpful feature pool. To achieve maximum intraclass separation while maintaining the original class information, linear discriminant analysis is then applied to the helpful feature pool. This combination of helpful feature pool formation and linear discriminant analysis forms the proposed application of PLDA. The reduced discriminant feature set obtained from PLDA is then classified using the k-nearest neighbor classification algorithm. The proposed method outperforms the previously presented methods in terms of classification accuracy.

INDEX TERMS Multistage centrifugal pump, fault diagnosis, modes of vibration, mechanical faults, Pearson linear discriminant analysis.

I. INTRODUCTION

Centrifugal pumps (CPs) are widely used in various household and industrial processes. A survey made by a major European organization has revealed that around 65% of the grid energy is consumed by machines driven by electric motors out of which CPs account for 80% of the total energy consumption [1], [2]. Due to the versatile applications of the CPs, the appearance of a fault in them may lead to severe consequences such as costly repairs, energy loss, economic losses, industrial process downtime, and reduced safety of the operating staff. To ensure reliable operation of the CPs, early fault diagnosis is of primary importance to minimize repair and maintenance costs.

Faults in the CP can be categorized into two main types: fluid flow related faults and mechanical faults (MFs) [3]. Mechanical seal related faults account for 34% of faults in CPs. In addition to the failure of mechanical seals, impeller faults in the CP can lead to mechanical or both mechanical and flow-related faults [4], [5]. Furthermore, MFs in CPs can cause either hard or soft failures. Hard failures are easy to identify; however, although soft failures cause performance degradation, the CP can remain in function [6]. Soft failures are crucial and must be identified promptly. This paper focuses on the identification of soft faults in the CP, typically due to a mechanical seal hole, mechanical seal scratch, or an impeller defect.

The associate editor coordinating the review of this manuscript and approving it for publication was Youqing Wang.

An MF in the CP changes the stiffness around the mechanical structure which produces a shock or an impulse. This shock results in variations of the amplitudes and distribution of the vibration signals [7]. Therefore, the vibration signal can be efficiently used for diagnosing faults in CPs [8], [9]. Indications of these shocks mainly occur at the specific fault characteristic frequencies (FCF), since these FCFs contain very little energy and are often overwhelmed by macro-structural vibration noise [10], [11]. A narrow band signal demodulation technique, such as multiband envelop detection, has been widely applied for detecting the FCFs [12]. However, it is known that the narrowband signal demodulation struggles with differentiating between fault impulses and interference impulses. To address this issue, the fast kurtogram and spectral kurtosis have been introduced by Antoni [13], [14] as an indicator to select the optimal frequency band for an FCF. Spectral kurtosis focuses on impulsiveness, which makes it sensitive to the outliers or periodic noises. Viet *et al.* [15] used blind source separation techniques for noise reduction in the acoustic emission signals for pressure vessel diagnosis. However, these kinds of techniques need a baseline or reference signals to ensure the appropriate noise reduction in the unimpaired signals. When applied to vibration signals, empirical mode decomposition has a persuasive noise reduction capability; however, its limited mathematical background, the presence of mode mixing, and extreme interpolation make this technique less attractive for solving the problems described above [16], [17]. Specifically, for CPs that consist of different mechanical components, it is hard to focus on one optimal narrow band for the FCF. Therefore, in this study, rather than defining specific frequency ranges, the modes of vibration will be defined where the FCF of CPs can be detected. Additionally, to overcome the macro-structure vibration noise, these specific modes of vibration will be filtered for fault feature extraction.

Signal-based intelligent fault diagnosis methods have been investigated for industrial devices for a long time [18]–[20]. These methods consist of three key steps: signal fault indicators (features) extraction, feature preprocessing, and fault classification [21]–[25]. Feature extraction from the signal is the most crucial step in fault diagnosis. In this step, raw statistical features in the time domain, frequency domain, and time-frequency domain are extracted for fault diagnosis [26]–[34]. However, the basic statistical features are neither sensitive to weak incipient faults nor appropriate for the severe faults [35]. To address this issue, in this paper, features from each domain are extracted and combined into a single feature pool called a raw hybrid feature pool. This hybrid feature pool obtained from the vibration signal in each domain has enough information to segregate different types of mechanical faults. However, raw hybrid statistical features result in a high-dimensional feature space and the conventional features might be both irrelevant or even redundant which can degrade the fault classification performance of the fault diagnosis framework. To enhance the classification accuracy, intrinsic information extraction from a hybrid feature pool or hybrid feature preprocessing is essential [36].

It is worth noting that from the literature review, the most recent studies focus on vibration signal and feature preprocessing for discriminant fault information extraction. In the past decades, several fault feature discrimination evaluation and dimensionality reduction methods have been proposed. Among them, principal component analysis (PCA) [37] and linear discriminant analysis (LDA) [38] are the most prominent ones. Sakthivel *et al.* [39] performed a comparison between dimensionality reduction techniques for CP fault diagnosis and found that PCA demonstrated promising results. For the study case of CP fault diagnosis, PCA results in principal components that carry information about diverse symptoms of machinery faults. However, PCA does not consider intraclass separability estimation. Moreover, information loss is also a significant drawback in PCA. Contrary to PCA, LDA finds an optimal reduced dimensional representation by considering inter-class scatteredness along with intraclass separability, if a sufficiently large labeled data set is provided. Several variants of LDA have been proposed in the last decades such as the trace ratio LDA [40], local sensitive discriminant analysis [41], and robust linear optimized LDA [42]. However, the penalty graph representing intraclass discrimination can affect classification accuracy. The above shortcomings indicate an urgent need to apply the new PLDA technique. PLDA first selects highly interclass-correlated features and add them to a helpful feature pool. To achieve a maximum intraclass separation while maintaining original class information, the LDA is then applied to this helpful feature pool. The combination of the helpful feature pool formation and LDA form the core of PLDA.

The major contributions of this paper are summarized as follows:

1. To overcome the noise from macrostructural vibration, fault characteristic modes of vibration are defined and filtered out from the vibration signal. To the best of the our knowledge, defining fault characteristic modes of vibration for CP fault diagnosis has not been reported so far.
2. PLDA is proposed for obtaining a low dimensional discriminant feature space for CP mechanical faults diagnosis. According to the literature review, the computation of low dimensional discriminant feature space for CP mechanical faults diagnosis using PLDA has not been reported yet.
3. Data collected from a real industrial CP test rig is utilized for validating the proposed fault diagnosis strategy that ensures the comparability of the achieved results with other fault diagnosis approaches.

The remaining parts of this manuscript are organized as follows. Section II provides a brief technical background of LDA. The CP test rig setup and experimental procedure are described in Section III. The proposed method is introduced and discussed in Section IV. The experimental results are presented and explained in Section V. Concluding remarks,

along with a discussion of future research directions, are provided in Section VI.

II. TECHNICAL BACKGROUND

A. MULTICLASS LDA

LDA is a supervised machine learning technique that projects high-dimensional data into a lower-dimensional space while preserving the discriminatory information for each class. The optimal transform matrix in LDA is obtained by maximizing the intraclass distance and minimizing the interclass scatteredness and, as a result, the maximum class discrimination is obtained. In a multiclass classification problem, the interclass scatter matrix w_s and intraclass scatter matrix b_s can be expressed as

$$w_s = \sum_{i=1}^{j} w_i^2 = \sum_{i=1}^{j} \sum_{l_i \in class_i} (l_i - \mu_i) - (l_i - \mu_i)^t, \quad (1)$$

where l_i are the samples of $class_i$, μ_i is the center of $class_i$ in (1).

$$b_s = \sum_{i=1}^{j} n_i(\mu_i - \mu)(\mu_i - \mu)^t \quad (2)$$

$$\mu = \frac{1}{N} \sum_{\forall l} l = \frac{1}{N} \sum_{i=1}^{j} n_i \mu_i. \quad (3)$$

Here, n_i is the sample number of clusters i, N is the sample number per class, and the maximum rank of b_s is $j - 1$ in (2) and (3).

Let T be the transformation matrix which will project the raw hybrid feature space into a lower-dimensional space. Based on (2) and (3) we can express the scatter matrix for the projected feature space as:

$$\tilde{w}_s = T^t w_s T \quad (4)$$

$$\tilde{b}_s = T^t b_s T. \quad (5)$$

Here, $\tilde{w}_s$ is the projected intraclass scatter matrix and $\tilde{b}_s$ is the interclass scatter matrix in (4) and (5). Thus, the objective function can be derived from (4) and (5), as shown in (6).

$$J(T) = \frac{\det(\tilde{w}_s)}{\det(\tilde{b}_s)} \quad (6)$$

$$b_s T = \lambda w_s T, \quad \left(\lambda = \frac{T^t b_s T}{T^t w_s T}\right) \quad (7)$$

Eq. (7) has $j - 1$ linearly independent eigenvectors, shown by $T_1, T_2, \ldots, T_{j-1}$, respectively. The optimal data projection matrix T for the K-dimensional subspace is given by the eigenvectors corresponding to the largest K eigenvalues. Therefore, multiclass LDA can project the original raw hybrid feature space into a subspace of dimension $j - 1$ at most, as explained by [43].

III. TEST RIG SETUP AND PROCEDURE

A CP test rig developed for the fault experiments is shown in Figure 1(a). The test rig consists of a PMT-4008 multistage CP, which is driven by a 5.5kW motor, a control panel, which consists of the temperature, flow rate, water supply controller, speed, an ON/OFF switch, and a crystal display. Two water tanks (a buffer tank and the main tank) were used in the experiment. To maintain the net positive suction head at the CP inlet, the two tanks were kept at a sufficient height. The water tanks were connected to the CP through unobstructed steel pipes with an installed water valve and pressure gauges. The schematics of the test rig are shown in Figure 1(b). After connecting the two tanks at the inlet and outlet of the CP, the CP was operated at a constant speed of 1733 revolutions per minute (RPM). The vibration data from the CP were collected by four accelerometers, two of which were mounted near the pump casing and the other two were mounted near the mechanical seal and near the impellers using adhesive. Each accelerometer recorded the vibration signal of CP using an independent channel. The CP acceleration data was digitized by the National Instrumentation (NI) device (NI 9234). The details for the accelerometers and the NI instrument device used in the experiment are provided in Table 1.

TABLE 1. Specification for Data Collection instruments.

622b01-Accelerometer	**Sensitivity:** (10.2mv/(m/s2) 100mV/g ± 5 **Frequency Range:** from 0.42 – 10kHz
NI-9234 DAQ	**Generator:** 24bits ADC resolution with 4 analog input channels **Frequency Range:** from 0 – 13.1MHz

CP vibration data were acquired for 300 seconds with a sampling rate of 25.6kHz. A total of 1200 samples were collected from the CP. Out of them, 300 samples were collected under the normal operating condition, 300 samples were recorded under mechanical seal hole defect, 300 samples were acquired under mechanical seal scratch defect, and 300 samples were collected under the impeller defect condition. The mechanical seal hole of diameter 2.8 mm and depth of 2.8 mm that has been created in a 38 mm inner diameter CP is shown in Figure 2(a). A scratch of 2.5 mm diameter, having a depth of 2.8 mm, was created in a 38 mm inner diameter portion of the mechanical seal, as presented in Figure 2(b). A 2.5 mm diameter metal piece with a length of 18 mm and a depth of 2.8 mm was removed from the impeller and a defect was created, as shown in Figure 2(c). Each fault was seeded once in the CP and vibration data were collected. The CP vibration response for the various conditions is shown in Figure 3.

IV. PROPOSED APPROACH

The flow of the proposed fault diagnosis strategy for the CP is illustrated in Figure. 4. The proposed strategy starts with CP vibration signal preprocessing and ends with CP mechanical faults classification. Each step of the proposed fault diagnosis strategy is described in detail in this section.

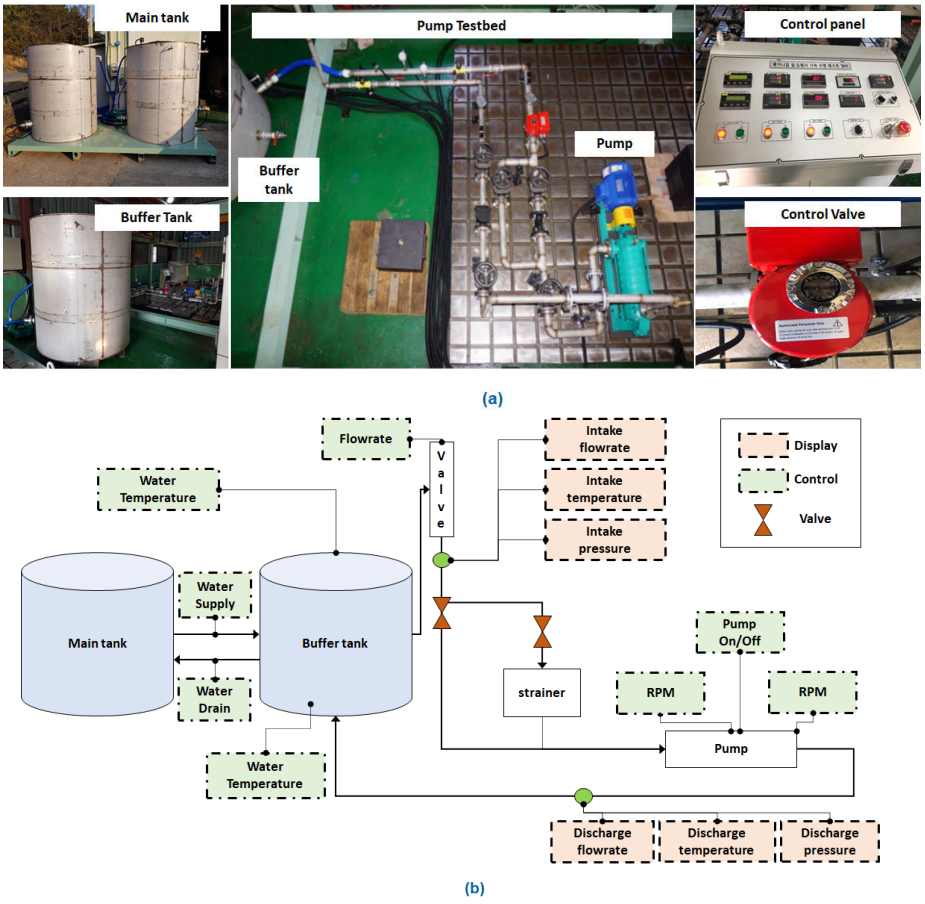

FIGURE 1. CP test rig (a) Pictorial diagram of the CP test rig setup (b) CP test rig schematics.

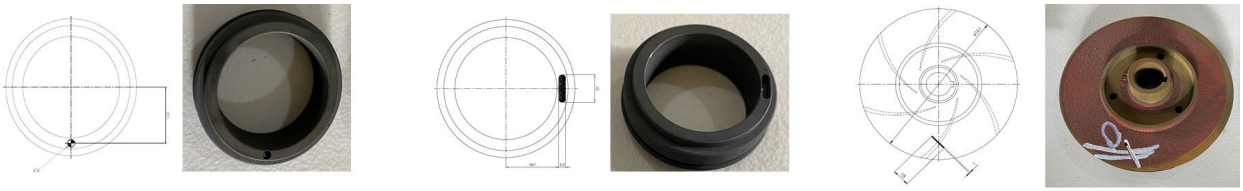

FIGURE 2. Mechanical faults seeded in the CP (a) Mechanical seal hole defect (b) Mechanical seal scratch defect (c) Impeller defect.

A. STEP 1: IDENTIFICATION AND SELECTION OF FAULT CHARACTERISTIC MODES OF VIBRATION

Whenever a mechanical fault occurs in the CP, the stiffness around the mechanical structure changes and a shock is produced. These shocks can be observed in the frequency spectrum at the specific FCFs. However, in the case of the CP, due to the complex fluid and mechanical components interactions, mechanical faults can lead to a hydraulic fault. Therefore, only focusing on the FCF may not help obtain discriminant features of the vibration signal, in addition to this, macrostructural vibration can also affect the quality of the statistical features obtained from the raw vibration signal. To overcome these issues, in this step, fault characteristic modes of vibration are identified and selected for further processing.

There are generated, excitation, and electronics frequencies in the CP. The generated and excitation frequency harmonics are valuable features to identify fault characteristic modes of vibration. As with the generated frequencies, the frequency harmonics generated by the CP due to an imbalance should be understood. These frequencies can be identified in the frequency spectrum if the parameters for the system, such as the rotating speed and the geometry of the CP are known.

In the case of impeller defect, the source frequencies are generated frequencies. Whenever an impeller defect occurs in the CP, an impeller imbalance appears in the vibration signature [5], [44]. This impeller imbalance can be observed at an FCF given as follows:

$$F_{i.defect} = z.N. \tag{8}$$

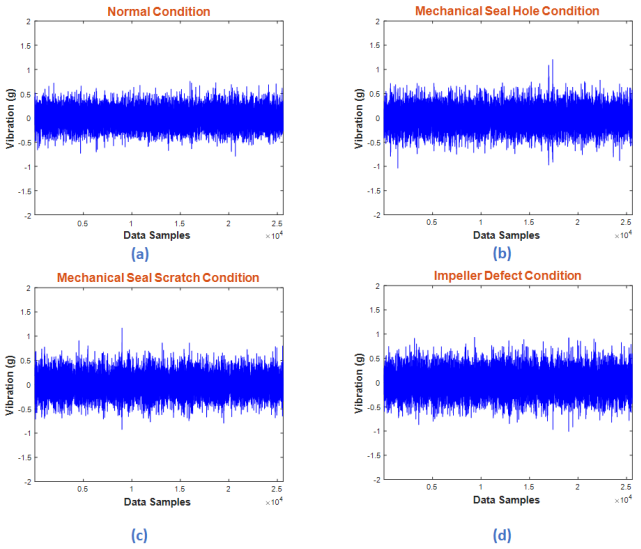

FIGURE 3. CP vibration signals under different operating conditions (a) Normal condition (b) Mechanical seal hole defect condition (c) Mechanical seal scratch defect condition (d) Impeller defect condition.

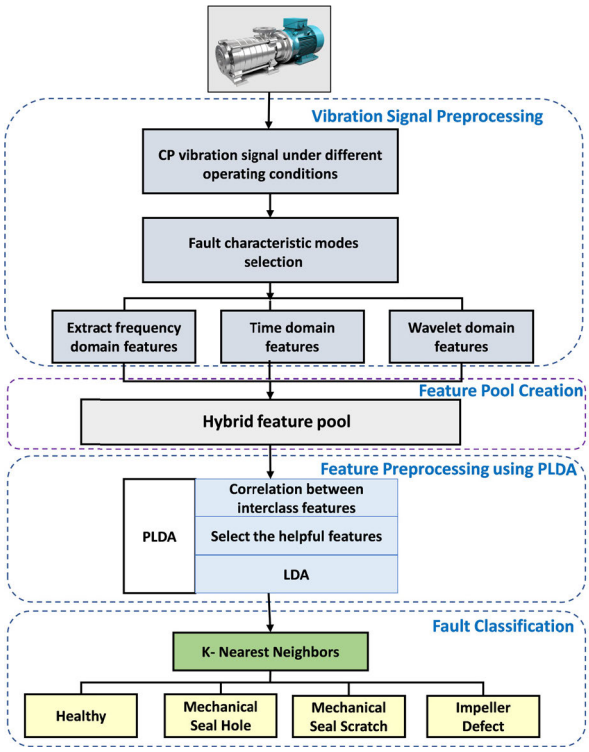

FIGURE 4. Proposed CP fault diagnosis strategy.

Here, z represents frequency harmonics and N is the operating speed of the pump expressed in Hz. Eq. (8) also satisfies ISO standard 13373-1 for rotating machinery [45]. Figure 5(a) presents the frequency spectrum of the CP under the normal operating condition. On the contrary, Figure 5(b) demonstrates the frequency spectrum of the CP under the impeller defect condition. $F_{i.defect}$ were calculated up to the 5th harmonic. It can be observed that the amplitudes of the 3rd, 4th, and 5th harmonics increased because of the impeller defect. Furthermore, some spikes appeared in the frequency spectrum after the impeller defect. The possible reason for this is the interaction of the fluid with the faulty impeller, as can be seen in Figure 5(b).

The excitation frequency in the CP frequency spectrum represents a mechanical seal defect. The excitation frequency is a single frequency harmonic present in the spectrum which reflects the amplified vibration of the CP. It can be calculated from the theory of vibration of a circular ring given in [46], [47].

From the energy conservation method,

$$\frac{d}{dt}((PE + KE) = 0, \tag{9}$$

where PE is the deformation potential energy and KE is the vibration kinetic energy. PE and KE are given by:

$$PE = \left(\frac{AEu_r^2}{2r^2}\right) 2\pi r \tag{10}$$

$$PE = \left(\frac{\rho A}{2}\right) \left((u_r')^2\right) 2\pi r. \tag{11}$$

Here, A represents the cross-sectional area of the ring, E is the elasticity modulus, u is the radial displacement, r is the centerline radius of the ring in (10) and (11). The unit elongation of the ring can be represented by u/r.

By solving (9), as a result, the equation of motion can be written as follows:

$$(u'') + (\omega_\circ^2)u = 0, \tag{12}$$

where $\omega_\circ$ represents the angular frequency. By solving (12), we can obtain the ring fundamental frequency as below:

$$f_{ring} = \left(\frac{1}{2\pi r}\right)\sqrt{\frac{E}{\rho}}. \tag{13}$$

As vibration is not random, it has appropriate modes. In the case of a mechanical seal, the in-plane bending mode and out-of-plane bending mode of vibration can be expressed as follows:

$$f_{inplane} = \left(\frac{2n\left(n^2-1\right)}{\pi}\right)\left(\frac{h}{d^2}\right) * \sqrt{\frac{E}{\rho\left(12\,n^2\right) + \left(\frac{2t\,h^3(1+\nu)}{c}\right)}} \tag{14}$$

Eq. (14) represents the in-plane bending mode of vibration for the mechanical seal and is presented below:

$$f_{outplane} = \left(\frac{n\left(n^2-1\right)}{\pi\, d^2\sqrt{n^2+1}}\right)\sqrt{\frac{E\,t^2}{3\rho}}. \tag{15}$$

Eq.(15) shows the out-of-plane vibration of mechanical seal, where h and t are the cross-sectional height and thickness of the ring, d is the diameter, c is the torsion constant, v is Poisson's ratio, and n represents the mode of vibration in (15) and (16).

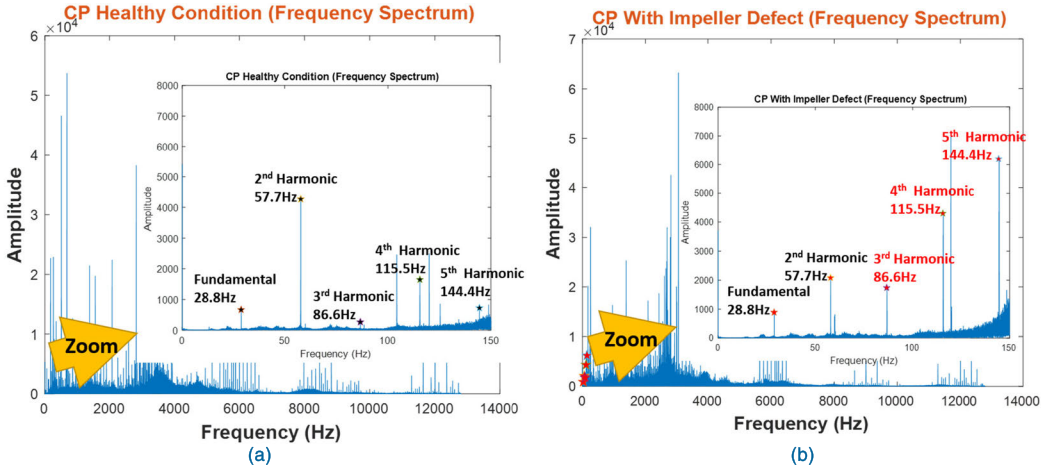

FIGURE 5. CP frequency spectrum (a) frequency spectrum under the normal condition (b) frequency spectrum under an impeller defect.

It is worth noting that the fundamental and in-plane vibrations occur at high frequencies. However, in the case of mechanical seal, lower frequencies can be found at out-of-plane bending modes of vibrations. Mechanical seal out-of-plane bending, or flexural vibration, and its consecutive modes are calculated using (15). It can be observed from Figures 6(a), 6(b), and 6(c) that the CP excitation frequency appears in between the 2nd and 3rd mode of flexural vibration with an amplitude almost double that of the normal condition whenever a mechanical seal defect occurs.

As mentioned earlier, to obtain discriminant features from vibration signals and to overcome macrostructural vibration, rather than focusing on specific frequencies, this study selects the CP vibration signal up to the 3rd mode of flexural vibration by using a lowpass filter with a cutoff frequency of 4.6 kHz. This vibration signal covers the FCFs of the impeller defect, corresponding hydraulic defects, and excitation frequencies of the CP, as can be seen in Figures 5 and 6.

B. STEP 2: RAW HYBRID FEATURE POOL CREATION

After filtering the vibration signal up to the 3rd mode of flexural vibration, raw statistical features in time, frequency, and time-frequency domains are extracted from the filtered vibration signal. The time-domain features used in this study are adapted from [35]. The frequency-domain features are employed from [36]. To extract features in the time-frequency domain, in this study, a wavelet packet transform (WPT) with a Daubechies (db4) mother wavelet is applied to the filtered vibration signal up to level 3. Features, like the mean, standard deviation, variance, root mean square, and kurtosis are extracted from the highest energy base of the WPT at level 3. A total of 22 features were extracted for each condition of the CP. All these features were combined into a single feature vector that formed a raw hybrid statistical feature pool.

C. STEP 3: PEARSON LINEAR DISCRIMENANT ANALYSIS

The raw hybrid statistical features result in a high-dimensional feature space. Moreover, some of the features might be irrelevant and can adversely affect the fault classification accuracy. To obtain discriminant features, in this step, a novel feature helpfulness-based dimensionality reduction technique called PLDA is introduced. The steps involved in PLDA are as follows:

1. A correlation is calculated between each feature (F_i) with other interclass features using the formulation presented below:

$$S_{F,G} = \frac{n\left(\sum_{i=1}^{k} F_i G_i\right) - \left(\sum_{i=1}^{k} F_i\right)\left(\sum_{i=1}^{k} G_i\right)}{\sqrt{\left[n\left(\sum_{i=1}^{k} F_i^2\right) - \left(\sum_{i=1}^{k} F_i\right)^2\right]\left[n\left(\sum_{i=1}^{k} G_i^2\right) - \left(\sum_{i=1}^{k} G_i\right)^2\right]}} \tag{16}$$

2. The correlation values F_i are summed up, which results in a helpfulness value ($h.value$) for F_i.

$$h.value = \sum_{j=1}^{l} s_{F,G}. \tag{17}$$

3. For a feature: if $h.value$ is greater than 0 then it will be selected and will be added to the helpful feature pool. Otherwise, the feature will be added to a less helpful feature pool as can be seen from eq. 18.

$$HF_i = \begin{cases} \forall F_i \text{ if } h.value > 0, & \text{helpfull} \\ otherwise, & \text{less helpfull} \end{cases} \tag{18}$$

4. For each class, the helpful feature pool is obtained and passed to the multiclass LDA (explained in section 2.1). The helpful feature pool helps multiclass LDA in improving the interclass scatteredness. Furthermore, it assists the multiclass LDA to overcome the issue of penalty graph for intraclass separation as the helpful feature pool only contains highly correlated features for each specific class.

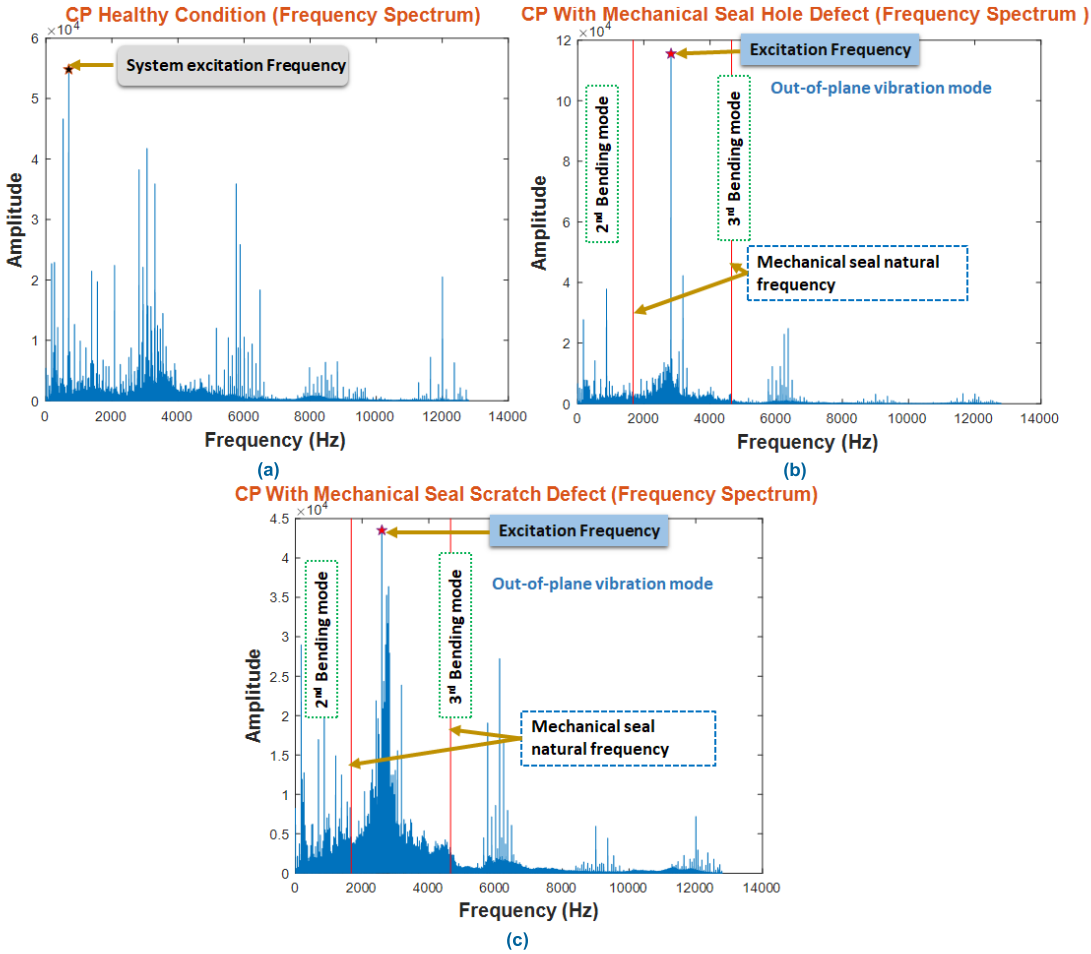

FIGURE 6. CP frequency spectrum (a) under the normal condition (b) under a mechanical seal hole defect (c) under mechanical seal scratch.

After applying PLDA to the raw hybrid feature pool, a reduced dimension discriminant feature space with reduced interclass scatteredness and high intraclass separability is obtained for CP fault diagnosis. To classify different CP conditions, this reduced-dimensionality feature space is provided as an input to the k-nearest neighbors (K-NN) classifier where K is equal to 3. The K-NN classifier has gained attention for solving classification problems due to its simple architecture and low computational complexity. However, the computational complexity of K-NN is strongly related to the dimensionality of feature space. Therefore, PLDA is used to reduce the dimensionality of the feature space while effectively increasing the separability of the features. The results obtained after implementing the proposed strategy and its comparison with other state-of-the-art methods are presented in the next section.

V. RESULTS AND DISCUSSION

The effectiveness of the proposed CP fault diagnosis strategy can be determined by an appropriate dataset configuration. In this study, a total of 1200 vibration samples were collected from CPs under normal, mechanical seal hole, mechanical seal scratch, and impeller defect conditions. Each signal class consists of 300 samples. The feature pool with statistical features extracted from the dataset comprises $C_{cp} \times S_{cp} \times f_{cp}$ features in total. Here, C_{cp} is the condition of the CP, S_{cp} is the vibration signal instances for each condition, and f_{cp} is the number of features extracted.

A k-fold cross-validation (k-CV) strategy of $k = 3$ was utilized for validating the proposed strategy during each trial. In k-CV, the dataset is divided into k folds, where each *k-th* fold is used as a testing subset, while the remaining *k-1* folds are used as a training subset. In this paper, for each class, out of 300 samples, 200 samples were randomly selected for training and the remaining 100 samples were used for testing. Thus, in total each testing set contained 400 samples while each training set consisted of the remaining 800 samples for the classification task.

A. PERFORMANCE EVALUATION OF THE PROPOSED CP FAULT DIAGNOSIS METHOD

To evaluate the quality of the proposed CP fault diagnosis strategy, this study compares the proposed method with one vibration signal features preprocessing technique [15], one supervised dimensionality reduction technique (Tr-LDA) [38], and one unsupervised dimensionality

TABLE 2. K-NN (K = 3) based classification results obtained from the proposed method and the reference methods.

Methods	TPRm (%)				ACA
	Healthy	Mechanical seal hole	Mechanical seal scratch	Impeller defect	
Proposed	100	100	100	100	100%
Viet et al [15]	96.90	90.09	97.82	100	96.20
Tr-LDA [34]	87.73	94.23	76.28	100	89.56
PCA [35]	80.35	97.70	64.04	85.71	81.95

reduction technique (PCA) [39]. To make the comparison fair and to ensure repeatability in results, the experiments were performed 15 times with a random combination of training and testing data. The true positive rate (TPR) or macro recall (Rm) for each class is calculated using the formulation as below:

$$TPR_m = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{(R_{TP}^{j,m})}{R_{TP}^{j,m} + R_{FN}^{j,m}}\right) \times 100(\%) \tag{19}$$

The average classification accuracy (ACA) for each method is computed by:

$$ACA = \frac{1}{k}\sum_{j=1}^{k}\left(\sum_{m=1}^{L}\frac{R_{TP}^{j,m}}{R_{sanples}}\right) \times 100(\%) \tag{20}$$

The macro precision (Pm) is calculated as follows:

$$P_m = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{(R_{TP}^{j,m})}{R_{TP}^{j,m} + R_{FP}^{j,m}}\right) \times 100(\%) \tag{21}$$

Finally, the error rate (Er) for each method is obtained using the expression below:

$$E_r = \frac{1}{k}\sum_{j=1}^{k}\left(\frac{(R_{TP}^{j,m}) + (R_{FN}^{j,m})}{(R_{TP}^{j,m}) + (R_{FN}^{j,m}) + R_{TN}^{j,m} + R_{FP}^{j,m}}\right) \times 100(\%), \tag{22}$$

where k represents the k folds of the CV, $(R_{TP}^{j,m})$ represents the true positives, $(R_{TN}^{j,m})$ are the true negatives, $(R_{FP}^{j,m})$ are the false positives, $(R_{FN}^{j,m})$ are the false negatives segregated by the classification algorithm as class m, j is the iteration of the k folds, and $R_{samples}$ is the total number of samples in a specific testing subset as expressed in (19), (20), (21), and (22), respectively.

The experimental results are presented in Table 2. The results demonstrate that the proposed CP fault diagnosis strategy outperforms the reference methods with an ACA of 100% with *Er* of 0%, and *Pm* of 100% as can be seen from Table 2 and Figure 7.

The obtained results can be explained as follows. MFs in the CP with varying severity cause random fluctuations in the vibration signal with specific frequencies. Due to complex fluid and mechanical component interactions in the CP, the fault frequencies or signatures are often overwhelmed

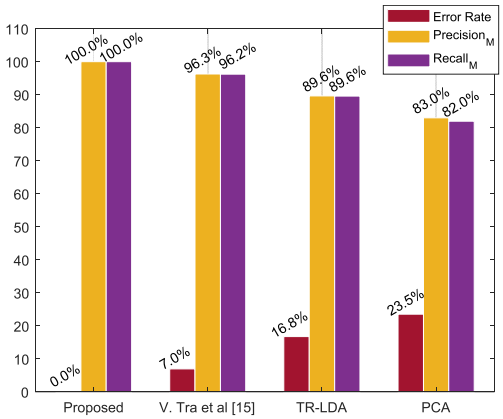

FIGURE 7. Proposed method evaluation against the reference methods.

by macrostructure vibrations. To extract discriminant intrinsic information from the vibration signal it is important to preprocess the raw vibration signal and the raw statistical features. Discriminant features help the classification algorithm to classify the faults efficiently. First, the proposed CP fault diagnosis strategy overcomes the issue of macrostructural vibration by selecting the fault characteristic modes of vibration. After filtering out the fault characteristic modes of vibration, the proposed strategy extracts raw hybrid statistical features in each domain (time, frequency, and time-frequency). A total of 22 features are extracted for each CP condition. To obtain intrinsic discriminant information from a raw hybrid statistical feature pool, a novel PLDA algorithm is applied. The PLDA first selects interclass highly correlated features from the raw hybrid feature pool and provides a highly discriminant feature space with reduced dimensionality after applying the transformation matrices of LDA. It can be seen from Figure 8(a) that the features from the proposed method after vibration signal and feature preprocessing are highly discriminant and separable. Furthermore, Figure 8(a) provides clear evidence for the effectiveness of the proposed method, along with its high classification accuracy.

The method of Viet *et al.* [15] first preprocesses the raw signal by removing the noise using the blind source separation (BSS) technique. Furthermore, raw hybrid features are extracted from the denoised signal in time, frequency, and the time-frequency domain. To increase the classification performance of K-NN, that method utilizes a genetic algorithm for discriminant feature selection. After applying the

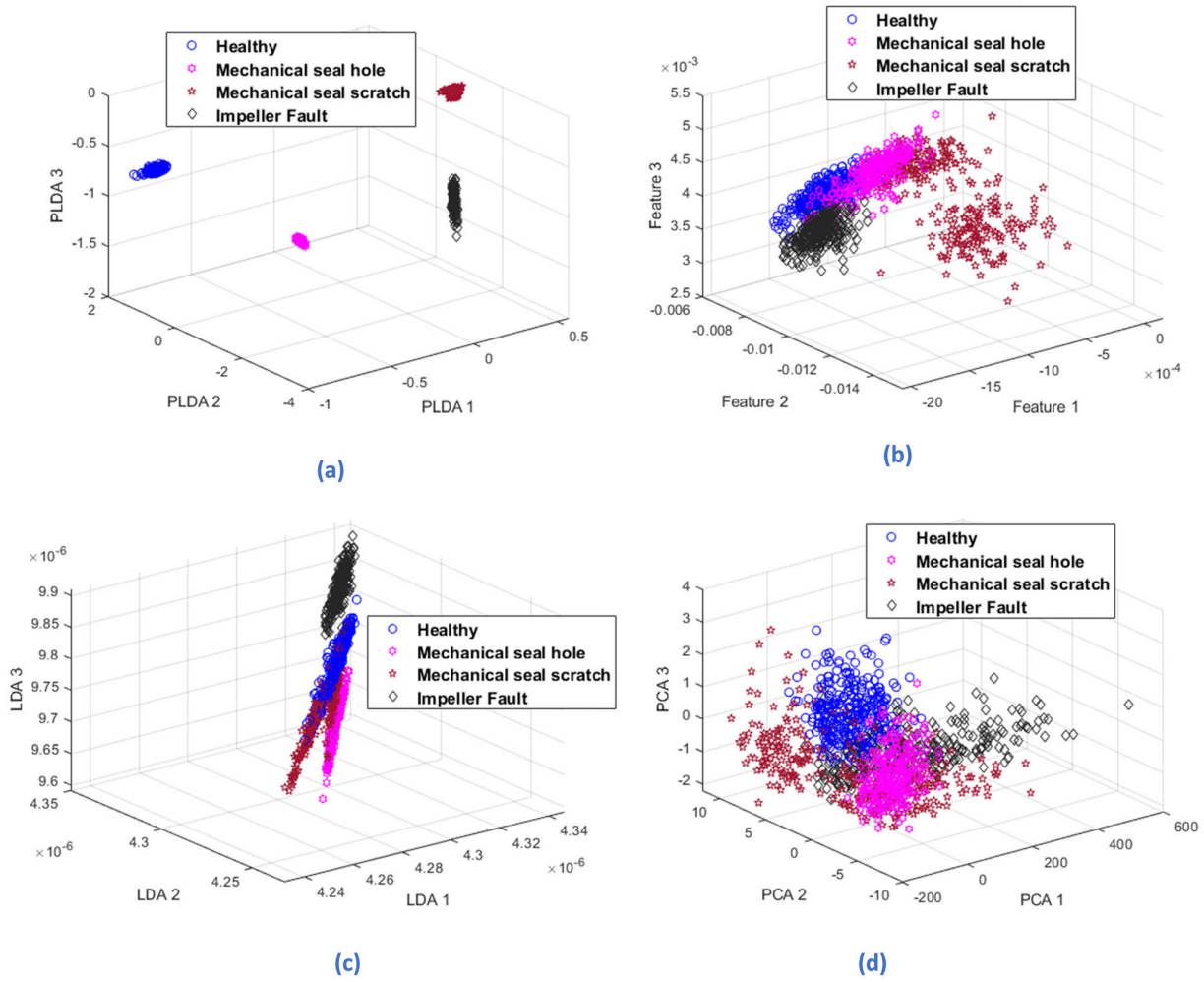

FIGURE 8. Three-dimensional representation of feature space (a) Obtained from the proposed method (b) Viet *et al.* [15] (c) Tr-LDA [38] (d) PCA [39].

presented method to our dataset for CP fault classification, we obtained an ACA of 96.20% with an *Er* of 7%, and *Pm* of 96.3%, as can be seen from Table 2 and Figure 7, which is slightly worse than the ACA of the proposed method. These results are expected because the BSS technique results in unimpaired signals, which often lose important information about the fault, as the algorithm obtains the unimpaired signal from a mixing matrix. Furthermore, genetic algorithm-based feature selection is very sensitive to the fitness function. Despite this drawback, it can be concluded that this method is effective, as it can be observed from Table 2 that the ACA is higher than 95%. Furthermore, the TPR for impeller fault and mechanical seal scratch condition classification is close to that of the proposed method. However, this method cannot effectively classify the weak incipient fault, such as differentiating between the mechanical seal hole and normal condition. The feature space obtained from the presented method is not fully separated which leads this technique to have an *Er* of 7%, as can be seen from Figure 8(b).

Another tested method is Tr-LDA, which is a linear dimensionality reduction technique that reduces the within-class scatteredness and increases the between-class separation using a trace ratio criterion. After implementing the algorithm provided in [38] to the dataset used in this paper, we obtained an ACA of 89.56% with *Er* of 16.75%, and *Pm* of 89.62%, as can be seen from Table 2 and Figure 7, an ACA that underperforms the proposed approach. These results are expected because Tr-LDA does not consider feature preprocessing for intrinsic discriminant information extraction from raw statistical features before the transformation. As can be observed from Figure 8(c), Tr-LDA reduces the inter-class scatteredness but cannot separate the classes efficiently which results in an *Er* of 16.75%.

PCA is a linear dimensionality reduction technique that constructs a lower-dimensional representation of the data based on the maximal variance in the data. By implementing the steps provided in [39] for PCA and by using the K-NN classifier, we obtained an ACA of 81.95% with an *Er* of 23.5%, and *Pm* of 83.0%, as can be seen from Table 2 and Figure 7 for our dataset. This is because PCA does not consider intraclass separability. Furthermore, the determination of the optimal number of components is also a

challenge for retaining fault feature information in PCA. As can be observed from Figure 8(d), the features are not well separated compared to the proposed method, which is why the classification error rate is high for PCA.

Overall, the proposed method for CP fault diagnosis is very effective for classifying CP faults of closely varying severities. The effectiveness of the proposed method arises from its main idea: preprocessing the vibration signal and the extracted raw hybrid statistical features. The proposed technique is easy to implement and is characterized by low computational cost. The results obtained and the properties of the proposed framework make it attractive for industrial use.

VI. CONCLUSION

In this paper, the classification of critical mechanical faults of the centrifugal pumps is attempted. A centrifugal pump fault classification strategy has been developed based on the vibration signal analysis and feature preprocessing. The developed strategy consists of three stages. In the first phase, the vibration signal is preprocessed by selecting the fault characteristic modes of vibration. In the second phase, raw statistical hybrid features are extracted from the filtered vibration signal. In the final phase, raw statistical features are preprocessed using a novel technique called Pearson linear discriminant analysis. The novel Pearson linear discriminant analysis method selects the highly correlated features from the hybrid feature pool and then reduces the selected feature pool dimension using the transformation matrices of linear discriminant analysis. In the experimental part of this study, the reduced discriminant feature space is provided as an input to the k-nearest neighbor algorithm to accomplish the task of fault classification. The experimental results demonstrate that the proposed approach segregates the mechanical faults of closely varying severity with a tested accuracy of 100% and thus outperformed the existing methods used for the comparison. In future research, the proposed method will be applied for diagnosing not only mechanical faults but for diagnostic both mechanical and hydraulic faults of centrifugal pumps.

REFERENCES

[1] H. Vogelesang, "An introduction to energy consumption in pumps," *World Pumps*, vol. 2008, no. 496, pp. 28–31, 2008, doi: 10.1016/S0262-1762(07)70434-0.

[2] V. K. Arun Shankar, S. Umashankar, S. Paramasivam, and N. Hanigovszki, "A comprehensive review on energy efficiency enhancement initiatives in centrifugal pumping system," *Appl. Energy*, vol. 181, pp. 495–513, Nov. 2016, doi: 10.1016/j.apenergy.2016.08.070.

[3] J. S. Rapur and R. Tiwari, "Experimental fault diagnosis for known and unseen operating conditions of centrifugal pumps using MSVM and WPT based analyses," *Measurement*, vol. 147, Dec. 2019, Art. no. 106809, doi: 10.1016/j.measurement.2019.07.037.

[4] S. M. Chittora, "Monitoring of mechanical seals in process pumps," KTH Roy. Inst. Technol., Stockholm, Sweden, Tech. Rep. 2018:686, 2018.

[5] K. McKee, G. Forbes, M. I. Mazhar, R. Entwistle, and I. Howard, "A review of major centrifugal pump failure modes with application to the water supply and sewerage industries," in *Proc. Asset Manage. Conf. (ICOMS)*, Gold Coast, QLD, Australia, 2011, pp. 1–12.

[6] M. C. Comstock, J. E. Braun, and E. A. Groll, "The sensitivity of chiller performance to common faults," *HVAC R Res.*, vol. 7, no. 3, pp. 263–279, Jul. 2001, doi: 10.1080/10789669.2001.10391274.

[7] M. A. Abu-Zeid and S. M. Abdel-Rahman, "Bearing problems' effects on the dynamic performance of pumping stations," *Alexandria Eng. J.*, vol. 52, no. 3, pp. 241–248, Sep. 2013, doi: 10.1016/j.aej.2013.02.002.

[8] J. S. Rapur and R. Tiwari, "Experimental time-domain vibration-based fault diagnosis of centrifugal pumps using support vector machine," *ASCE-ASME J. Risk Uncertainty Eng. Syst. B, Mech. Eng.*, vol. 3, no. 4, pp. 1–7, Dec. 2017, doi: 10.1115/1.4035440.

[9] E. Ebrahimi and M. Javidan, "Vibration-based classification of centrifugal pumps using support vector machine and discrete wavelet transform," *J. Vibroeng.*, vol. 19, no. 4, pp. 2586–2597, Jun. 2017, doi: 10.21595/jve.2017.18120.

[10] H. Sun, S. Yuan, and Y. Luo, "Characterization of cavitation and seal damage during pump operation by vibration and motor current signal spectra," *Proc. Inst. Mech. Eng. A, J. Power Energy*, vol. 233, no. 1, pp. 132–147, Feb. 2019, doi: 10.1177/0957650918769761.

[11] P. W. Tse, Y. H. Peng, and R. Yam, "Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—Their effectiveness and flexibilities," *J. Vib. Acoust.*, vol. 123, no. 3, pp. 303–310, Jul. 2001, doi: 10.1115/1.1379745.

[12] J. Duan, T. Shi, H. Zhou, J. Xuan, and Y. Zhang, "Multiband envelope spectra extraction for fault diagnosis of rolling element bearings," *Sensors*, vol. 18, no. 5, pp. 1–21, 2018, doi: 10.3390/s18051466.

[13] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 108–124, Jan. 2007, doi: 10.1016/j.ymssp.2005.12.002.

[14] J. Antoni, "The spectral kurtosis: A useful tool for characterising non-stationary signals," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 282–307, Feb. 2006, doi: 10.1016/j.ymssp.2004.09.001.

[15] V. Tra and J. Kim, "Pressure vessel diagnosis by eliminating undesired signal sources and incorporating GA-based fault feature evaluation," *IEEE Access*, vol. 8, pp. 134653–134667, 2020, doi: 10.1109/access.2020.3010871.

[16] L.-Q. Zuo, H.-M. Sun, Q.-C. Mao, X.-Y. Liu, and R.-S. Jia, "Noise suppression method of microseismic signal based on complementary ensemble empirical mode decomposition and wavelet packet threshold," *IEEE Access*, vol. 7, pp. 176504–176513, 2019, doi: 10.1109/ACCESS.2019.2957877.

[17] T. Jin, Q. Li, and M. A. Mohamed, "A novel adaptive EEMD method for switchgear partial discharge signal denoising," *IEEE Access*, vol. 7, pp. 58139–58147, 2019, doi: 10.1109/ACCESS.2019.2914064.

[18] Q. Hu, X.-S. Si, A.-S. Qin, Y.-R. Lv, and Q.-H. Zhang, "Machinery fault diagnosis scheme using redefined dimensionless indicators and mRMR feature selection," *IEEE Access*, vol. 8, pp. 40313–40326, 2020, doi: 10.1109/ACCESS.2020.2976832.

[19] A. Rai and J.-M. Kim, "A novel health indicator based on information theory features for assessing rotating machinery performance degradation," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6982–6994, Sep. 2020, doi: 10.1109/tim.2020.2978966.

[20] M. Sohaib and J. Kim, "Fault diagnosis of rotary machine bearings under inconsistent working conditions," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3334–3347, 2020.

[21] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587, doi: 10.1016/j.ymssp.2019.106587.

[22] X. Ma, Y. Si, Z. Yuan, Y. Qin, and Y. Wang, "Multistep dynamic slow feature analysis for industrial process monitoring," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9535–9548, Dec. 2020, doi: 10.1109/tim.2020.3004681.

[23] Z. Lou and Y. Wang, "New nonlinear approach for process monitoring: Neural component analysis," *Ind. Eng. Chem. Res.*, Sep. 2020, doi: 10.1021/acs.iecr.0c02256.

[24] M. Cui, Y. Wang, X. Lin, and M. Zhong, "Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine," *IEEE Sensors J.*, early access, Oct. 14, 2020, doi: 10.1109/jsen.2020.3030910.

[25] Y. Si, Y. Wang, and D. Zhou, "Key-performance-indicator-related process monitoring based on improved kernel partial least squares," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2626–2636, Mar. 2021, doi: 10.1109/tie.2020.2972472.

[26] S. Wang, J. Xiang, Y. Zhong, and H. Tang, "A data indicator-based deep belief networks to detect multiple faults in axial piston pumps," *Mech. Syst. Signal Process.*, vol. 112, pp. 154–170, Nov. 2018, doi: 10.1016/j.ymssp.2018.04.038.

[27] J. S. Rapur and R. Tiwari, "Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 40, no. 6, pp. 1–21, Jun. 2018, doi: 10.1007/s40430-018-1202-9.
[28] S. Li, N. Chu, P. Yan, D. Wu, and J. Antoni, "Cyclostationary approach to detect flow-induced effects on vibration signals from centrifugal pumps," *Mech. Syst. Signal Process.*, vol. 114, pp. 275–289, Jan. 2019, doi: 10.1016/j.ymssp.2018.05.027.
[29] M. A. S. ALTobi, G. Bevan, P. Wallace, D. Harrison, and K. P. Ramachandran, "Fault diagnosis of a centrifugal pump using MLP-GABP and SVM with CWT," *Eng. Sci. Technol., Int. J.*, vol. 22, no. 3, pp. 854–861, Jun. 2019, doi: 10.1016/j.jestch.2019.01.005.
[30] X. Tang, Z. Zhang, Q. Huang, and Y. Gong, "Fault location and fault type recognition of power system based on wavelet transform," in *Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT Asia)*, May 2019, pp. 689–692, doi: 10.1109/ISGT-Asia.2019.8881101.
[31] Y. Jin, C. Shan, Y. Wu, Y. Xia, Y. Zhang, and L. Zeng, "Fault diagnosis of hydraulic seal wear and internal leakage using wavelets and wavelet neural network," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 4, pp. 1026–1034, Apr. 2019, doi: 10.1109/TIM.2018.2863418.
[32] B. Bessam, A. Menacer, M. Boumehraz, and H. Cherif, "Wavelet transform and neural network techniques for inter-turn short circuit diagnosis and location in induction motor," *Int. J. Syst. Assurance Eng. Manage.*, vol. 8, no. S1, pp. 478–488, Jan. 2017, doi: 10.1007/s13198-015-0400-4.
[33] S. Alabied, U. Haba, A. Daraz, F. Gu, and A. D. Ball, "Empirical mode decomposition of motor current signatures for centrifugal pump diagnostics," in *Proc. 24th Int. Conf. Autom. Comput. (ICAC)*, Sep. 2018, pp. 1–6, doi: 10.23919/IConAC.2018.8749109.
[34] A. E. Prosvirin, M. M. M. Islam, and J.-M. Kim, "An improved algorithm for selecting IMF components in ensemble empirical mode decomposition for domain of rub-impact fault diagnosis," *IEEE Access*, vol. 7, pp. 121728–121741, 2019, doi: 10.1109/access.2019.2938367.
[35] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Oxford, U.K.: Elsevier, 2016.
[36] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, p. 2876, Dec. 2017, doi: 10.3390/s17122876.
[37] H. Roopa and T. Asha, "A linear model based on principal component analysis for disease prediction," *IEEE Access*, vol. 7, pp. 105314–105318, 2019, doi: 10.1109/access.2019.2931956.
[38] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2441–2451, May 2014, doi: 10.1109/TIE.2013.2273471.
[39] N. R. Sakthivel, B. B. Nair, M. Elangovan, V. Sugumaran, and S. Saravanmurugan, "Comparison of dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibration signals," *Eng. Sci. Technol., Int. J.*, vol. 17, no. 1, pp. 30–38, Mar. 2014, doi: 10.1016/j.jestch.2014.02.005.
[40] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8, doi: 10.1109/CVPR.2007.382983.
[41] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. 20th Int. Joint Conf. Artif. Intell.*, Jan. 2007, pp. 708–713.
[42] L. Yang, X. Liu, F. Nie, and Y. Liu, "Robust and efficient linear discriminant analysis with $L_{2,1}$-norm for feature selection," *IEEE Access*, vol. 8, pp. 44100–44110, 2020, doi: 10.1109/ACCESS.2020.2978287.
[43] D. Li, G. Hu, and C. J. Spanos, "A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis," *Energy Buildings*, vol. 128, pp. 519–529, Sep. 2016, doi: 10.1016/j.enbuild.2016.07.014.
[44] S. Yedidiah, *Centrifugal Pump User's Guidebook: Problems and Solutions*. New York, NY, USA: Springer, 2012.
[45] *Condition Monitoring and Diagnostics of Machines—Vibration Condition Monitoring—Part 1: General Procedures*, Standard ISO 13373-1:2002, 2002. Accessed: Sep. 11, 2020. [Online]. Available: https://www.iso.org/standard/21831.html
[46] D. H. Y. Weaver, Jr., William, and S. P. Timoshenko, *Vibration Problems in Engineering*, vol. 207, no. 2. Hoboken, NJ, USA: Wiley, 1929.
[47] Z. Ahmad, A. Rai, A. S. Maliuk, and J.-M. Kim, "Discriminant feature extraction for centrifugal pump fault diagnosis," *IEEE Access*, vol. 8, pp. 165512–165528, 2020, doi: 10.1109/access.2020.3022770.

ZAHOOR AHMAD received the B.S. degree in computer engineering from the COMSATS Institute of Information Technology (CIIT), currently COMSATS University Islamabad (CUI), Attock, Pakistan, in 2016, and the M.S. degree in electronics and information engineering from Korea Aerospace University (KAU), Goyang, South Korea. He is currently pursuing the Ph.D. degree in computer engineering with the University of Ulsan, South Korea, where he has been a Graduate Research Assistant with the Ulsan Industrial Artificial Intelligence (UIAI) Laboratory, since 2019.

He worked on the development of advanced algorithms for UAV path planning. His current research interests include artificial intelligence, signal processing, fault diagnosis, vibration-based condition monitoring of industrial machinery, and fault feature extraction.

Mr. Ahmad received the institution's highest prize of Gold Medal for his B.S. degree from CUI, Attock. He received the fully-funded scholarship for the M.S. degree from a Korean Government project. His recent awards and honors include the National Research Foundation (NRF) of Korea Research Fellowship Scholarship and the University of Ulsan President Excellence Scholarship for pursuing his Ph.D. Program.

ALEXANDER E. PROSVIRIN received the Engineer's degree in specialization of "control and informatics in technical systems" from the Moscow State University of Mechanical Engineering "MAMI" (MSUME-MAMI), currently Moscow Polytechnic University, Moscow, Russia, in 2013. He is currently pursuing the Ph.D. degree in computer engineering with the University of Ulsan, Ulsan, South Korea, where he has already defended his Ph.D. dissertation.

Since 2016, he has been working as a Graduate Research Assistant with the Ulsan Industrial Artificial Intelligence (UIAI) Laboratory, Department of Electrical, Electronics, and Computer Engineering, University of Ulsan. He is the author of more than ten scientific articles and four book chapters. His research interests include the data-driven fault diagnosis and anomaly detection of complex engineering systems using artificial intelligence, machine learning, advanced signal processing, fault feature extraction, and feature engineering.

Mr. Prosvirin has served as a reviewer for several recognized journals.

JAEYOUNG KIM (Member, IEEE) received the B.S. and M.S. degrees in electrical, electronics, and computer engineering from the University of Ulsan, South Korea, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree in electrical, electronics, and computer engineering.

His current research interests include artificial intelligence, signal processing, fault diagnosis, and the prognosis of industrial machinery.

JONG-MYON KIM (Member, IEEE) received the B.S. degree in electrical engineering from Myongji University, Yongin, South Korea, in 1995, the M.S. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2000, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2005.

He is currently a Professor with the Department of IT Convergence, University of Ulsan, Ulsan, South Korea. His research interests include multimedia-specific processor architecture, fault diagnosis and condition monitoring, parallel processing, and embedded systems.

Dr. Kim is a member of the IEEE Industrial Electronics Society.