Three-Stage Hybrid Fault Diagnosis for Rolling Bearings With Compressively Sampled Data and Subspace Learning Techniques

Hosameldin O. A. Ahmed and Asoke K. Nandi, *Fellow, IEEE*

***Abstract*—To avoid the burden of much storage requirements and processing time, this paper proposes a three-stage hybrid method, compressive sampling with correlated principal and discriminant components (CS-CPDC), for bearing faults diagnosis based on compressed measurements. In the first stage, CS is utilized to obtain compressively sampled signals from raw vibration data. In the second stage, an effective multistep feature learning algorithm obtains fewer features from correlated principal and discriminant attributes from the compressively sampled signals, which are then concatenated to increase the performance. In the third stage, with these concatenated features, multiclass support vector machine is used to train, validate, and classify bearing faults. Results show that the proposed method, CS-CPDC, offers high classification accuracies, reduced computation time, and storage requirement, with fewer measurements.**

***Index Terms*—Bearing fault classification, canonical correlation analysis (CCA), compressive sampling (CS), linear discriminant analysis (LDA), machine condition monitoring (MCM), principal component analysis (PCA).**

I. Introduction

ROTATING machines are widely used in industry for various tasks. Unforeseen machine failures may affect production schedules, product quality, and production costs. Condition monitoring of a rotating machine can play an important role in machine availability. Rolling element (RE) bearings are the critical components in a rotating machine and their failures may lead to more major failures in machines. With the development of sensing systems, various forms of signals can be collected for machine condition monitoring (MCM) [1]–[3]. The analysis of vibration signals can focus on three main groups: time domain, frequency domain, and time–frequency domain [4]–[8].

Techniques beyond bandlimited sampling [9] offer lower sampling rates and reduced amount of data. To address the challenges of analyzing a large amount of acquired vibration data, various techniques have been used to reduce signals dimensionality. For example, principal component analysis (PCA) [10], linear discriminant analysis (LDA) [11], independent component analysis [12], and genetic algorithm (GA) [13] are among the most commonly used methods. For instance, Malhi and Gao [14] presented a PCA-based approach to select the most representative features for bearing faults classification that has shown to improve classification accuracy of feedforward neural network (NN) and radial basis function (RBF) network. Jin *et al.* [15] introduced the trace ratio LDA method to reduce dimension of high-dimension non-Gaussian data for fault classification. Ciabattoni *et al.* [16] introduced a novel LDA-based algorithm to deal with fault data dimension reduction and fault detection issues. Jack and Nandi [13] examined the use of a GA to select the most significant input features from a large set of possible features in MCM contexts.

In recent years, several publications have proposed new methods for bearing fault diagnosis. For example, Amar *et al.* [17] suggested a novel bearing fault classification approach combining vibration spectrum imaging and artificial NN. Li *et al.* [18] presented a semisupervised diagnosis method based on a distance-preserving self-organizing map for classifying different bearing faults. Soualhi *et al.* [19] examined the combination of Hilbert–Huang transform, support vector machine (SVM), and support vector regression, and showed its efficiency for the monitoring of ball bearing. Chen and Li [20] proposed a multisensory feature fusion method using Sparse AutoEncoder (SAE) and deep belief network that outperform some other feature fusion methods. Zhang *et al.* [21] presented a hybrid intelligent fault diagnosis method integrating permutation entropy, ensemble empirical mode decomposition, and optimized SVM. Recently, Lei *et al.* [22] proposed a two-stage learning method based on sparse filtering to learn features from mechanical vibration signals for machine fault diagnosis.

Compressive sampling (CS) [23] has been developed for sensing and compression. The efficiency of CS in machine fault diagnosis has been validated in several studies. For instance, in an analysis of the effects of CS on the classification of bearing

Manuscript received November 29, 2017; revised March 22, 2018, June 4, 2018, and July 23, 2018; accepted August 20, 2018. Date of publication September 13, 2018; date of current version February 28, 2019. This work was supported in part by the National Science Foundation of China under Grant 61520106006, and in part by the National Science Foundation of Shanghai under Grant 16JC1401300. *(Corresponding author: Asoke K. Nandi.)*

H. O. A. Ahmed is with the Department of Electronic and Computer Engineering, Brunel University London, London UB8 3PH, U.K. (e-mail: hosameldin.ahmed@brunel.ac.uk).

A. K. Nandi is with the Department of Electronic and Computer Engineering, Brunel University London, London UB8 3PH, U.K., and also with the Key Laboratory of Embedded Systems and Service Computing, College of Electronic and Information Engineering, Tongji University, Shanghai 200092, China (e-mail: asoke.nandi@brunel.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2018.2868259

faults, Wong *et al.* [24] found small performance degradation using entropy-based features computed from CS-based recovered signal. Zhang *et al.* [25] proposed a method based on compressed vibration signal using several overcomplete dictionaries trained by a dictionary learning method. Each of these dictionaries can be effective in sparse signal decomposition for a specific bearing condition. Several studies have shown that it is possible to learn directly from compressed measurements without reconstructing the original signal. In [26], an intelligent condition monitoring method for bearing faults based on CS and sparse overcomplete feature learning algorithm using SAE was proposed. In a recent paper by Ahmed *et al.* [27], three approaches to process compressed vibration measurements were proposed for classification of bearing faults, including using the compressed measurements directly as the input to the classifier, and extracting features from these compressed measurements using PCA and LDA.

Although the above investigations reported many interesting results, to the best of authors' knowledge, there is little work in the literature on learning from compressively sampled signals. Also, no research exists on the combination of CS, PCA, LDA, and canonical correlation analysis (CCA). Motivated by the idea of CS and the advantages of PCA, LDA, and CCA, we present a new method for intelligent fault diagnosis. PCA can be applied to a compressively sampled vibration signal to extract the components. However, by selecting only the larger components of PCA, we may remove some useful features. On the other hand, LDA works well with a larger number of samples. Intuitively, these two different sets of features reflect different characteristics of the original signals and may be complementary. Hence, we integrate them in a combined framework, by forming a relationship between them via CCA and concatenating their resulting linear combinations into a feature vector. The proposed method, CS with correlated principal and discriminant components (CS-CPDC), needs fewer measurements, reduces computation time, storage requirement and the bandwidth for transmitting compressively sampled data. The contributions of this paper are summarized as follows.

1) A new three-stage intelligent hybrid method, CS-CPDC, for fault diagnosis is presented.
   a) The aim is to reduce computations, to reduce transmission costs, and to reduce demands on the environment compared to other techniques. CS-CPDC reduces the measured vibration data using CS that linearly maps the original vibration data into a lower dimension space of compressively sampled data.
   b) The aim is to obtain a set of features that achieves superior classification. While an individual set of features (e.g., either PCA or LDA) can be good for representations, it may not be good for classifications. Thus, we propose to combine PCA and LDA features via CCA in a three-step process to transform the characteristic space of compressively sampled signal into a low-dimensional space of correlated important and discriminant attributes, which are then concatenated to form a vector of useful features.

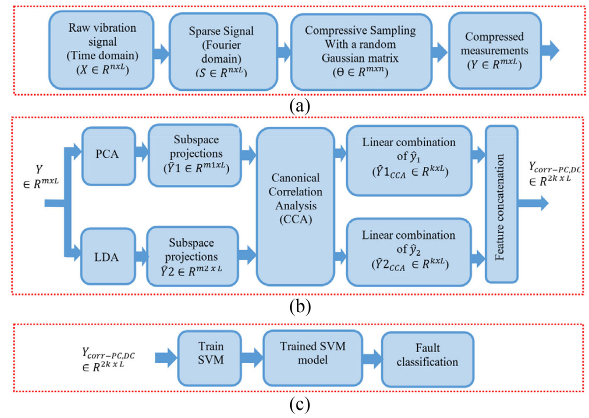


Fig. 1. Training of our proposed method. (a) First stage. (b) Second stage. (c) Third stage.

   c) For fault classification, multiclass SVM based on error-correcting output codes (ECOC) is employed to classify the bearing health condition using the learned feature vector.
2) Two case studies of bearing datasets (real and not simulated) are used to validate the efficiency of CS-CPDC. First, the CS model of CS-CPDC and the effect of the amount of compressively sampled data are thoroughly studied to verify that our CS model generates compressively sampled signals that possess the quality of the original signals. Second, various scenarios of the proposed three-step learning algorithm of CS-CPDC are studied to test the classification accuracy in each stage. In addition, CS-CPDC is compared with related work. The comparison results show that CS-CPDC is able to achieve high levels of classification accuracy with fewer measurements compared with existing methods.
3) We have studied the advantages of CS in our proposed method. We demonstrate that CS reduced computational time and yet provided high classification accuracy.

The rest of this paper is structured as follows. Section II is devoted to descriptions of the proposed method. Sections III and IV are dedicated to a description of the performed experiments and datasets of the two faults classification case studies of bearings datasets and the corresponding experimental results. Finally, Section V draws some conclusions from this study.

II. Proposed Method

This section describes the CS-CPDC method for machine fault detection and classification. In the first stage [see Fig. 1(a)], CS is used to obtain compressively sampled raw vibration signals. In the second stage [see Fig. 1(b)], a proposed multistep approach of PCA, LDA, and CCA is used to extract features from the obtained compressively sampled signals. In the third stage [see Fig. 1(c)], SVM is applied to classify bearing health condition using the learned features from the previous stage.

A. First Stage: CS

In this stage, CS-CPDC acquires compressively sampled signals using CS framework [23], [28]. CS is an extension of sparse

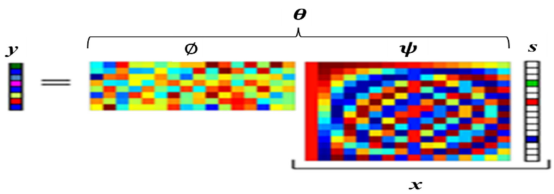


Fig. 2. CS framework.

Algorithm 1: CS Stage.

Input: $X \in R^{n\ x\ L}, \Theta \in R^{mxn}, \alpha$
Output: $Y \in R^{m\ x\ L}$
1: sparMethod(X) $\longrightarrow S \in R^{nxL}$
2: Project S into Θ with compressed sampling rate α to obtain compressively-sampled signal $Y \in R^{mxL}$

representations and special case of it. The simple idea of CS is that many real-world signals have sparse representations in some domain, e.g., Fourier transform (FT), can be recovered from fewer measurements under certain conditions. In fact, CS is based on two principles: 1) sparsity of the signal of interest; and 2) the measurements matrix that satisfies the data minimal information loss, i.e., fulfill restricted isometry property (RIP). Concisely, we describe the sparsity as follows:

Assume that $\mathrm{x} \in R^{nx1}$ be an original time indexed signal. Given a sparsifying transform matrix $\psi \,\epsilon\, R^{nxn}$ whose columns are the basis elements $\{\psi_i\}_{i=1}^{n}$. Based on this basis, x can be represented as follows:

$$x = \sum_{i=1}^{n} \psi_i s_i. \tag{1}$$

Or more efficiently

$$x = \psi s. \tag{2}$$

Here, s is $n^{*}1$ column vector of coefficients. If the basis ψ produces q-sparse representations of x, then (1) can be rewritten as follows:

$$x = \sum_{i=1}^{q} \psi_{ni} s_{ni} \tag{3}$$

where ni is the index of the basis elements and the coefficients corresponding to the q nonzero elements. So, $s \,\epsilon\, R^{nx1}$ is a vector column with only q nonzero elements and represents the sparse representation vector of x.

Based on CS framework, $m << n$ projections of the vector x with a group of measurement vectors $\{\emptyset_j\}_{j=1}^{m}$ and the sparse representations s of x can be produced from

$$y = \emptyset\psi s = \theta s. \tag{4}$$

Here, y is $m^{*}1$ column vector of the compressed measurements and $\theta = \emptyset\psi$ is the measurement matrix. To produce good compressed measurements, the measurement matrix θ has to satisfy the data minimal information loss, i.e., satisfy the RIP.

Definition 1.1: The measurement matrix θ satisfies the RIP if there exists a parameter $\delta \in (0, 1)$ such that

$$(1 - \delta)\, \|s\|_{l_2}^2 \le \theta\, \|s\|_{l_2}^2 \le (1 + \delta)\, \|s\|_{l_2}^2 . \tag{5}$$

The size of the measurement matrix $(m^{*}n)$ depends on the CS rate (α) (i.e., $m = \alpha^{*}n$). Fig. 2 shows an illustration of the CS framework. The model described above is meant to be single measurement vector CS that recovers one vector from its corresponding compressed measurement vector. But, multiple measurement vectors CS (MMV-CS) is considered for signals that are represented as a matrix with a set of jointly sparse vectors such that

$$Y = \Theta S \tag{6}$$

where $Y \in R^{mxL}$, m is number of compressed measurements and L is number of observations, $\Theta \in R^{mxn}$ is a dictionary, and $S \in R^{nxL}$ is a sparse representation matrix. Several studies have been conducted to reconstruct jointly sparse signals (S) given multiple compressed measurement vector [29], [30]. In our proposed method, MMV-CS has been used to obtain compressively sampled signals since the dataset consists of a matrix of multiple measurements. Also, since it is possible to recover the original signal (X) from the compressed data (Y), this indicates that (Y) possesses the quality of the original signal (X). So, in this study, we use the compressed measurements directly.

As shown in Fig. 1(a), first the sparse representations ($S \in R^{nxL}$) that consists of only a small number of $q \ll n$ of nonzero coefficients are obtained from raw vibration signals ($X \in R^{nxL}$) using an appropriate basis representation, e.g., FT, wavelet transform, etc. Then, the obtained (S) is projected into a suitable measurement matrix ($\Theta \in R^{mxn}$), e.g., random matrix with i.i.d. Gaussian entries and Bernoulli (± 1) matrix, with the compressed sampling rate (α) to generate the compressively sampled signals ($Y \in R^{mxL}$) where m is the number of compressed signal elements (i.e., $m = \alpha^{*}n$). These procedures are summarized next in Algorithm 1 [28].

B. Second Stage: Feature Learning

While the CS projections obtained in the first stage help to recover the original signal from low-dimensional features, they may not be the best from a discriminant point of view. Furthermore, the size of the CS projections may still represent a large amount of data collected in real operating condition. Consequently, techniques to extract fewer features of the CS projections are required. Accordingly, PCA and LDA are commonly used. However, while an individual set of features (e.g., either PCA or LDA) can be good for representations, it may not be good for classifications. Thus, the aim of the second stage is to generate features for superior classification accuracy.

The second stage consists of three steps as shown in Fig. 1(b). In the first step, CS-CPDC finds two feature representations from the compressively sampled signals using PCA and LDA. Hence, we transform the characteristic space of the compressively sampled signal into a low-dimensional space defined by those basis vectors corresponding to larger eigenvalue components (PCA). Furthermore, we augment these basis vectors with discriminant attributes learned through supervised learning

(LDA). Let us consider a set of compressively sampled signals $Y \in R^{mxL}$; here, Y can be presented as $Y = [y_1, y_2, \ldots, y_l]$ where $1 \leq l \leq L$, and let each of these signals fit in with one of the c classes of machine conditions. To extract feature representations of these signals, CS-CPDC performs a linear transformation to map the m-dimensional space of the compressively sampled vibration to a lower dimensional feature space, using the following equation:

$$\hat{y}_r = W^T\, y_r. \tag{7}$$

Here, $r = 1, 2, \ldots, L$, $\hat{y}_r$ is the transformed feature vector with reduced dimension, and W is a transformation matrix.

To find the larger attributes of the compressively sampled vibration signals using (7), we used PCA to compute W projection matrix using the scatter matrix, i.e., the covariance matrix C of the compressively sampled data, which can be computed as follows:

$$C = \frac{1}{L}\sum_{i=1}^{L}(y_i - \bar{y})(y_i - \bar{y})^T. \tag{8}$$

Here, $\bar{y}$ is the mean of all samples. In the produced projection matrix W, successive column vectors from left to right correspond to decreasing eigenvalues. We select the $m1$ eigenvectors corresponding to the $m1$ largest eigenvalues. Hence, a new $m1$-dimensional space $\hat{Y}1 \in R^{m1xL}$ is produced from $Y \in R^{mxL}$, where $m1 << m$.

Furthermore, we employed LDA to compute discriminant attributes from the compressively sampled signals. LDA considers maximizing the Fisher criterion function $J(W)$, i.e., the ratio of the between the class scatter (S_B) to the within class scatter (S_w) such that

$$J(W) = \frac{|W^T S_B W|}{|W^T S_w W|} \tag{9}$$

where

$$S_B = \frac{1}{L}\sum_{i=1}^{c} l_i \left(\mu^i - \mu\right)\left(\mu^i - \mu\right)^T \tag{10}$$

$$S_w = \frac{1}{L}\sum_{i=1}^{c}\sum_{j=1}^{l_i}(y_j^i - \mu^i)(y_j^i - \mu^i)^T \tag{11}$$

where μ^i is the mean vector of class i, $y \in R$ of size $L^* m$ is the training dataset, y_1^i represents the dataset belong to the cth class, n_i is the number of measurements of the ith class, μ^i is the mean vector of class i, and μ is the mean vector of all training dataset. LDA projects the space of the compressively sampled data onto a $(c - 1)$-dimension space by finding the optimal projection matrix W by maximizing $J(W)$. Now, W is composed of the selected eigenvectors $(\hat{w}_1, \ldots, \hat{w}_{m2})$ with the first $m2$ largest eigenvalues $(m2 = c - 1)$. Consequently, a new $m2$-dimensional space of discriminant attributes $\widehat{Y}2 \in R^{m2xL}$ is produced from $Y \in R^{mxL}$, where $m2 << m$.

These different feature representations extracted from the same dataset always reflect different characteristics of the original signals. The best combination of them retains the multiple features of the integration that can be used effectively for classification. We propose CCA [31] to combine PCA and LDA features to obtain superior classification.

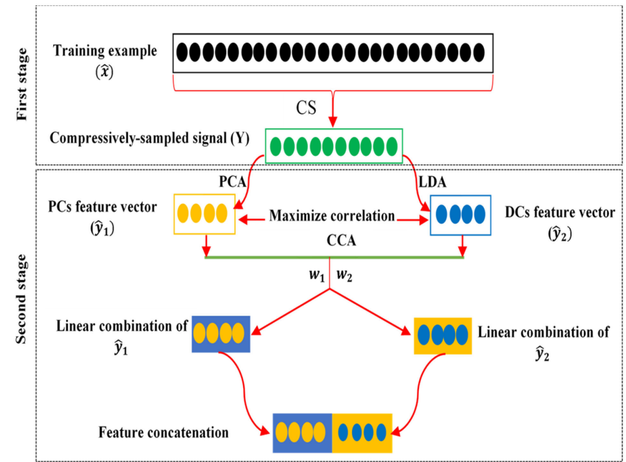


Fig. 3. Illustration of the training process of the first and second stages of the proposed method.

The second step of the multistep procedure of CS-CPDC utilizes CCA to combine the different feature representations $\widehat{Y}1$ and $\widehat{Y}2$ by forming the relationship between them, i.e., maximizing the overlapping variance between $\widehat{Y}1$ and $\widehat{Y}2$. The main idea is to find linear combinations of $\hat{Y}1$ and $\hat{Y}2$ that can maximize the correlation between them based on the following objective function:

$$(W_1, W_2) = \arg\max_{W_1, W_2}\ \acute{W}_1 C_{\hat{Y}1\hat{Y}2}$$

$$\text{s.t. } \acute{W}_1 C_{\hat{Y}1\hat{Y}1} W_1 = 1, \acute{W}_2 C_{\hat{Y}2\hat{Y}2} W_2 = 1 \tag{12}$$

where $C_{\hat{Y}1\hat{Y}2}$ is the cross-covariance matrix of $\hat{Y}1$ and $\hat{Y}2$ that can be computed using the following equation:

$$C\left(\hat{Y}1, \hat{Y}2\right) = \hat{E}\left[\begin{pmatrix}\hat{Y}1\\ \hat{Y}2\end{pmatrix}\begin{pmatrix}\hat{Y}1\\ \hat{Y}2\end{pmatrix}'\right] = \begin{bmatrix}C_{\hat{Y}1\hat{Y}1} & C_{\hat{Y}1\hat{Y}2}\\ C_{\hat{Y}2\hat{Y}1} & C_{\hat{Y}2\hat{Y}2}\end{bmatrix}. \tag{13}$$

The resulting linear combinations of $\hat{Y}1$ ($\hat{Y}1_{\text{CCA}} = W_1 * \hat{Y}1$) and the $\hat{Y}2$ ($\hat{Y}2_{\text{CCA}} = W_2 * \hat{Y}2$) will maximize their correlation.

Finally, in the third step, the learned features $\hat{Y}1_{\text{CCA}}$ and $\hat{Y}2_{\text{CCA}}$ are concatenated to obtain a vector ($Y_{\text{corr - PC, DC}} \in R^{Lx2k}$) that comprises highly correlated representations of principal and discriminative components where k is equal to minimal dimension size of $m1$ and $m2$. These procedures are summarized in Algorithm 2. Fig. 3 shows an illustration of the training process of the first and second stages of our proposed method.

C. Third Stage: Fault Detection and Classification

For our multiclass problem, we employed multiclass SVM classifier. The simple idea of SVM is that it can find the best hyperplane(s) to separate two classes. Based on the features of the data, SVM can make linear or nonlinear classifications by different kernel functions, e.g., RBF, polynomial function,

Algorithm 2: Feature Learning Stage.

Input: $Y \in R^{mxL}, y \in R^{1xL}$: label information vector for each data points, c: number of classes, $m1$: selected number of principal components
Output: $Y_{\text{corr - PC, DC}} \in R^{Lx2k}$
1: $\text{PCA}(Y) \longrightarrow E1 \in R^{mxm1}$
2: $\hat{Y}1 = Y^T * E1$
3: $\text{LDA}(Y, y) \longrightarrow E2 \in R^{mxm2}$; $m2 = c - 1$.
4: $\hat{Y}2 = Y^T * E2$
5: $\text{CCA}(\hat{Y}1, \hat{Y}2) \longrightarrow w_1, w_2 \in R^{Lxk}$; $k = \min(m1, m2)$.
6: $\hat{Y}1_{\text{CCA}} = w_1 * \hat{Y}1$, $\hat{Y}2_{CCA} = w_2 * \hat{Y}2$
7: $Y_{\text{corr - PC, LD}} = [\hat{Y}1_{\text{CCA}}\ \hat{Y}2_{\text{CCA}}]$

and Sigmoid function [32]. Multiclass SVM includes multiple two-class subproblems, i.e., SVM classifiers that can be easily combined together using one-versus-one and one-versus-all coding design.

In our case, we applied “fitcecoc” function [33] on the learned features from the second stage. It uses $c(c-1)/2$ binary SVM models using one-versus-one coding design, where c is the number of unique class labels. This will return a fully trained ECOC multiclass model that cross-validated using tenfold cross validation.

III. First Case Study

A. Data Description

The vibration data used in this case were recorded from experiments on a small test rig that mimics operating roller bearings' environment. Six conditions of roller bearings health conditions have been recorded. These contain, two normal conditions, that is, a brand new condition (NO) and a worn but undamaged condition (NW), as well as four fault conditions, including inner race (IR) fault, an outer race (OR) fault, RE fault, and cage (CA) fault.

The test rig used to collect the vibration data involves a dc motor driving the shaft through a flexible coupling, with the shaft supported by two Plummer bearing blocks. A series of damaged bearing was inserted in one of the Plummer blocks, and the resultant vibrations in the horizontal and vertical planes were measured using two accelerometers. The output from the accelerometers was fed back through a charge amplifier to a Loughborough Sound Images DSP32 ADC card (using a low-pass filter with a cutoff 18 kHz), and sampled at 48 kHz. The machine was run at a series of 16 different speeds ranging between 25 and 75 r/s, and ten-time series were taken at each speed. This gave a total of 160 examples of each condition, and a total of 960 raw data files to work with.

To apply CS-CPDC in this case study, we started by acquiring the compressively sampled vibration signal from the high-dimensional data X with 6000 time samples for each of the 960 observations. First, we used FFT basis as sparse representations of X. Then, we applied CS framework with different sampling rates (α) using a random Gaussian matrix that satisfies the RIP. The size of the Gaussian matrix is m by n, where n is the length of the original vibration signal measurements and m is the number of compressed signal elements (i.e., $m = \alpha^* n$). Based on CS framework, multiplying this matrix with our signal sparse representations generates different sets of compressed measurements of the vibration signal.

B. Experimental Results

First, 50% of the total observations were randomly selected for training and the other 50% for testing. Then, we examined the selection of the compressed sampling rate (α) using different values (0.01, 0.02, 0.03, 0.04, 0.5, 0.1, and 0.2) to generate compressively sampled vibration signals.

To ensure that our CS model generates enough samples for the purpose of bearing fault classification, we used the generated compressively sampled signals in the first stage to reconstruct the original signal X by applying the compressive sampling matching pursuit (CoSaMP) algorithm [34]. For example, with $\alpha = 0.1$, the average percentage reconstruction errors for the six conditions of bearings are 1.8% (NO), 0.9% (NW), 3.3% (IR), 1.6% (OR), 0.7% (RE), and 2.6% (CA), which indicate good signal reconstruction.

To learn features from the training set with compressed measurements, we proposed a multistep approach, i.e., the second stage of our proposed method described in Fig. 1(b) using $(c-1)$ components for LDA and 40 principal components for PCA for each of the α values described above. These learned features from the second stage were used to train the multiclass SVM. To achieve better evaluations of the trained ECOC multiclass model, we applied tenfold cross validation in all our experiments. The training dataset is randomly subdivided into ten subsets. Each subset is validated on the classifier that is trained using the other nine subsets. The process is repeated 20 times and the training classification accuracy is the average taken from these 20 trials.

To evaluate the performance of the proposed method, we first compressively sampled each testing signal using the same values of α used to sample training set, and then the trained multistep algorithm is used to obtain the learned features of the testing set. Once the features were learned, the trained SVM is used to classify the testing signals. The overall results are shown in Table I, where the classification accuracy is the average of 20 trials for each experiment, and the time is obtained by averaging the testing time of these 20 trials.

Table I shows that the value of α affects not only the classification accuracy results but also the time required by the CS-CPDC method to complete the classification task. It can be clearly seen that the larger the value of α, the better the classification accuracy and the more time the method requires. However, high levels of classification accuracy achieved with less than 25% of the original data samples. In particular, accuracies from our proposed method are 99.9%, 99.8%, and 99.3% for only 20%, 10%, and 5% of the whole data, respectively.

C. Effect of Numbers of Principal Components on Classification Accuracy

To determine the effect of the number of PCs on classification accuracy of CS-CPDC, we tested it with

TABLE I
CLASSIFICATION RESULTS WITH THEIR CORRESPONDING RMSE AND COMPUTATIONAL TIME USING VARIOUS COMPRESSED SAMPLING RATES

Compressed Sampling rate (α)	The cross-validated accuracy (%)	Testing Classification Accuracy (%)	Testing Time (s)
α = 0.2 (1200 samples)	100.0 ± 0.0	99.9 ± 0.1	7.8± 0.01
α = 0.1 (600 samples)	100.0 ± 0.0	99.8 ± 0.2	6.7± 0.07
α = 0.05 (300 samples)	99.8 ± 0.2	99.3 ± 0.6	4.9 ± 0.03
α = 0.04 (240 samples)	99.4 ± 0.6	98.8 ± 1.2	4.3 ± 0.11
α = 0.03 (180 samples)	99.1 ± 0.8	98.4 ± 1.4	3.8 ± 0.03
α = 0.02 (120 samples)	99.1 ± 0.9	97.8 ± 1.3	3.2 ± 0.09
α = 0.01 (60 samples)	98.3 ± 1.6	96.4 ± 0.6	3.1 ± 0.16

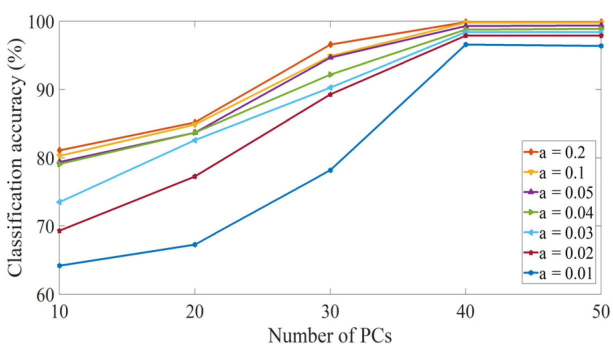


Fig. 4. Classification accuracies for different compressively sampled signals versus the number of PCs.

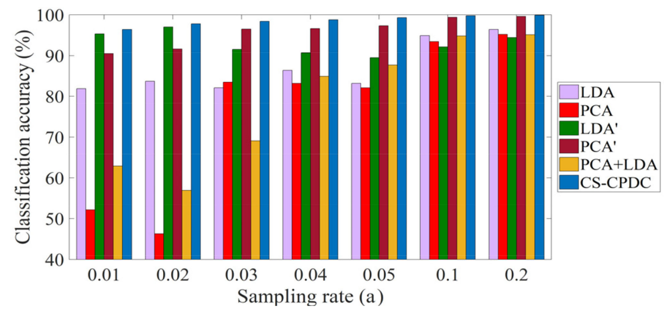


Fig. 5. Comparison of classification performance using individual and combined features.

TABLE II
CLASSIFICATION COMPARISON OF BEARING FAULTS

Methods	Sampling rate (α)	Testing Accuracy (%)
Compressed Sensed followed by signal recovery [24]	0.5 0.25	92.4 ± 0.5 84.6 ± 3.4
CS-CPDC	0.2 0.1	99.9 ± 0.1 99.8 ± 0.2

TABLE III
CLASSIFICATION COMPARISON OF BEARING FAULTS

Methods		Testing Accuracy (%) α = 0.1
DNN based Over-complete feature representations [26]	2 - layers	99.8 ± 0.1
	3 - layers	100 ± 0.0
	4 - layers	100 ± 0.0
CS [27]		98.6± 0.3
CS-PCA [27]		98.5± 0.4
CS-LDA [27]		89.8± 3.5
CS-CPDC		99.8± 0.2

$\alpha = 0.2, 0.1, 0.05, 0.04, 0.03, 0.02,$ and 0.01 using different number of PCs in the range 10–50. Fig. 4 shows the classification accuracies versus the number of PCs for each value of α. It is clear that most of the compressively sampled signals require no more than 40 PCs to achieve high classification accuracy.

D. Comparison of Classification Performance Using Individual and Combined Features

In the second stage of the proposed multistep features learning approach, four groups of features were extracted individually before the features concatenation step. These include PCA-based features $(\widehat{Y1})$, LDA-based features $(\widehat{Y2})$, and the linear combinations features $\hat{Y}1_{\mathrm{CCA}}$ and $\hat{Y}2_{\mathrm{CCA}}$ of $\widehat{Y1}$ and $\widehat{Y2}$; these features will be referred as PCA' and LDA', respectively. Experiments are conducted using these features based on PCA, LDA, PCA', LDA', and concatenated features of PCA and LDA (PCA+LDA) with CS-CPDC to classify bearing faults. The test classification results are displayed in Fig. 5 and achieved by averaging the results of 20 trials for each experiment. Closer inspection of Fig. 5 shows significant improvements in the classification accuracy achieved by PCA' and LDA' compared to PCA and LDA, respectively. However, classification results from the CS-CPDC method achieved the best classification results for each value of α.

E. Comparison of Results

Different from the methods in [24] that recovered the original signals from the compressively sampled signals, CS-CPDC learns features directly from the compressed measurements. To show the superiority of CS-CPDC over the methods in [24], Table II shows classification results of bearing faults using the same dataset used in the first case study of this work. It is clear that all results from CS-CPDC are better than those in [24]. For further verification of the efficiency of the CS-CPDC method, Table III presents the comparisons with some recently published results of CS-based methods [26], [27]. It is clear that results from CS-CPDC remain very competitive with results reported in [26]. Moreover, the classification accuracy obtained using our proposed method with 10% of the original data is better than the results reported in [27] and the improvement is statistically significant. In particular, results from [27] are 98.6, 98.5, and 89.8 for CS, CS-PCA, and CS-LDA, while our proposed method achieved 99.8%. Additionally, the root mean square

TABLE IV
COMPARISON RESULTS TO EXAMINE THE SPEED AND ACCURACY PERFORMANCE OF OUR PROPOSED METHOD WITH CS AND WITHOUT CS

Methods	Classification accuracy (%)	Time (s)
CS-CPDC without CS (with 6000 inputs from FT)	99.4 ± 0.5	64.9± 0.3
CS-CPDC with CS α = 0.1 (600 samples)	99.8± 0.2	6.7± 0.1

error (RMSE) of this method (0.2%) is smaller than the reported RMSEs in [27] (e.g., 0.3%, 0.4%, and 3.5%).

F. Need for CS

In CS-CPDC, CS is employed to obtain compressively sampled signals in the first stage motivated by the following: 1) reduced computations—we used CS to reduce a large amount of the acquired vibration data. This reduction in the amount of vibration data resulted in much reduced computation, i.e., less than 15% of the computation load of not using CS; 2) reduced transmission costs—in the cases of having to send vibration data from remote places by wireless (e.g., in the case of offshore wind turbines) or wired transmissions, the cost of transmission will be less as CS reduces the amount of vibration data; and 3) benefits to the environment—as the application of CS results in reduced computations, it helps to reduce the amount of power needed for both computations and transmission. In consequence, CS offers much benefit to the environment.

Several experiments were conducted to classify bearing fault using CS-CPDC without the compression in the first stage, i.e., with all the 6000 original samples from their sparse (FT) domain. Table IV contains the results where the classification accuracy is the average of 20 trials of testing accuracy. Two things are clear from the results presented in Table IV. First, CS-CPDC with CS is significantly faster than (or requires less than 15% of the time of) the proposed method without CS. Second, CS-CPDC achieved better classification results with small RMSE. Of the remaining 0.6% (100%–99.4%) accuracy, our method can make up $2/3\,(=(99.8-99.4)/(100-99.4))$ of the missed accuracy, and it does so with significantly lower RMSE, i.e., 0.2 compared to 0.5. Thus, the increase in accuracy of 0.4% with RMSE of 0.2% is statistically significant.

IV. SECOND CASE STUDY

The bearing datasets [35] used in this case come from a motor driving mechanical system where the faults were seeded into the drive-end bearing of the motor. The bearing vibration signals were collected under several conditions, grouped by fault width (0.18, 0.36, and 0.53 mm) and four motor loads (0, 1, 2, and 3 hp) with different shaft speeds (1797, 1772, 1750, and 1730 r/min). The sampling rate used was 12 kHz [35], [36].

In this study, a motor bearing dataset composed of these vibration signals with 10 bearing health conditions and 100 signal examples for each health condition per load. Thus, the total dataset contains 400 examples for each health condition, with 1200 data points for each signal example. The description of this dataset is presented in Table V.

TABLE V
DESCRIPTION OF BEARING HEALTH CONDITIONS

Health Condition	Number of samples	Fault Width (mm)	Classification Label
NO	400	0	1
OR1	400	0.18	2
OR2	400	0.36	3
OR3	400	0.53	4
IR1	400	0.18	5
IR2	400	0.36	6
IR3	400	0.53	7
RE1	400	0.18	8
RE2	400	0.36	9
RE3	400	0.53	10

TABLE VI
CLASSIFICATION RESULTS WITH THEIR CORRESPONDING RMSE AND COMPUTATIONAL TIME USING VARIOUS COMPRESSED SAMPLING RATES

Sampling rate (α)	The cross-validated accuracy (%)	Training time (s)	Testing accuracy (%)	Testing time (s)
α = 0.1 120 samples	100	4.83 ± 0.01	99.9 ± 0.1	1.62 ± 0.02
α = 0.09 108 samples	100	4.72 ± 0.01	99.8 ± 0.2	1.55 ± 0.04
α = 0.08 96 samples	99.9 ± 0.1	4.62 ± 0.12	99.8 ± 0.1	1.47 ± 0.08
α = 0.07 84 samples	99.9 ± 0.1	4.61 ± 0.01	99.6 ± 0.2	1.41 ± 0.02
α = 0.06 72 samples	99.8 ± 0.2	4.50 ± 0.01	99.6 ± 0.2	1.32 ± 0.07
α = 0.05 60 samples	99.6 ± 0.2	4.36 ± 0.09	99.2 ± 0.3	1.24 ± 0.07
α = 0.04 48 samples	99.3 ± 0.3	4.24 ± 0.07	99.2± 0.5	1.19 ± 0.14
α = 0.03 36 samples	99.1 ± 0.4	4.20 ± 0.10	98.5± 0.8	1.15 ± 0.02

To classify the motor bearing health condition, the same steps as in the first case study were followed to apply the CS-CPDC method. Half of the signal examples are selected randomly for training, and the rest of the signal examples are utilized for testing performance. Different compressed samples with 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, and 0.1 sampling rates (α) of the original signals and ten selected PCs are used for the overall classification results and their related RMSEs of 20 trials as shown in Table VI.

In Table VI, CS-CPDC delivers high classification accuracies with small RMSEs. In particular, the classification accuracy for $\alpha = 0.1$ is 99.9%, and the testing time required by CS-CPDC is only 1.62 s. In general, the computation time increased slightly with the increase in α value. For example, the total time for training and testing with the smallest value of $\alpha = 0.03$ (5.35 s) increased by less than 20% compared with the total time required by the largest value of $\alpha = 0.1$ (6.45 s).

For further evaluation of CS-CPDC method, experiments were conducted for $\alpha = 0.1$ with training size of 10% and 40%, and 20 trials for each experiment with tenfold cross validation.

TABLE VII
OUR RESULTS WITH TENFOLD CROSS VALIDATION AND RESULTS FROM LITERATURE ON CASE WESTERN RESERVE UNIVERSITY VIBRATION DATASETS OF ROLLER BEARINGS

	Methods	Training size (%)	No of classes	Load (hp)	Testing accuracy (%)
Fixed load	[37]	10	10	3	92.5
	[38]	35	10	3	92.65
	[39]	N/A	10	0	97.89
	[40]	75	10	0	88.9
	CS-CPDC	10	10	3	99.8 ± 0.2
				0	99.8 ± 0.1
Variable loads	[18]	40	4	0,1,2,3	95.8
	[41]	75	9	0,1,2,3	97.59
	[21]	40	11	0,1,2,3	97.91±0.09
	[42]	70	10	0,1,2,3	99.44
	CS-CPDC	40	10	0,1,2,3	99.9 ± 0.1

TABLE VIII
COMPARISON RESULTS TO EXAMINE THE SPEED AND ACCURACY PERFORMANCES OF OUR PROPOSED METHOD WITH TENFOLD CROSS VALIDATION

Methods		TsCA	Time (s)
[22]		99.66 ± 0.19	12
CS-CPDC	α = 0.1	99.8 ± 0.1	5.28
	α = 0.09	99.8 ± 0.1	4.92
	α = 0.08	99.7 ± 0.2	4.74

The results are compared to some recently published results [18], [21], [37]–[42] in Table VII. The first column refers to the scenarios of the motor operation and load conditions (fixed load and variable loads) in which the bearing samples collected. The second column defines the methods used for classification. The third column records the percentage of samples used to train these methods. The fourth column defines the related load of the data, and the fifth column records testing accuracies obtained using these methods.

Compared with the methods presented in Table VII, CS-CPDC with the smallest percentage (10%) of samples of the original data achieved the highest accuracy in both motor operation condition, i.e., fixed load and variable load.

For additional comparison, several experiments were conducted for $\alpha = 0.1, 0.09$, and 0.08, using variable loads bearing dataset (0, 1, 2, 3 loads) with ten classes to examine the speed and accuracy of CS-CPDC compared to the method in [22] that used the same data. To match the experimental setup in [22], only 10% of signal examples are used for training with tenfold cross validation; the testing classification accuracies and computation times were obtained by averaging 20 trials in each experiment. The results, as shown in Table VIII, indicate that the CS-CPDC method is significantly faster than the method in [22] and yet our classification accuracies, for all values of α, are better than the results reported in [22].

V. CONCLUSION

A three-stage hybrid method CS-CPDC for bearing fault diagnosis was proposed. In this method, CS model was used to generate compressively sampled signals. Then, a multistep feature learning algorithm was used to learn fewer features from correlated principal and discriminant attributes from the compressively sampled signals, which were then concatenated to increase the performance. These concatenated features were input to the multiclass SVM to train, validate, and classify bearing health conditions. Our method offers the highest accuracy, and requires least execution time and storage.

One of the more significant findings from this study is that CS-CPDC outperforms other CS-based methods that used principal and discriminant components individually as in [27]. As well as achieving satisfactory high classification results using the learned linear combinations of the principal and discriminant components in the second step of the second stage of the proposed method, the concatenated features in the final step of the second stage achieves even higher classification accuracies. Moreover, compared to existing non-CS-based methods, the proposed method offers superior performance in both increased accuracy and reduced computation time.

ACKNOWLEDGMENT

The authors would like to acknowledge interesting discussions with Dr. M. Abbod of Brunel University London at the early stages of this work.

REFERENCES

[1] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.
[2] P. J. Tavner, "Review of condition monitoring of rotating electrical machines," *IET Elect. Power Appl.*, vol. 2, pp. 215–247, 2008.
[3] M. Kang, J. Kim, L. M. Wills, and J. M. Kim, "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7749–7761, Dec. 2015.
[4] A. K. Nandi, C. Liu, and M. Wong, "Intelligent vibration signal processing for condition monitoring," in *Proc. Int. Conf. Surveillance 7*, 2013, pp. 1–5.
[5] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
[6] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.
[7] Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, 2013.
[8] H. Guo, L. B. Jack, and A. K. Nandi, "Feature generation using genetic programming with application to fault classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 1, pp. 89–99, Feb. 2005.
[9] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.
[10] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
[11] P. Comon, "Independent component analysis a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.
[12] S. Balakrishnama, and A. Ganapathiraju, *Linear Discriminant Analysis—A Brief Tutorial*. Starkville, MS, USA: Inst. Signal Inf. Process., Mississippi State Univ., 1998.
[13] L. B. Jack and A. K. Nandi, "Genetic algorithms for feature selection in machine condition monitoring with vibration signals," *Proc. IEE—Vis. Image Signal Process.*, vol. 147, no. 3, pp. 205–212, 2000.

[14] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[15] 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.

[16] L. Ciabattoni, G. Cimini, F. Ferracuti, A. Freddi, G. Ippoliti, and A. Monteriú, "A novel LDA-based approach for motor bearing fault detection," in *Proc. 2015 IEEE 13th Int. Conf. Ind. Informat.*, Cambridge, MA, USA, 2015, pp. 771–776.

[17] M. Amar, I. Gondal, and C. Wilson, "Vibration spectrum imaging: A novel bearing fault classification approach," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 494–502, Jan. 2015.

[18] W. Li, S. Zhang, and G. He, "Semisupervised distance-preserving self-organizing map for machine-defect detection and classification," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 869–879, May 2013.

[19] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform support vector machine and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.

[20] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[21] X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.

[22] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[23] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[24] M. L. D. Wong, M. Zhang, and A. K. Nandi, "Effects of compressed sensing on classification of bearing faults with entropic features," in *Proc. 2015 23rd Eur. Signal Process. Conf.*, Nice, France, 2015, pp. 2256–2260.

[25] X. Zhang, N. Hu, L. Hu, L. Chen, and Z. Cheng, "A bearing fault diagnosis method based on the low-dimensional compressed vibration signal," *Adv. Mech. Eng.*, vol. 7, no. 7, pp. 1–12, 2015.

[26] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features," *Mech. Syst. Signal Process.*, vol. 99, pp. 459–477, 2018.

[27] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Compressive sensing strategy for classification of bearing faults," in *Proc. 2017 IEEE Int. Conf. Acoust., Speech, Signal Process.*, New Orleans, LA, USA, 2017, pp. 2182–2186.

[28] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[29] J. Chen, J. Ye, L. Sun, and J. Liu, "Efficient recovery of jointly sparse vectors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1812–1820.

[30] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[31] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2664–2699, 2004.

[32] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[33] Fit Multiclass Models for Support Vector Machines or Other Classifiers—MATLAB fitcecoc, Mathworks.com, 2017. [Online]. Available: http://uk.mathworks.com/help/stats-/fitcecoc.html

[34] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[35] Case Western Reserve University Bearing Data Center. [Online]. Available: http://csegroup.case.edu/bearingdatacenter/home. Accessed on: 21 Sep. 2017.

[36] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64/65, pp. 100–131, 2015.

[37] 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.

[38] M. Van and H. J. Kang, "Wavelet kernel local fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 12, pp. 3588–3600, Dec. 2015.

[39] M. Van and H. J. Kang, "Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition," *Proc. Inst. Mech. Eng. C—J. Mech. Eng. Sci.*, vol. 230, pp. 291–302, 2016.

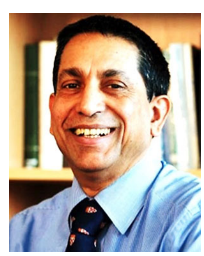
[40] W. Du, J. Tao, Y. Li, and C. Liu, "Wavelet leaders multifractal features based fault diagnosis of rotating mechanism," *Mech. Syst. Signal Process.*, vol. 43, no. 1, pp. 57–75, 2014.

[41] M. Xia, T. Li, L. Liu, L. Xu, and C. W. de Silva, "Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder," *IET Sci. Meas. Technol.*, vol. 11, no. 6, pp. 687–695, 2017.

[42] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.

Hosameldin O. A. Ahmed received the B.Sc. (Hons.) degree in engineering technology, specialization in electronic engineering, and the M.Sc. degree in computer engineering and networks from the Faculty of Science and Technology, University of Gezira, Wad Madani, Sudan, in 1999, and 2010, respectively. He is currently working toward the Ph.D. degree in electronic and computer engineering with Brunel University London, London, U.K.

Since 2014, he has been working closely with Prof. Asoke K. Nandi on the area of machine condition monitoring. Their collaboration has made an important contribution to the advancement of vibration-based machine condition monitoring using compressive sampling and modern machine learning algorithms. His work has been published in two journal papers and several high-quality international conferences. His research interests include signal processing, compressive sampling, and machine learning with application to vibration-based machine condition monitoring.



Asoke K. Nandi (F'11) received the Ph.D. degree in physics from the University of Cambridge (Trinity College), Cambridge, U.K., in 1979.

He held academic positions in several universities, including Oxford, Imperial College London, Strathclyde, and Liverpool as well as Finland Distinguished Professorship in Jyvaskyla (Finland). In 2013, he moved to Brunel University London, London, U.K., to become the Chair and Head of Electronic and Computer Engineering. In 1983, he codiscovered the three fundamental particles known as $W+$, $W-$ and Z0 at CERN, for which the Nobel Prize for Physics in 1984 was awarded to his two team leaders for their decisive contributions. He has authored more than 550 technical publications, including 220 journal papers as well as four books, e.g., *Automatic Modulation Classification: Principles, Algorithms and Applications* (New York, NY, USA: Wiley, 2015) and *Integrative Cluster Analysis in Bioinformatics* (Wiley, 2015). The h-index of his publications is 70 (Google Scholar) and ERDOS number is 2. His current research interests include signal processing and machine learning, with many data science applications.

Dr. Nandi is a Fellow of the Royal Academy of Engineering (U.K.) and of seven other institutions. He is an IEEE Distinguished Lecturer (2018–2019).