Received July 2, 2018, accepted August 2, 2018, date of publication August 13, 2018, date of current version September 5, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2865116

Compressive Sampling and Feature Ranking Framework for Bearing Fault Classification With Vibration Signals

HOSAMELDIN AHMED[1] AND ASOKE K. NANDI[1,2], (Fellow, IEEE)

[1]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, U.K.

[2]Key Laboratory of Embedded Systems and Service Computing, College of Electronic and Information Engineering, Tongji University, Shanghai 200092, China

Corresponding author: Asoke K. Nandi (asoke.nandi@brunel.ac.uk)

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.

ABSTRACT Failures of rolling element bearings are amongst the main causes of machines breakdowns. To prevent such breakdowns, bearing health monitoring is performed by collecting data from rotating machines, extracting features from the collected data, and applying a classifier to classify faults. To avoid the burden of much storage requirements and processing time of a tremendously large amount of vibration data, the present paper proposes a combined compressive sampling (CS)-based on multiple measurement vector (MMV) and feature ranking framework to learn optimally fewer features from a large amount of vibration data from which bearing health conditions can be classified. The CS-based on the MMV model is the first step in this framework and provides compressively sampled signals based on compressed sampling rates. In the second step, the search for the most important features of these compressively sampled signals is performed using the feature ranking and selection techniques. For that purpose, we have investigated the following: 1) two compressible representations of vibration signals that can be used within CS framework, namely, fast Fourier transform-based coefficients and thresholded Wavelet transform-based coefficients and 2) several feature ranking and selection techniques, namely, three similarity-based techniques, fisher score, Laplacian score, Relief-F; one correlation-based technique, Pearson correlation coefficients; and one independence test technique, Chi-Square (Chi-2) to select fewer features that can sufficiently represent the original vibration signals. These selected features, in combination with three of the popular classifiers–multinomial logistic regression classifier, artificial neural networks, and support vector machines, have been evaluated for the classification of bearing faults. Results show that the proposed framework achieves high classification accuracies with a limited amount of data using various combinations of methods, which outperform recently published results.

INDEX TERMS Bearing fault classification, multiple measurement vector, compressive sampling, feature ranking, classification. algorithms.

I. INTRODUCTION

Rotating machines are at the core of most engineering process in industry and are used to accomplish numerous tasks. Unexpected machine failures or breakdowns will affect production plans, product quality, and production costs. For that reason, it is essential for industrialists to monitor machine health condition to avoid machine breakdowns. Rolling element bearings are critical components in rotating machine and their failures are amongst the main causes of machine breakdowns. Bearing vibration levels, coolant temperatures, line currents, and voltages are among the most quantities measured and used for rotating machine Condition Monitoring (CM) [1]. Of these measurements, bearing vibration signals provide various features that make it one of the most widely used techniques for fault diagnosis [2].

The aim of vibration based CM is to classify the acquired vibration signal into the matching condition correctly by means of a classification algorithm which is usually a multi-class classification problem [3]. To monitor machine health condition using vibration signals the following

procedure is commonly used. First, typical vibration signals need to be collected from an operating machine of interest through vibration sensors, e.g., displacement sensors, velocity sensors, and accelerometers. Second, the characteristics of the vibration signals need to be examined by using signal processing techniques. However, in most of the modern industrial rotating machines, the acquired vibration signal represents a large amount of time series data that make the processing to become extremely difficult. Accordingly, rather than processing a large amount of vibration data, the common methodology is to extract certain features of the raw vibration signals that are able to adequately describe the signal of the machine health condition. Also, depending on the number of the extracted features, previous research has established that one may possibly want to implement more filtering to select a minimum subset of features using a feature selection algorithm. Finally, with these features, the presence of a fault in the machine can be detected in advance before the machine breakdown happens using a classification algorithm that has the ability to classify the health condition of the machine of interest. The overall framework of machine CM using this methodology is presented in Fig. 1.

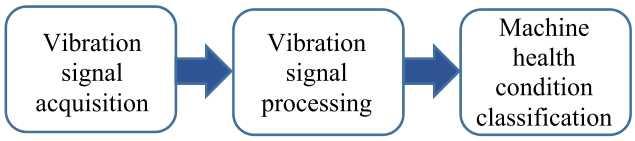

FIGURE 1. The overall framework of vibration based machine condition monitoring.

Features from the raw vibration signal can be extracted using various techniques that based in three main domains. First, time domain based techniques, that extract features utilizing some statistical factors, e.g., Impulse factor, skewness, kurtosis, peak-to-peak value, crest factor, root mean square, etc. Second, frequency-domain based techniques that can be used to observe frequency features, e.g., baseband auto-spectral density, linear frequency spectrum, and phase-averaged linear spectra, which can be generated using Fourier Transform (FT) [4]. Third, time-frequency based techniques that have been used for the non-stationary signal type that is very common once fault takes place in a rotating machine. The literature on time-frequency based techniques has highlighted several techniques including, Short Time Fourier Transform (STFT), Wavelet Transform (WT), Hilbert-Huang Transform (HHT), Local Mean Decomposition (LMD), Winger-Ville Distribution, etc. [5]. Based on the time-frequency domain, Spectral Kurtosis (SK) based techniques that have the ability to automatically identify which frequency bands of a vibration signal have larger impulsivity have been widely used in fault diagnosis [6]–[8].

As previously stated, the modern large-scale rotating machines in industry generate large amounts of vibration signals for the purpose of CM. As a result, various dimensionality reduction techniques of features extraction and features selection have been proposed and effectively used in machine fault diagnosis. For instance, Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA) are amongst the most frequently utilised techniques. For example, Malhi and Gao [9] developed a PCA-based approach to select the most representative features for classification of faults in three types of roller bearings. Jin *et al.* [10] introduced trace ratio LDA to deal with high-dimensional non-Gaussian fault data of roller bearings. Ciabattoni *et al.* [11] introduce a novel LDA based algorithm to deal with fault data dimension reduction and fault detection problems. Widodo *et al.* [12] developed a method that combined ICA and SVM for fault diagnosis of induction motors. In a similar way, Chang and Jiao [13] also found a combination of Neural Network (NN) and ICA can achieve a considerable classification accuracy of rotating machinery fault diagnosis. Ahmed *et al.* [14] conducted a series of trials in which Deep Neural Network (DNN) is employed to extract features from vibration signals in order to classify bearing faults.

Also, feature selection techniques that have the ability to remove the irrelevant or redundant features that may slow the learning process can play an important role in machine fault diagnosis. In most cases, feature selection techniques are performed after the feature extraction step. For example, Van and Kang [15] proposed a method comprises three main steps, first, a feature extraction technique based on non-local-means denoising and EMD is used for feature extraction. Second, a feature selection technique combining Distance Evaluation Technique (DET) and Particle Swarm Optimization (PSO) is employed to select the superior feature subset. Jack and Nandi [16] apply Genetic Algorithm (GA) to select features from different feature sets using different forms of preprocessing.

In recent years, there has been an increasing amount of literature on roller bearings fault diagnosis using vibration signals. For example, Amar *et al.* [17] suggested a novel bearing fault classification approach combining Vibration Spectrum Imaging (VSI) and Artificial Neural Network (ANN). In another study, Li et al. [18] presented a semi-supervised diagnosis method based on a distance-preserving Self-Organizing Map (SOM) for classifying different bearing faults. Soualhi *et al.* [19] examined the combination of Hilbert-Huang Transform (HHT), SVM, and Support Vector Regression (SVR), and showed its efficiency for the condition monitoring of ball bearing. In a different study, Chen and Li [20] proposed a multisensory feature fusion method for bearing fault using Sparse Autoencoder (SAE) and Deep Belief Network (DBN) that outperform some other feature fusion methods. Zhang *et al.* [21] presented a hybrid intelligent fault diagnosis method integrating Permutation Entropy (PE), Ensemble Empirical Mode Decomposition (EEMD), and optimized SVM. 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.

More recently, Zhang *et al.* [23] proposed a transfer learning method based on neural networks for fault diagnosis of roller bearings. Yu *et al.* [24] presented a procedure for bearing fault conditions classification using Empirical Mode Decomposition (EMD), Feature Selection by Adjusted Index and Standard Deviation Ratio (FSASR), Support Margin Local Fisher Discriminant Analysis (SM-LFDA) as feature dimensionality reduction technique, and SVM for fault classification. Nayana and Geethanjali [25] investigated several statistical time domain features including mean absolute value (MAV), simple sign integral (SSI), waveform length (WL), Wilson amplitude (WAMP), zero crossing (ZC), slope sign changes (SSC) for bearing faults identification using LDA, NB, and SVM classifiers.

All the above methods use data that have been collected satisfying the Shannon/Nyquist sampling theorem, in which the sampling rate must be at least twice the maximum frequency present in the signal. It is clear that collecting a large amount of data requires large storage and time for signal processing and this also may limit the number of machines that can be monitored remotely across wireless sensor networks (WSNs) due to bandwidth and power constraints. A reasonable approach to tackle the challenges involved in dealing with a large amount of data could be to compress the data. Recently, Compressive Sampling [26] has been developed for sensing and compression.

The efficiency of CS in machine fault diagnosis has been validated in a number of studies. For instance, Wong *et al.* [27] studied the effects of CS on the classification of bearing faults and found a small performance degradation when using entropic features computed from CS based recovered signal. Xinpeng *et al.* [28] developed a bearing fault detection method based on CS and Matching Pursuit (MP) reconstructing algorithm. Tang *et al.* [29] proposed an interesting approach in which authors attempted to observe the characteristic harmonics from sparse measurements through a compressive matching pursuit strategy during the process of incomplete reconstruction. Zhang *et al.* [30] developed a technique based on compressed vibration signal by using several over-complete dictionaries that can be effective in sparse signal decomposition for a specific bearing condition. Tang *et al.* [31] proposed a sparse classification strategy that sampled the original characteristics of a vibration signal by applying a small number of random projections and then constructed a learning redundant dictionary to sparsely represent the vibration signal. Moreover, in the literature on CS based fault diagnosis methods, several attempts have been made to learn directly from compressed measurements without reconstructing the original signal. For example, in [32] an intelligent condition monitoring method for bearing faults based on CS and sparse over-complete feature learning algorithm using SAE was proposed. In a recent paper by Ahmed *et al.* [33], three approaches to process compressed vibration measurements were proposed for classification of bearing faults, using the compressed measurements directly as the input to the classifier and extracting features from these compressed measurements using PCA and LDA.

Even though the aforementioned studies reported many interesting results, there are two main problems with these studies: (1) CS-based sparse signal reconstruction is a complex computational problem that depends on the sparsity of the measured vibration signal. Therefore, CS-based signal recovery methods may not be useful in reducing computational complexity for condition monitoring of rolling bearings, and (2) most of the methods that are based on learning directly from the compressed measurements achieved good classification accuracy but by increasing the sampling rate, thereby requiring higher computational complexity.

In this work, we argue that despite the fact that the obtained CS-based compressed measurements are able to recover the original signal, they may not provide the best bearing fault classification. Moreover, these compressed measurements may still represent a large amount of data collected in real operating condition. In our earlier work [55] and [56] where FFT-based CS is combined with LS and FS, and the classification is achieved using SVM. This paper proposes a combined CS based on MMV model and Feature Ranking (FR) framework to classify bearing health conditions. In this framework, CS is used to reduce the amount of the original signal by obtaining compressively-sampled signals that possess the quality of the original signal. Then, a feature ranking technique is employed to further filter the obtained compressively-sampled signals by ranking their features and select a subset of fewer most significant features. In this manner, we are able to reduce the large amount of the collected vibration signals and avoid spending much time on computing eigenvalues that are included in most feature extraction algorithms, e.g., PCA and LDA.

Based on our proposed framework, we considered two techniques of feature selection to select fewer features of the compressively sampled signals. These are:

(1) Similarity based methods: that assign similar values to the compressively sampled signals that are close to each other. Three algorithms (LS, FS, and Relief-F algorithms) were investigated to select fewer features based on similarity.

(2) Statistical based methods: that measure the importance of feature of the compressively sampled signals using different statistical measures. Two algorithms, PCC and Chi-2 were investigated to select fewer features based on correlation and independence test respectively.

Various experiments were conducted to: (1) validate our proposed framework by investigating different scenarios of combinations of CS, feature ranking techniques, and classifiers. These are, a) FFT-based CS and thresholded WT-based CS to obtain compressively sampled measurements, b) several feature ranking techniques with different feature selection criterion, namely, FS, LS, Relief-F, PCC, and Chi-2 to rank and select fewer features from the compressively sampled measurements, and c) LRC, ANN, and SVM classifiers to deal with the classification problem, and (2) observe the

best combinations of MVM based-CS, feature ranking and selecting techniques, and classifiers with reduced complexity and improved classification accuracy.

The remainder of this paper is organised as follows. Section II describes briefly the theoretical background of CS, different feature ranking methods, and different classification algorithms used in this study. Section III is devoted to a description of the proposed framework. The experiments on two case studies of bearing faults classification is presented in section IV, with the corresponding results and comparisons with recently published results using the same datasets. Finally, section V draws some conclusions.

II. METHODS

A. COMPRESSIVE SAMPLING (CS)

Compressive sampling, also named "compressed sensing" or "compressed sampling" [26], [34], is an extension of sparse representation. The simple idea of CS is based on the fact that many real-world signals have sparse representations in some domain, e.g., Fourier Transform (FT), and can be recovered from a small number of samples in certain conditions. CS is based on two concepts: (1) sparsity of the signals, and (2) the measurements matrix to be used for compression of the original signals based on their sparse representations. This measurement matrix must satisfy the data minimal information loss, i.e., satisfy Restricted Isometry Property (RIP) to ensure the signal recovery from the compressed measurements. Briefly, we describe the standard CS framework as follows. Assume that we have the original signal vector (x) where $x \epsilon R^{nx1}$. To compute a set of sparse representations of x we need to apply a sparsifying transform ψ such that

$$x = \psi s \tag{1}$$

Here s is a $n \times 1$ column vector that has a small number of nonzero coefficients and means the sparse features. Based on CS theory, the signal x can be recovered from its compressed samples $y \epsilon R^{mx1}$ ($m \ll n$) that can be computed using the following equation:

$$y = \phi \psi s \tag{2}$$

Here ϕ is the measurement matrix that has to be incoherent with the sparsifying transform ψ, i.e., satisfy Restricted Isometry Property (RIP).

Definition 1: The measurement matrix ϕ satisfies the Restricted Isometry Property (RIP) if there be existent a parameter $\delta \in (0, 1)$ such that

$$(1 - \delta) \|s\|_2^2 \leq \|\phi s\|_2^2 \leq (1 + \delta) \|s\|_2^2 \tag{3}$$

The measurement matrix size is (mxn) and is based on the compressive sampling rate (α) (i.e., $m = \alpha^* n$). The estimation of s can be performed by solving an optimization problem using L1-norm such that

$$\hat{s} = \min_{s \in R^N} \frac{1}{2} \|\phi \psi s - y\|_{l2}^2 + \gamma \|s\|_{l1} \tag{4}$$

With $\|\phi \psi s - y\|_{l2}^2 \leq \varepsilon > 0$, and $\gamma > 0$ is a regularization parameter. Hence, the original vector x can be recovered by applying the inverse of the sparsifying transform ψ^{-1} to $\hat{s}$ such that

$$\hat{x} = \psi^{-1} \hat{s} \tag{5}$$

This model of CS is based on Single Measurement Vector Compressive sampling (SMV) that recovers one vector from its corresponding compressed measurement vector.

On the other hand, Multiple Measurement Vectors Compressive Sampling (MMV) model is considered where the data can be represented as a matrix with a set of jointly sparse vectors. Thus, CS based on MMV can be computed as follows:

$$Y = DS \tag{6}$$

where $Y \in R^{mxL}$ is multiple measurement vectors, $D \in R^{mxn}$ is a dictionary, and $S \in R^{nxL}$ is a sparse representation matrix. Furthermore, matrices Y and S can be represented as follows: $Y = \{y_1, y_2, \ldots, y_L\}$, and $S = [s_1, s_2, \ldots, s_l]$, where $1 \leq l \leq L$, y_l's and s_l's are column vectors. Several studies have been conducted to reconstruct jointly sparse signals (S) given multiple compressed measurement vectors [35], [36]. Similarly, the original signal matrix X can be recovered utilising the inverse of the sparsifying transform and $\hat{S}$ such that

$$\hat{X} = \psi^{-1} \hat{S} \tag{7}$$

Here $\hat{X}$ and $\hat{S}$ is the estimation of X and S respectively.

In this study, we used CS based on MMV model since the measured vibration signals is a mixture of vibrations from several rotating components or from different positions with one rotating part. Besides, the signal recovery in both types of CS models indicate that the compressed measurements have enough information of the original signal, i.e., possess the quality of the original signal.

B. FEATURE RANKING AND SELECTION

Feature selection techniques are used to select a subset of features that can sufficiently represent the characteristic of the original features. In view of that, this will reduce the computational cost, and may remove irrelevant and redundant features. Feature selection methods can be categorised into three groups, supervised, unsupervised, and semi-supervised feature selection techniques. Also, it can be further grouped into filter models, wrapper models, embedded models, and hybrid models. Of these models, filter-based techniques are fast and require low computational complexity. The filtering can be performed using univariate feature filters that rank each single feature or using multivariate feature filters which evaluate a feature subset [37], [38]. This section gives brief descriptions of the feature ranking methods that used to rank the compressively-sampled signals features in this study.

1) FISHER SCORE

Fisher score (FS) [39] is a filter-based feature selection method and one of the commonly used supervised feature selection methods. The main idea of FS is to compute a subset of features with a large distance between data points in different classes and small distance between data points in the same class. To describe briefly FS method, assume the input matrix $Y \in R^{mxL}$ reduces to $Z \in R^{dxL}$ matrix. The FS of the *i-th* feature can be computed by the following equation:

$$FS(Y^i) = \frac{\sum_{c=1}^{C} L_c(\mu_c^i - \mu^i)^2}{(\sigma^i)^2} \tag{8}$$

where $Y^i \in R^{1\ xL}$, L_c is the size of the *c-th* class, $(\sigma^i)^2 = \sum_{c=1}^{C} L_c(\sigma_c^i)^2$, μ_c^i and σ_c^i are the mean and standard deviation of *c-th* class corresponding to the *i-th* feature; μ^i and σ^i are the mean and standard deviation of the entire dataset corresponding to the *i-th* feature.

Usually, FS of each feature is computed independently. As shown in [36] to generalise FS to select features jointly that maximize the lower bound of FS, the following optimization problem is introduced,

$$\begin{aligned} \mathrm{FS}(W, p) = tr\Big\{ & (W^T diag(p) S_b diag(p) W) \\ & (W^T diag(p) (S_t + \gamma I) diag(p) W)^{-1} \Big\}, \\ \text{s.t., } & p \in \{0,1\}^m, \quad p^T 1 = d \end{aligned} \tag{9}$$

where p is the feature selection vector, and d is the number of features to be selected. Also, the optimal p that maximizes Eq. (9) is the same as the optimal p that minimizes the following problem

$$\begin{aligned} & \min_{p,W} \frac{1}{2} \left\| W^T diag(p) Y - G \right\|_F^2 + \gamma \|W\|_F^2 \\ & s.t., \ p \in \{0,1\}^m, \quad p^T 1 = d \end{aligned} \tag{10}$$

Here G is a specific class indicator matrix such that

$$\mathrm{G(i, c)} = \begin{cases} \sqrt{\frac{L}{L_c}} - \sqrt{\frac{L_c}{L}} & if\ Y_i \in c \\ -\sqrt{\frac{L_c}{L}} & otherwise \end{cases} \tag{11}$$

2) LAPLACIAN SCORE

Laplacian Score (LS) is an unsupervised filter based technique that rank features depending on their locality preserving power. In fact, LS is mainly based on Laplacian Eigenmaps and Locality Preserving Projection, and can be briefly described as follows [40].

Given a dataset $Y = [y_1, y_2, \ldots, y_n]$, where $Y \in R^{mxn}$, suppose the Laplacian Score of the *r-th* feature is L_r and f_{ri} represent the *i-th* sample of the *r-th* feature where $i = 1, \ldots, m$ and $r = 1, \ldots, n$. First LS algorithm constructs the nearest neighbour graph G with m nodes, where the *i-th* node corresponds to y_i. Next, an edge between nodes i and j is placed, if y_i is among k nearest neighbors of y_j or vice versa, then i and j are connected. The elements of weight matrix of graph G is S_{ij} and can be defined as follows:

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & y_i = y_j \\ 0, & otherwise \end{cases} \tag{12}$$

The Laplacian score L_r for each sample can be computed as follows:

$$L_r = \frac{\tilde{f_r}^T L \tilde{f_r}}{\tilde{f_r}^T D \tilde{f_r}} \tag{13}$$

where $D = diag(S\mathbf{1})$ is the identity matrix, $\mathbf{1} = [1, \ldots, 1]^T$, $L = D - S$ is the graph Laplacian matrix, and $\tilde{f_r}$ can be calculated using the following equation:

$$\tilde{f_r} = f_r - \frac{f_r^T D \mathbf{1}}{\mathbf{1}^T D \mathbf{1}} \tag{14}$$

More details of the mathematical formulation of LS for feature selection can be found in [40].

3) RELIEF-F

Relief-F is a supervised feature ranking algorithm that commonly used as a pre-processing technique for a feature subset selection. Relief-F is an extension of the traditional Relief algorithm [41] that has the ability to deal with noisy, incomplete, and multi-class datasets. It uses a statistical approach to select the important features based on their weight W. The main idea of Relief-F is to randomly compute examples from the training data and then calculate their nearest neighbours from the same class, also called the nearest hit, and the other nearest neighbours from different class, also called the nearest miss. The procedure of Relief-F algorithm is summarized below in algorithm 1 [41].

Further explanation of the mathematical formulation of the Relief-F algorithm can be found in [41].

4) PEARSON CORRELATION COEFFICIENT

Pearson Correlation Coefficient (PCC) [41] is a supervised filter-based ranking technique that examines the relationship between two variables according to their correlation coefficient (r), -1 ≤ r ≤ 1. Here the negative values indicate inverse relations, the positive values indicate a correlated relation, and the value 0 indicates no relation. PCC can be computed as follows:

$$r(i) = \frac{cov(x_i, y)}{\sqrt{var(x_i) * var(y)}} \tag{15}$$

Here x_i is the i_{th} variable, y is the class labels.

5) CHI-SQUARED

Feature ranking and selection using chi-square (chi-2) is based on the χ^2 test statistics [42]. Chi-2 evaluate the importance of a feature by calculating the χ^2 test with respect to the

Algorithm 1 Relief-F

Input: ℓ learning instances, m features and c classes; Probabilities of classes $p_{y;}$ Sampling parameter a; Number of nearest instances from each class d;
Output: for each feature f_i a feature weight $-1 \leq W[i] \leq 1$;

1 for $i = 1$ to m do $W[i] = 0.0$; end for;
2 for $h = 1$ to a do
3 randomly compute an instance $\mathbf{x}_k$ with class y_k;
4 for y = 1 to c do
5 find d nearest instances x[j, y] from class y, $j = 1 \ldots d$;
6 for $i = 1$ to m do
7 for $j = 1$ to d do
8 if $y = y_k$ {nearest hit}
9 then $W[i] = W[i]$ – diff (i, $\mathbf{x}_k$, x[j, y])/ (a*d);
10 else $W[i] = W[i] + p_y/(1-p_{yk})$* diff($i$, $\mathbf{x}_k$, x[j, y])/ (a*d);

11 end if;
12 end for; {j} end for; {i}
13 end for; {y} end for; {h}
14 return (W);

class labels. The χ^2 value for each feature f in a class labels group c can be computed using the following equation:

$$\chi^2(f, c) = \frac{L(E_{c,f}E - E_cE_f)^2}{(E_{c,f} + E_c)(E_f + E)(E_{c,f} + E_f)(E_c + E)} \quad (16)$$

where L is the total number of examples, $E_{c,f}$ is the number of times f and c co-occur, E_f is the number of time the feature f occurs without c, E_c is the number of times c occurs without f, and E is the number of times neither f nor c occurs. The bigger value of χ^2 indicates that the features are highly related.

In this study, we applied the cross-tabulation function [43] that returns the chi-square statistic, and its p-value. The obtained values of chi-2 are sorted in descending order to create a new feature vector with ranked features.

C. CLASSIFICATION ALGORITHMS

In this study, we applied three supervised classifiers (LRC, ANN, and SVM) to classify between c classes based on an input vector of selected fewer feature x = [x_1, …, x_k]. The details are as follows.

1) MULTINOMIAL LOGISTIC REGRESSION CLASSIFIER (LRC)

Multinomial logistic regression [44], also called Softmax regression in ANN, is a linear supervised regression model that generalizes the logistic regression where labels are binary, i.e., $c^{(i)} \in \{0, 1\}$ to multi-classification problems that have labels $\{1 \ldots c\}c^{(i)} \in \{1, \ldots, K\}$ where c is the number of classes. Briefly, we present the simplified multinomial logistic regression model as follows:

Let there be a training set $\{(x^{(1)}, c^{(1)}), \ldots, (x^{(L)}, c^{(L)})\}$ of L labeled examples and input features $x^{(i)} \in R^n x^{(i)} \in R^k$. In logistic regression with binary labels,$c^{(i)} \in \{0, 1\}$, our hypothesis can be written as follows:

$$h_0(x) = \frac{1}{1 + exp(-\theta^T x)} \quad (17)$$

Here θ are model parameters that are trained to minimize the cost function $J(\theta)$ defined by the following equation

$$J(\theta) = -\left[\sum_{i=1}^{L} c^{(i)} \log h_\theta\left(x^i\right) + \left(1 - c^{(i)}\right) \times \log\left(1 - h_\theta\left(x^i\right)\right)\right] \quad (18)$$

In multinomial logistic regression with multi-labels $c^{(i)} \in \{1, \ldots, c\}c^{(i)} \in \{1, \ldots, K\}$ the aim is to estimate the probability $P(c = c^{(i)}|x)$ P(c = k|x) for each value of $c^{(i)} = 1$ to c, such that

$$h_\theta(x) = \begin{bmatrix} P(c = 1 \mid x; \theta) \\ P(c = 2 \mid x; \theta) \\ \cdot \\ \cdot \\ \cdot \\ P(c = K|x; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} \exp\left(\theta^{(j)T} x\right)} \begin{bmatrix} \exp\left(\theta^{(1)T} x\right) \\ \exp\left(\theta^{(2)T} x\right) \\ \cdot \\ \cdot \\ \cdot \\ \exp\left(\theta^{(K)T} x\right) \end{bmatrix} \quad (19)$$

where $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(K)} \in R^n$ are the parameters of the multinomial logistic regression model.

2) ARTIFICIAL NEURAL NETWORK (ANN)

ANN is a supervised learning algorithm that has the ability to learn real, discrete, and vector-valued target function [2]. It has been used successfully in bearing fault diagnosis, e.g., [45]–[48]. There are different types of ANN, e.g., Radial Basis Function (RBF), Probabilistic Neural Network (PNN), Multi-Layer Perceptron (MLP), etc. The MLP ANN (Fig. 2) is one of the most commonly used methods. As shown in Fig. 2, it involves an input layer, one to several hidden layers, and an output layer. Each layer consists of a number of neurons. The neuron receive inputs, multiply it by the weights

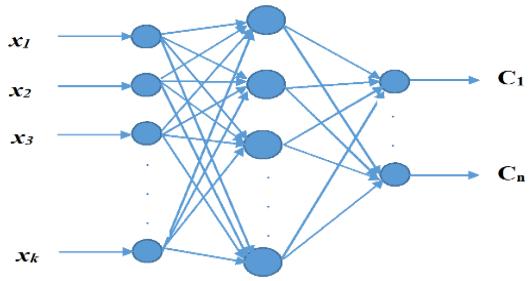

FIGURE 2. A Multilayer Perceptron Model for ANN.

of each input and combine the results of the multiplication. Then, the combined multiplications of the signals and weights are then passed to a transfer function to generate the output of the neuron (Fig. 3).

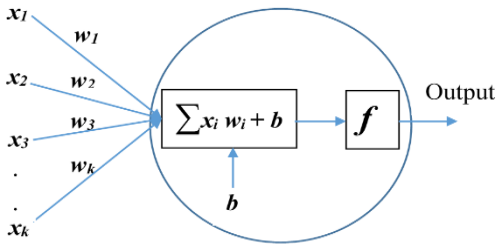

FIGURE 3. Model of an artificial neuron.

In this study, we used pattern recognition networks that are feedforward networks with one hidden layer and 10 neurons that trained using Scaled Conjugate Gradient (SCG) back-propagation function [49].

3) SUPPORT VECTOR MACHINE (SVM)

SVM is a supervised machine learning method that was first proposed for binary classification problem [50]. The basic idea of SVM is that it can find the best hyperplane(s) to separate the two classes. Based on the features of the data, SVM can make linear or non-linear classifications by using different kernel functions, e.g., Radial basis function (RBF), Polynomial function (PF), and Sigmoid function (SF) [51]. This is illustrated in Fig. 4.

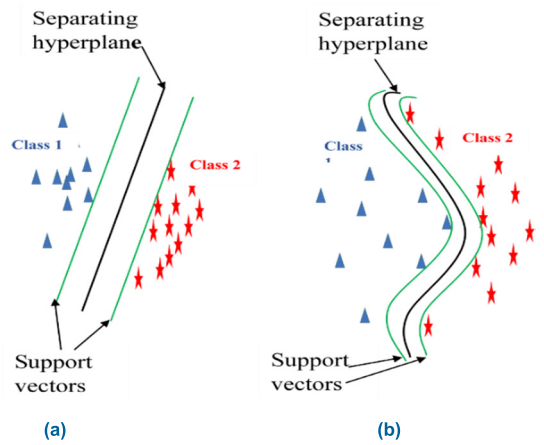

FIGURE 4. (a) Linear classifier, (b) Non-linear classifier.

For multiclass classification problems, several SVM classifiers can together deal with the multiclass problems. For example, one-against-all and one-against-one methods based on binary classification are commonly used in multiclass classification problems. A comparison of methods for multiclass SVMs can be found in [52].

In this study, we applied "fitcecoc" function [53] on the selected features using the aforementioned feature selection methods. The "fitcecoc" function 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 error-correcting output code (ECOC) multiclass model that is cross-validated using 10-fold cross-validation.

III. PROPOSED FRAMEWORK

In this study, we propose an original CS and feature ranking methods framework for bearing fault classification using vibration signals. Vibration signals are usually collected through vibration sensors. The three main types of vibration sensors are displacement sensors, velocity sensors, and accelerometers. As shown in Fig. 5, the proposed framework first compress the vibration data and then rank the features of the compressed data from which the most significant features can be selected to be used for classification. The details are as follows:

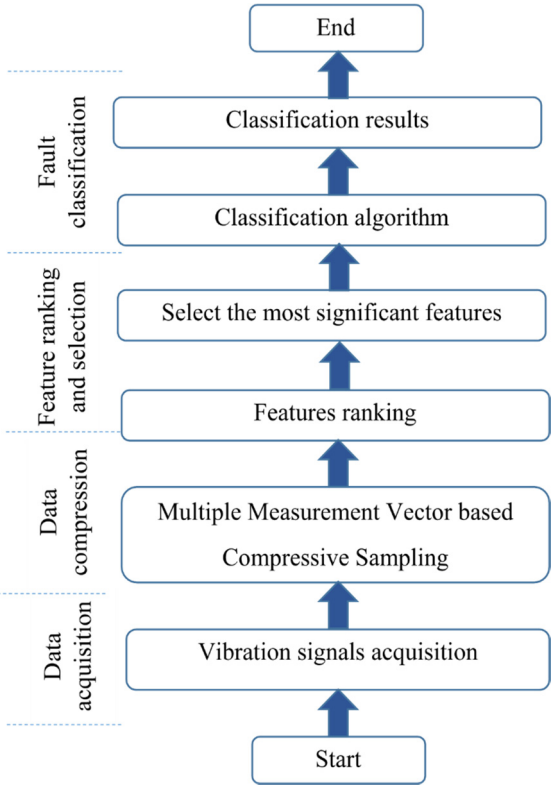

FIGURE 5. The proposed framework.

(1) Vibration data compression: CS based on MMV model is employed to produce compressively-sampled signals, i.e., compressed data $Y = \{y_1, y_2, \ldots, y_L\} \in R^m$ that have enough information of the original bearing raw data $X = \{x_1, x_2, \ldots, x_L\} \in R^n$. Here m $\ll$ n.

(2) Feature ranking and selection: as long as the compressively-sampled signals produced by CS model have enough information about the original vibration signals, we may further filter the

compressively-sampled signals using feature ranking and selection techniques to rank and select fewer features from the compressively-sampled signals that can sufficiently represent characteristics of bearing health conditions.

(3) Fault classification: with these fewer selected features a classifier is used to classify bearing health condition.

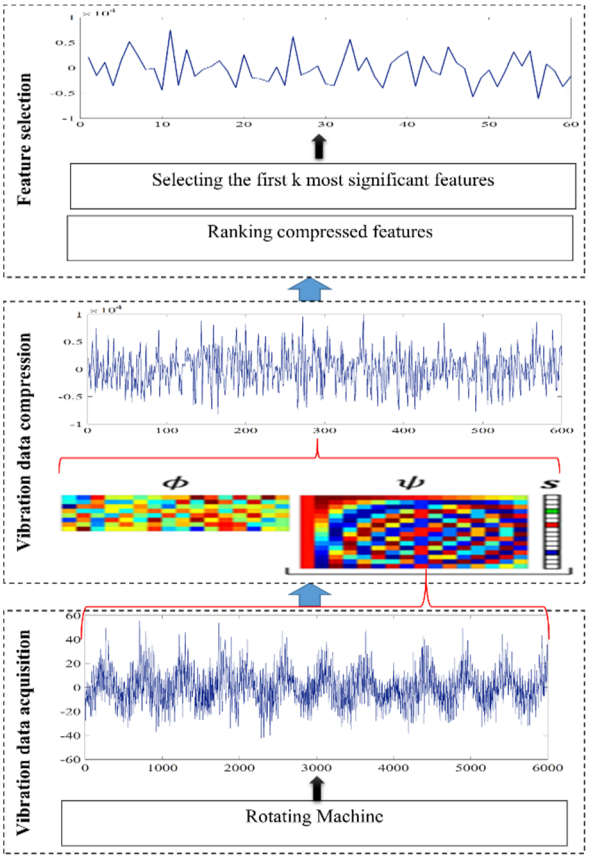

FIGURE 6. Illustration of the data compression and feature selection.

Fig. 6 shows an illustration of the data compression and feature selection process in the proposed framework.

IV. EXPERIMENTAL STUDY

Owing to the importance of roller bearings (Fig. 7) in rotating machines, two case studies of vibration signals generated by different health conditions in roller bearings have been used to evaluate the efficiency of the proposed framework. Based on the proposed framework, different scenarios of methods combinations have been investigated. In order to obtain compressively-sampled signals from the raw vibration signals, CS based on MMV model is designed using two different sparse representations based methods including thresholded WT and FFT. With regard to feature ranking and selection, five techniques have investigated including FS, LS, Relief-F, PCC, and Chi-2. In terms of fault classification, three classification algorithms have been tested; these include multinomial LRC, ANN, and SVM.

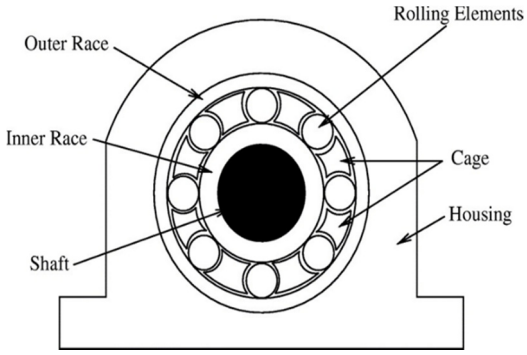

FIGURE 7. Typical roller bearing.

A. FIRST CASE STUDY

1) DATA DESCRIPTION

The vibration data used in this case study were collected from experiments on a small test rig. Six conditions of roller bearings status have been recorded and examined. These include, two normal conditions, namely, a brand new (NO), and a worn but undamaged condition (NW); and four fault conditions containing, inner race (IR) fault, an outer race (OR) fault, rolling element (RE) fault, and cage (CA) fault. As shown in Table 1, each condition has its corresponding unique characteristic.

TABLE 1. The characteristics of bearings health conditions in the first case study bearing dataset.

Condition	Characteristic
NO	The bearing is a brand new and in perfect condition.
NW	The bearing is in service for some period of time but in good condition.
IR	Inner race fault. This fault is created by cutting a small groove in the raceway of the inner race.
OR	Outer race fault. This fault is created by cutting a small groove in the raceway of the outer race.
RE	Roller element fault. This fault created by using electrical etcher to mark the surface of the balls, simulating corrosion.
CA	Cage fault. This fault is created by removing the plastic cage from one of the bearings, cutting away a section of the cage so that two of the balls were not held at a regular space and free to move.

Data were recorded at 16 different speeds. Fig. 8 depicts some typical time series plots for the six different aforementioned conditions. Depending on the fault conditions, the defects modulate the vibration signals with their own patterns. The inner and outer race fault conditions have a fairly periodic signal; the rolling element fault may or may not be periodic, dependent upon several reasons including

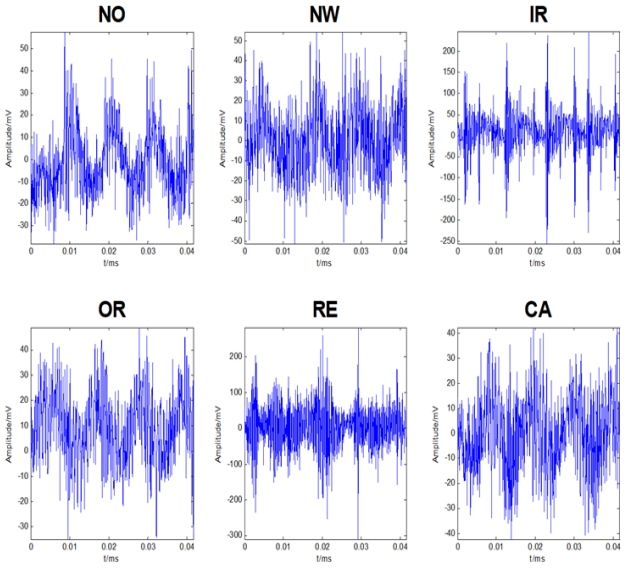

FIGURE 8. Typical time-domain vibration signals for the six different bearing health conditions.

the level of damage to the rolling element, the loading of the bearing, and also the track that the ball describes within the raceway itself. The cage fault generates a random distortion, which also depends on the degree of damage and the bearing loading.

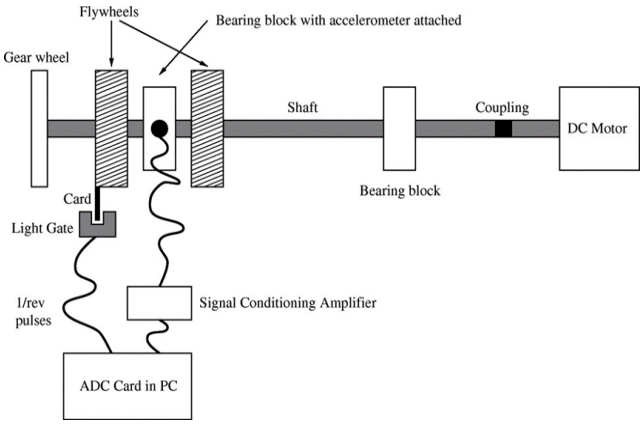

FIGURE 9. The test rig used to collect the vibration data of roller bearings.

Fig. 9 shows the test rig to collect the vibration data of bearings. The test rig consists of a DC motor driving the shaft through a flexible coupling, with the shaft supported by two Plummer bearing blocks. A series of damaged bearings were 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 cut-off 18 kHz), and sampled at 48 kHz, giving a slight oversampling. The machine was run at a series of 16 different speeds ranging between 25 and 75 rev/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.

2) RESULTS

To apply our proposed framework in this case study, fifty percent of the vibration data is randomly selected for training and the other 50% are used for testing the performance. To obtain compressively-sampled signals from the original vibration signals of roller bearings, MMV based CS model with two different sparse representations techniques, i.e., thresholded Haar WT and FFT are used. First, we used the Haar wavelet basis with five decomposition levels as sparsifying transform where the wavelet coefficients are thresholded using the penalized hard threshold to obtain wavelet coefficients based sparse representations of the original vibration signals. Second, we used the Fast Fourier Transform (FFT) to obtain the sparse components. Then we applied compressive sampling framework with different sampling rates (α) (0.1, 0.2, 0.3 and 0.4) with 600, 1200, 1800, and 2400 compressed measurements from our original vibration signals using a random matrix with i.i.d. Gaussian entries which satisfy the RIP.

To ensure that our CS model generates enough samples for the purpose of bearing fault classification, we used the generated compressively-sampled signals in each of the sparse representation methods to reconstruct the original signal X by applying the Compressive Sampling Matching Pursuit (CoSaMP) algorithm [54]. The reconstruction errors measured by Root Mean Squared Error (RMSE). For example, by using thresholded WT based CS with $\alpha = 0.1$ the RMSE for the six conditions of bearings are 8.5% (NO), 24.6% (NW), 15.23% (IR), 12.71% (OR), 11.87 % (RE), and 5.29% (CA); this has been studied in details in our previous work in [29]. While for FFT based CS using the same sampling rate $\alpha = 0.1$ the RMSE values are 4.8% (NO), 8.9% (NW), 6.3% (IR), 5.6% (OR), 4.7 % (RE), and 3.6% (CA), which indicate good signal reconstruction.

Based on the theory of CS these compressively-sampled signals possess the quality of the original signals. For further filtering, we applied FS, LS, Relief-F, PCC, and Chi-2 to select fewer features (k) from these compressively-sampled signals. Finally, with these selected features, we applied multinomial LRC, ANN, and SVM with ten-fold cross-validation to deal with the classification problem. The classification accuracy levels are obtained by averaging the results of twenty trials for each classifier and for each experiment. Table 2, Table 3, Table 4 present testing classification results for LRC, ANN, and SVM respectively using two different sparsifying transforms, i.e., FFT and WT, to obtain the compressively-sampled signals using the aforementioned compressive sampling rates.

It can be seen from the results in Table 2, Table 3, and Table 4 that among the various proposed combinations of CS with FFT, feature selection techniques, and classifiers, most of the combinations with LRC and ANN achieved better results than with SVM with "fitcecoc" function. In particular,

TABLE 2. Classification Accuracy of Roller Bearings Health conditions for LRC with different Combinations of MMV-CS and Feature Ranking and Selection Techniques (all classification accuracies ≥ 99% in bold).

Sparsifying method		FFT				WT			
Sampling rate (α)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Classifier: LRC									
Method	k								
CS-FS	60	94.3 ± 5.6	98.7 ± 0.7	98.8 ± 0.5	98.1 ± 0.9	92.2 ± 7.5	96.4 ± 3.9	97.2 ± 2.8	98.0 ± 2.2
	120	**99.7 ± 0.4**	**99.8 ± 0.3**	**99.8 ± 0.2**	**99.9 ± 0.4**	96.3 ± 3.8	96.2 ± 4.0	**99.2 ± 0.8**	**99.9 ± 0.1**
	180	**99.9 ± 0.1**	**99.9 ± 0.2**	**99.9 ± 0.1**	**100.0 ± 0.0**	98.9 ± 1.2	97.8 ± 2.2	**99.7 ± 0.3**	**99.9 ± 0.2**
CS-LS	60	95.8 ± 0.9	93.4 ± 1.2	95.2 ± 3.7	97.4 ± 2.8	91.9 ± 8.2	93.3 ± 6.9	95.1 ± 4.8	96.6 ± 3.5
	120	**99.5 ± 0.3**	**99.8 ± 0.6**	**99.4 ± 0.6**	**99.9 ± 0.1**	92.8 ± 7.2	94.2 ± 5.7	96.6 ± 3.4	98.0 ± 2.1
	180	**99.8 ± 0.2**	**99.9 ± 0.1**	**99.9 ± 0.1**	**99.9 ± 0.1**	95.4 ± 4.6	95.5 ± 4.4	96.5 ± 3.7	98.9 ± 1.1
CS-Relief-F	60	**99.5 ± 0.3**	**99.3 ± 0.6**	**99.5 ± 0.2**	**99.4 ± 0.3**	94.1 ± 5.1	93.7 ± 6.4	95.8 ± 3.4	96.6 ± 3.5
	120	**99.8 ± 0.3**	**99.9 ± 0.2**	**99.9 ± 0.1**	**99.9 ± 0.1**	96.4 ± 3.2	95.9 ± 4.2	98.3 ± 1.8	**99.9 ± 0.1**
	180	**99.9 ± 0.1**	**99.9 ± 0.1**	**100.0 ± 0.0**	**100.0 ± 0.0**	98.4 ± 1.6	**99.3 ± 0.7**	**100.0 ± 0.0**	**100 ± 0.0**
CS-PCC	60	98.5 ± 1.3	**99.3 ± 0.5**	98.8 ± 0.4	98.4 ± 0.9	94.1 ± 5.7	94.3 ± 4.6	97.4 ± 2.3	98.6 ± 1.5
	120	**99.8 ± 0.3**	**99.5 ± 0.2**	**99.8 ± 0.2**	**99.9 ± 0.1**	95.0 ± 4.9	95.0 ± 4.9	98.1 ± 1.7	98.9 ± 1.3
	180	**99.9 ± 0.1**	**99.9 ± 0.2**	**99.9 ± 0.2**	**99.9 ± 0.1**	98.8 ± 1.2	**99.0 ± 1.1**	98.8 ± 1.2	**99.4 ± 0.6**
CS-Chi-2	60	98.0 ± 2.2	98.7 ± 1.4	98.8 ± 1.2	98.8 ± 1.2	95.1 ± 4.5	96.7 ± 2.7	98.3 ± 2.8	98.9 ± 1.2
	120	**99.5 ± 0.5**	**99.5 ± 0.4**	**99.5 ± 0.3**	**99.5 ± 0.5**	96.8 ± 2.9	97.9 ± 3.2	96.6 ± 3.4	**99.9 ± 0.1**
	180	**99.9 ± 0.1**	**99.9 ± 0.1**	**100.0 ± 0.0**	**99.9 ± 0.1**	**99.5 ± 0.4**	**99.3 ± 0.7**	**100.0 ± 0.0**	**100.0 ± 0.0**

TABLE 3. Classification Accuracy of Roller Bearings Health conditions for ANN with different Combinations of MMV-CS and Feature Ranking and Selection Techniques (all classification accuracies ≥ 99% in bold).

Sparsifying method		FFT				WT			
Sampling rate (α)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Classifier: ANN									
Method	k								
CS-FS	60	93.8 ± 6.3	98.3 ± 0.8	**99.0 ± 1.1**	98.9 ± 0.4	91.6 ± 8.3	97.1 ± 2.8	98.9 ± 1.2	97.7 ± 2.3
	120	**99.2 ± 1.3**	**99.7 ± 0.7**	**99.7 ± 0.6**	**99.6 ± 0.9**	97.2 ± 2.8	98.3 ± 1.7	**99.0 ± 0.9**	**99.3 ± 0.7**
	180	**99.8 ± 0.5**	**99.8 ± 0.3**	**99.7 ± 0.3**	**100.0 ± 0.0**	98.2 ± 1.9	**99.1 ± 0.9**	**99.3 ± 0.8**	**99.7 ± 0.3**
CS-LS	60	98.6 ± 0.9	97.3 ± 2.1	97.9 ± 2.2	95.8 ± 3.3	92.9 ± 7.0	93.7 ± 6.3	94.4 ± 5.5	97.1 ± 2.7
	120	**99.9± 0.2**	**99.4 ± 0.7**	**99.9 ± 0.1**	**99.9 ± 0.1**	95.3 ± 4.8	95.2 ± 4.8	96.9 ± 3.2	98.3 ± 1.7
	180	**99.9 ± 0.1**	**99.9 ± 0.2**	**99.9 ± 0.1**	**99.9 ± 0.1**	96.8 ± 2.9	95.9 ± 4.1	97.8 ± 2.1	98.9 ± 1.0
CS-Relief-F	60	**99.7 ± 0.6**	**99.7 ± 0.3**	**99.2 ± 0.9**	**99.6± 0.3**	94.6 ± 5.4	94.4 ± 5.7	93.2 ± 6.7	97.1 ± 2.7
	120	**99.9 ± 0.2**	**99.7 ± 0.4**	**99.6 ± 0.9**	**99.7 ± 0.4**	97.2 ± 2.8	97.0 ± 3.1	**99.5 ± 0.5**	**99.8 ± 0.2**
	180	**99.9± 0.3**	**100 ± 0.0**	**100.0 ± 0.0**	**99.9 ± 0.3**	**99.1 ± 0.9**	**99.9 ± 0.1**	**100.0 ± 0.0**	**100.0 ± 0.0**
CS-PCC	60	**99.4 ± 0.3**	**99.4 ± 0.6**	**99.2 ± 0.9**	**99.5 ± 0.5**	94.9 ± 5.3	95.7 ± 4.3	98.5 ± 1.6	94.4 ± 5.5
	120	**99.8 ± 0.3**	**99.7 ± 0.4**	**99.9 ± 0.3**	**99.9 ± 0.1**	96.4 ± 3.6	96.3 ± 3.8	**99.0 ± 0.9**	98.5 ± 1.4
	180	**99.7± 0.5**	**100 ± 0.0**	**100 ± 0.0**	**99.9 ± 0.2**	**99.3 ± 0.7**	**99.6 ± 0.3**	**99.3 ± 0.7**	**99.9 ± 0.1**
CS-Chi-2	60	96.9± 2.0	95.2± 3.9	97.8 ± 0.9	98.3 ± 0.9	95.8 ± 3.9	96.3 ± 3.7	98.5 ± 2.7	98.9 ± 1.2
	120	**99.2 ± 0.9**	**99.4 ± 0.8**	**99.0 ± 1.1**	**99.3± 0.8**	98.8 ± 0.4	98.1 ± 2.9	**99.5 ± 0.5**	**99.9 ± 0.1**
	180	**99.9± 0.3**	**99.9 ± 0.3**	**99.6 ± 0.6**	**99.9 ± 0.2**	**99.2 ± 0.7**	**99.9 ± 0.1**	**100.0 ± 0.0**	**100.0 ± 0.0**

results from CS-Chi-2 and CS-Relief-F for all values of the sampling rate (α) and the number of selected features (k) with both LRC and ANN are better than with SVM. Also, all the combinations of CS with FFT and the considered feature selection techniques with LRC and ANN achieved high classification accuracies (all above 99%) for all values of α with $k = 120$. Also, CS-FS, CS-LS, and CS-PCC with LRC and ANN achieved better results than with SVM for all values of α with k = 60 and 120. Moreover, all classification accuracies are above 99% for all the classifiers considered with CS-FS, CS-Relief-F, CS-PCC, and CS-Chi-2 with both WT and FFT sparse representations techniques using $\alpha = 0.4$ and k = 180. For CS-LS all considered classifiers achieved accuracy results above 99% using $\alpha = 0.4$ and k = 180 with FFT only.

SVM achieved good results in several scenarios with the larger number of selected features, i.e., k = 180: (1) using CS-FS with FFT, $\alpha = 0.2$, 0.3, and 0.4, and with thresholded WT and $\alpha = 0.4$, (2) using CS-LS with FFT and all values of α, (3) using CS-Relief-F with WT and $\alpha = 0.3$ and 0.4, (4) CS-PCC with FFT, and with WT for $\alpha = 0.4$, (5) using CS-Chi-2 with WT and all values of α.

Generally, the classification accuracies of all the proposed methods, i.e., CS-FS, CS-LS, CS-Relief-F, CS-PCC, and

TABLE 4. Classification Accuracy of Roller Bearings Health conditions for SVM with different Combinations of MMV-CS and Feature Ranking and Selection Techniques (all classification accuracies ≥ 99% in bold).

Sparsifying method		FFT				WT			
Sampling rate (α)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Classifier: SVM Method	k								
CS-FS	60	85.8 ± 8.4	89.5 ± 5.1	74.8 ± 5.6	69.9 ± 2.6	91.5 ± 8.4	95.2 ± 4.7	95.9 ± 3.7	95.7 ± 3.9
	120	89.6 ± 11.4	92.3 ± 6.2	98.4 ± 1.6	97.9 ± 3.9	94.9 ± 4.9	94.8 ± 5.2	97.5 ± 2.3	97.9 ± 2.4
	180	97.4 ± 5.2	**99.1 ± 0.9**	**99.3 ±0.7**	**99.8 ± 1.2**	95.8 ± 3.9	96.3 ± 3.8	98.8 ± 1.4	**99.4 ± 0.7**
CS-LS	60	92.7 ± 7.0	92.1 ± 5.1	92.0 ± 0.7	95.3 ± 4.7	92.6 ± 7.4	92.2 ± 7.7	91.9 ± 8.3	92.7 ± 6.9
	120	98.7 ± 1.4	98.9 ± 1.1	**99.3 ± 0.8**	**99.7 ± 0.3**	91.4 ± 6.9	93.5 ± 6.4	94.7 ± 5.2	95.7 ± 4.2
	180	**99.2 ± 0.8**	**99.2 ± 0.7**	**99.2 ± 0.9**	**99.9 ± 0.1**	94.4 ± 5.4	93.2 ± 6.9	95.3 ± 4.6	97.8 ± 2.3
CS-Relief-F	60	85.8 ± 13.7	77.9 ± 1.5	68.4 ± 16.1	67.8 ± 16.4	94.7 ± 5.3	94.2 ± 5.9	92.7 ± 7.4	94.2 ± 4.8
	120	89.3 ± 8.4	90.9 ± 7.8	83.9 ± 12.2	78.5 ± 9.7	95.3 ± 4.7	95.5 ± 4.4	97.5 ± 2.7	98.8 ± 1.3
	180	95.6 ± 5.8	96.2 ± 4.8	88.0 ± 9.4	96.8 ± 4.5	97.7 ± 2.1	98.8 ± 1.2	**99.3 ± 0.6**	**99.0 ± 1.2**
CS-PCC	60	78.3 ± 14.0	73.2 ± 7.4	69.4 ± 13.1	65.1 ± 15.4	71.5 ± 17.8	79.9 ± 17.8	85.8 ± 13.2	94.7 ± 5.3
	120	93.2 ± 7.5	93.8 ± 7.9	93.3 ± 8.5	79.9 ± 2.2	78.2 ± 16.2	82.2 ± 15.8	88.2 ± 10.9	92.4 ± 7.5
	180	97.7 ± 5.1	96.7 ± 5.7	97.8 ± 2.8	**99.3 ± 0.7**	83.5 ± 13.7	87.4 ± 11.6	92.7 ± 7.6	**99.3 ± 0.7**
CS-Chi-2	60	64.2 ± 7.1	68.3± 7.6	63.7± 9.3	69.1 ± 18.4	94.3 ± 3.8	95.2 ± 3.2	97.7 ± 3.4	98.2 ± 2.8
	120	83.2 ± 9.0	76.9 ± 9.3	74.8 ± 11.9	71.4 ± 7.1	96.2 ± 3.7	97.1 ± 3.4	98.5 ± 1.7	98.8 ± 1.2
	180	95.2 ± 5.2	94.9 ± 6.3	90.7 ± 7.0	82.3 ± 14.4	**99.6 ± 0.3**	**99.6 ± 0.2**	**99.8 ± 0.2**	**99.9 ± 0.2**

CS-Chi-2 are based on the compressed sampling rate (α), the feature selection method, and the number of the selected features (k). However, the features are selected from random compressed projections of length (m) that do not include representations of all the attributes in the original data of length (n), i.e., m $\neq$ n. Therefore, the assumption that when α gets larger the accuracy gets better may not apply in every set of selected features. For example, it can be clearly seen from the results in Table 2 that the classification accuracy of all the proposed methods with FFT sparsifying method becomes better when α becomes larger with k = 180. While with k = 120, only one variation was found using CS-PCC with accuracy 99.8% and 99.5% for α equal to 0.1 and 0.2 respectively. With k = 60, five variations were found; one variation using CS-FS with accuracy 98.8% and 98.1% for α equal to 0.3 and 0.4 respectively; one variation using CS-LS with accuracy 95.8% and 93.4% for α equal to 0.1 and 0.2 respectively; two variations using CS-Relief-F, first with accuracy 99.5%, 99.3% for α equal to 0.1 and 0.2, and second with 99.5% and 99.4% accuracy for 0.3, and 0.4 respectively; one variation using CS-PCC with 98.8% and 98.4% for 0.3, and 0.4 respectively.

Taken together, these results show that the proposed framework with various methods studied here has the ability to classify bearing health conditions with a high classification accuracy with the following comments:

1) FFT as a sparsifying transform method for our proposed MMV based CS model can achieve better results than thresholded WT.
2) LRC and ANN have the ability to achieve high classification accuracy with different values of the sampling rate (α) and a number of selected features (k) for all the considered CS and feature selection techniques combinations.
3) SVM has the ability to achieve good classification accuracy with the larger number of selected features, i.e., k = 180, and larger values of α, e.g., $\alpha = 0.4$, for certain combinations. This can be clearly seen in Table 3 and previously published results in [55] and [56].
4) With the larger number of selected features, all the proposed methods achieved high classification accuracy. Thus, for the application of the proposed methods in fault diagnosis, we recommend selecting a larger number of features from compressively-sampled signals.

3) COMPARISONS OF RESULTS

For further verification of the efficiency of the proposed framework, complete comparison results of the classification accuracy using the different combinations based on the proposed framework compared with some recently published results using the same vibration dataset, for instance in [27] results reported for three methods, one method uses all the original vibration data from which entropic features are extracted, and the other two uses compressed measurements to recover the original vibration signals and from the recovered signals entropic features are extracted. With the extracted features SVM used to classify bearing health conditions. These results reported in [27].

Moreover, three CS based techniques have been used to classify bearing health conditions, using the compressed measurements as the input to LRC classifier, combining CS and PCA, CS and LDA to extract features from raw vibration data and then use LRC to deal with the classification problem. These results are reported in [33]. Also, a hybrid model consisting of the Fuzzy Min-Max (FMM) neural network and Random Forest (RF) with Sample Entropy (SampEn) and Power Spectrum (PS) features is used to classify bearing

TABLE 5. A comparison with the classification results from literature on bearing dataset.

		Classification accuracy (%)
Raw Vibration with entropic features [27]		98.9 ± 1.2
Compressed sampled with α = 0.5 followed by signals reconstruction [27]		92.4 ± 0.5
Compressed sampled with α = 0.25 followed by signals reconstruction [27]		84.6 ± 3.4
CS [33]	α = 0.1	98.6± 0.3
CS-PCA [33]	α = 0.1	98.5± 0.4
CS-LDA [33]	α = 0.1	89.8± 3.5
FMM-RF (SampEn + PS) [57]		99.81± 0.41
GP generated feature sets (un-normalised) [58]		
	ANN	96. 5
	SVM	97.1
Our proposed framework		
With FFT, α = 0.1, k = 120		
CS-FS	LRC	99.7 ± 0.4
	ANN	99.2 ± 1.3
CS-LS	LRC	99.5 ± 0.3
	ANN	99.9± 0.2
CS-Relief-F	LRC	99.8 ± 0.3
	ANN	99.9 ± 0.2
CS-PCC	LRC	99.8 ± 0.3
	ANN	99.8 ± 0.3
CS-Chi-2	LRC	99.5 ± 0.5
	ANN	99.2 ± 0.9

health conditions and the results reported in [57]. In [58] a Genetic Programming (GP) based approach is proposed for feature extraction from raw vibration data and with extracted features SVM and ANN are used to classify bearing health conditions. Table 5 presents classification results of bearing health conditions using our proposed methods with $\alpha = 0.4$, 0.3, and 0.1 and the reported results in [27], [33], [57], and [58] using the same dataset used in this case study.

It can be clearly seen that the classification results of our proposed methods are better than those reported in [27], [33], and [58]. Also, our results are as good as, if not better than results reported in [57] although we are using only 10% ($\alpha = 0.1$) of the original vibration data that is not matched by the method in [57] using all the raw vibration data.

This section has validated the proposed framework and has shown that the many combinations of CS and feature ranking methods achieved high classification accuracies of bearing faults. The next section of this paper will validate the usage of our proposed framework using publicly available bearing vibration dataset. The advantage of the shared dataset is that we can compare the results of other researchers easily.

B. SECOND CASE STUDY

The bearing vibration data used in this case study is provided by the Case Western Reserve University (CWRU) Bearing Data Center [59]. This data is freely available and commonly used in roller bearings fault diagnosis field. Fig. 10 shows the test rig that is used to acquire this vibration data. It is comprised of a 2 horsepower electric motor driving a shaft that contains a torque transducer and encoder. A dynamometer and electronic control system are used to apply torque to the shaft. A series of faults with width ranging from 0.18 to 0.71 mm (0.007 to 0.028 in) were seeded on the drive end bearing (in this case SKF deep-groove ball bearings 6205-2RS JEM were used) of the electric motor utilising electro-discharger machining.

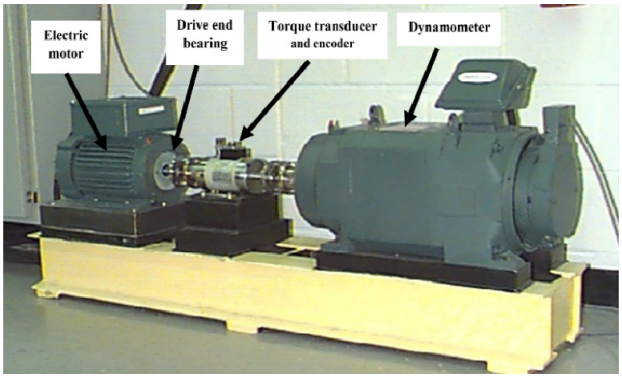

FIGURE 10. The test rig used to collect the first vibration data of bearings (CWRU Bearing Data Center).

The seeded faults include rolling elements, inner race, and outer race faults, and each faulty bearing was run for motor loads 0 – 3 horsepower at a constant speed in the range 1730 – 1797 rev/m, and the sampling rates used were 12 kHz. The bearing vibration signals were acquired under normal NO, IR, OR, and RE conditions for four different speeds. In this case study, bearing dataset is chosen from the data files of vibration signals that were sampled at 12 kHz with fault size (0.18, 0.36, 0.53, and 0.71), load 2-horsepower, and the number of examples chosen is 60 per condition. This gave a dataset with 720 total number of examples and 2000 data points for each signal. The description of the used bearing vibration dataset is presented in Table 6.

TABLE 6. Description of the bearing health conditions of the used bearing vibration dataset.

Health condition	Faults width (mm)	Classification label
NO	0	1
RE1	0.18	2
RE2	0.36	3
RE3	0.53	4
RE4	0.71	5
IR1	0.18	6
IR2	0.36	7
IR3	0.53	8
IR4	0.71	9
OR1	0.18	10
OR2	0.36	11
OR3	0.53	12

Of these raw vibration signals, 240 examples are randomly selected for training and 480 examples are used for testing. We applied the MMV-CS model with FFT to obtain compressively-sampled signals from the raw vibration signals using α equal to 0.1 and the same feature selections methods as in the first case study to select fewer features of these compressively-sampled signals. With these fewer selected features, we employed LRC, ANN, and SVM to deal with the classification problem. The classification accuracies are achieved by averaging the results of twenty trials for each classifier and for each experiment. Table 7 shows the accuracy of all experiments with a different number of selected features (k = 25 and 50). Fig. 11 and Fig. 12 show a column chart representations of the classification results presented in Table 7.

TABLE 7. Classification Accuracy of Roller Bearings Health conditions of the second case study with different Combinations of MMV-CS and Feature Ranking and Selection Techniques (all classification accuracies ≥ 99% in bold).

Method	LRC	ANN	SVM
CS-FS			
(k = 25)	98.4 ± 1.6	**99.6 ± 0.5**	97.4 ± 2.7
(k = 50)	**99.9 ± 0.1**	**100 ± 0.0**	**99.9 ± 0.1**
CS-LS			
(k = 25)	**99.1 ± 0.8**	**99.2 ± 0.8**	98.5 ± 1.6
(k = 50)	97.5 ± 2.6	**99.4 ± 0.7**	98.3 ± 1.7
CS-Relief-F			
(k = 25)	**99.3 ± 0.6**	**99.2 ± 0.8**	97.8 ± 2.4
(k = 50)	**99.4 ± 0.5**	**100 ± 0.0**	**99.5 ± 0.5**
CS-PCC			
(k = 25)	**99.2 ± 0.8**	**99.5 ± 0.6**	97.9 ± 1.9
(k = 50)	**99.3 ± 0.7**	**99.9 ± 0.1**	97.5 ± 2.4
CS-Chi-2			
(k = 25)	97.5 ± 2.6	**99.9 ± 0.1**	94.7 ± 5.4
(k = 50)	**99.1 ± 0.9**	**100 ± 0.0**	**99.9 ± 0.1**

As follows from the Table 7, all combinations of CS and feature ranking and selection techniques with ANN and a different number of selected features, i.e., k = 25 and 50 achieved high classification accuracies (all over 99%). As well as achieving high classification accuracy with ANN using 25 features, all combinations with ANN and 50 selected features are able to achieve even higher classification accuracy. In particular, results of CS-FS, CS-Relief-F, and CS-Chi-2 with ANN and 50 selected features achieved 100% classification accuracy for every single run in our investigations.

The average classification accuracies of LRC obtained using 50 selected features using CS-FS, CS-Relief-F, CS-PCC, and CS-Chi-2 are above 99%. Also, with 25 selected features based on CS-LS, CS-Relief-F, and CS-PCC, LRC is able to achieve over 99% classification accuracy. In addition, with SVM and 50 selected features based on CS-FS, CS-Relief-F, and CS-Chi-2, the average

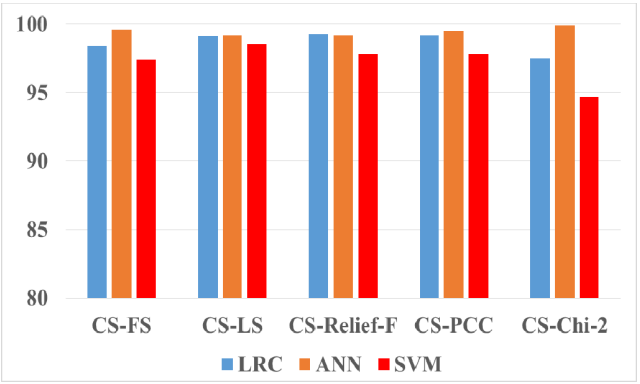

FIGURE 11. Classification accuracy rates of 25 selected features.

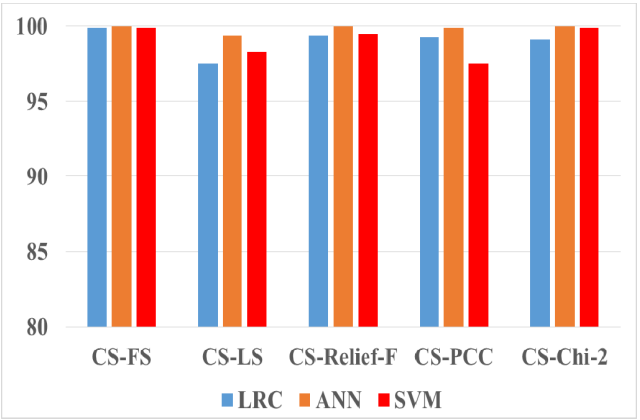

FIGURE 12. Classification accuracy rates of 50 selected features.

classification accuracy rates are generally above 99%. These observations can be clearly seen in Fig. 11 and 12 below. However, from Table 7 it can be clearly seen that for k = 25, the classification accuracies of CS-LS-LRC, CS-LS-SVM, and CS-PCC-SVM methods are 99.9%, 98.5%, and 97.9% respectively; while for k=50, the classification accuracies for the same methods are 97.5%, 98.3%, and 97.5% respectively. Therefore, for a fixed compressed signal size m, there is an optimal number of features k that makes the classification accuracy higher than other classification accuracies achieved using a different number of features that may be bigger or smaller than k.

For further evaluation of the efficiency of the proposed MMV-CS and feature ranking analysis-based framework. Table 8 presents the comparisons with some recently published results [24] with the same bearing dataset used in this case study. One method uses Feature Selection by Adjunct Rand Index and Standard Deviation Ratio (FSAR) to select features from the original feature set (OFS). Other methods, use PCA, LDA, Local Fisher Discriminant Analysis (LFDA), and Support Margin LFDA (SM-LFDA) to reduce the dimension of selected features using FSAR. With the selected features, SVM is used for the purpose of classification. The results for different numbers of selected features are reported

TABLE 8. A comparison with the classification results from literature on bearing dataset.

		Classification accuracy (%)
OFS-FSAR-SVM [24]		
(25 selected features)		91.46
(50 selected features)		69.58
OFS-FSAR-PCA-SVM [24]		
(25 selected features)		91.67
(50 selected features)		69.79
OFS-FSAR-LDA-SVM [24]		
(25 selected features)		86.25
(50 selected features)		92.7
OFS-FSAR-LFDA-SVM [24]		
(25 selected features)		93.75
(50 selected features)		94.38
OFS-FSAR-(SM-LFDA)-SVM [24]		
(25 selected features)		94.58
(50 selected features)		95.63
Our proposed framework		
With FFT, $\alpha = 0.1$, $k = 25$		
CS-FS	LRC	98.4 ± 1.6
	ANN	99.6 ± 0.5
	SVM	97.4 ± 2.7
CS-LS	LRC	99.1 ± 0.8
	ANN	99.2 ± 0.8
	SVM	98.5 ± 1.6
CS-Relief-F	LRC	99.3 ± 0.6
	ANN	99.2 ± 0.8
	SVM	97.8 ± 2.4
CS-PCC	LRC	99.2 ± 0.8
	ANN	99.5 ± 0.6
	SVM	97.9 ± 1.9
CS-Chi-2	LRC	97.5 ± 2.6
	ANN	99.9 ± 0.1
	SVM	94.7 ± 5.4

in [24]. It is clear that all the results from our proposed framework outperforms results reported in [24].

V. CONCLUSION

An original framework based on CS using MMV and feature ranking has been proposed for bearing fault classification. In this framework, CS based on MMV is used to reduce a large amount of bearing vibration signals by producing compressively-sampled signals from the raw vibration signals. For further filtering, a feature ranking technique is used to rank the features of the compressively-sampled signals from which the most important features are selected. With these selected features, a classification algorithm has been used to classify bearing faults. We have investigated different combinations of MMV based CS and feature ranking techniques (CS-FS, CS-LS, CS-Relief-F, CS-PCC, and CS-Chi-2) with two different bearing fault classification tasks. Three classification algorithms (LRC, ANN, and SVM) have been tested to evaluate the proposed framework for the classification of bearing faults.

The CS and feature ranking framework is able to achieve high classification accuracy in all of the faults studied here. Moreover, the various combinations of CS and feature ranking techniques investigated in this study offer higher classification accuracies with LRC, ANN, and SVM even with a limited amount of data compared to recently published results.

REFERENCES

[1] P. J. Tavner, L. Ran, J. Penman, and H. Sedding, *Condition Monitoring of Rotating Electrical Machines*. Stevenage, U.K.: IET, 2008.

[2] A. K. Nandi, C. Liu, and M. L. D. Wong, "Intelligent vibration signal processing for condition monitoring," in *Proc. Int. Conf. Surveill.*, vol. 7, 2013, pp. 1–15.

[3] C. Mechefske, "Machine condition monitoring and fault diagnostics," in *Vibration and Shock Handbook*. Boca Raton, FL, USA: CRC Press, 2005.

[4] M. P. Norton and D. G. Karczub, *Fundamentals of Noise and Vibration Analysis for Engineers*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[5] 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.

[6] Y. Wang, J. Xiang, R. Markert, and M. Liang, "Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications," *Mech. Syst. Signal Process.*, vols. 66–67, pp. 679–698, Jan. 2016.

[7] V. C. M. N. Leite *et al.*, "Detection of localized bearing faults in induction machines by spectral kurtosis and envelope analysis of stator current," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1855–1865, Mar. 2015.

[8] L. Saidi, J. Ben Ali, and F. Fnaiech, "The use of spectral kurtosis as a trend parameter for bearing faults diagnosis," in *Proc. 15th Int. Conf. Sci. Techn. Autom. Control Comput. Eng. (STA)*, Hammamet, Tunisia, 2014, pp. 394–399.

[9] 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.

[10] 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.

[11] 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. IEEE 13th Int. Conf. Ind. Inform. (INDIN)*, Cambridge, U.K., Jul. 2015, pp. 771–776.

[12] A. Widodo, B.-S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 299–312, 2007.

[13] Y. Chang and W. Jiao, "ICA-ANN method in fault diagnosis of rotating machinery," in *Proc. IEEE Int. Conf. Comput. Sci. Automat. Eng. (CSAE)*, Zhangjiajie, China, 2012, pp. 236–240.

[14] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Effects of deep neural network parameters on classification of bearing faults," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc., (IECON)*, Florence, Italy, Oct. 2016, pp. 6329–6334.

[15] M. Van and H. J. Kang, "Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection," *IET Sci., Meas. Technol.*, vol. 9, no. 6, pp. 671–680, 2015.

[16] L. B. Jack and A. K. Nandi, "Genetic algorithms for feature selection in machine condition monitoring with vibration signals," *IEE Proc.-Vis., Image Signal Process.*, vol. 147, no. 3, pp. 205–212, Jun. 2000.

[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, Y. Zang, 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, Jun. 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] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14347–14357, 2017.

[24] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, "Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection," *IEEE Access*, vol. 6, pp. 3715–3730, 2017.

[25] B. R. Nayana and P. Geethanjali, "Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5618–5625, Sep. 2017.

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

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

[28] Z. Xinpeng, H. Niaoqing, and C. Zhe, "A bearing fault detection method base on compressed sensing," in *Engineering Asset Management—Systems, Professional Practices and Certification*. Cham, Switzerland: Springer, 2015, pp. 789–798.

[29] G. Tang, W. Hou, H. Wang, G. Luo, and J. Ma, "Compressive sensing of roller bearing faults via harmonic detection from under-sampled vibration signals," *Sensors*, vol. 15, no. 10, pp. 25648–25662, 2015.

[30] 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.

[31] G. Tang, Q. Yang, H.-Q. Wang, G.-G. Luo, and J.-W. Ma, "Sparse classification of rotating machinery faults based on compressive sensing strategy," *Mechatronics*, vol. 31, pp. 60–67, Oct. 2015.

[32] 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, Jan. 2018.

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

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

[35] 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.

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

[37] J. Tang, S. Alelyani, H. Liu, and C. Aggarwal, "Feature selection for classification: A review," *Data Classification: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2014.

[38] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[39] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," in *Proc. 27th Conf. Annu. Conf. Uncertainty Artif. Intell.*, 2011, pp. 266–273.

[40] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. NIPS*, vol. 186, 2005, pp. 507–514.

[41] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. New York, NY, USA: Chapman & Hall, 2007.

[42] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proc. ICML*, 1997, pp. 412–420.

[43] Mathworks.com. (2017). *Cross-Tabulation—MATLAB Crosstab*. [Online]. Available: https://uk.mathworks.com/help/stats/crosstab.html

[44] D. W. Hosmer, Jr., and S. Lemeshow, *Applied Logistic Regression*. Hoboken, NJ, USA: Wiley, 2004.

[45] L. B. Jack and A. K. Nandi, "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mech. Syst. Signal Process.*, vol. 16, nos. 2–3, pp. 373–390, 2002.

[46] L. B. Jack, A. K. Nandi, and A. C. McCormick, "Diagnosis of rolling element bearing faults using radial basis function networks," *EURASIP J. Appl. Sig. Process.*, vol. 6, no. 1, pp. 25–32, 1999.

[47] A. C. McCormick and A. K. Nandi, "Classification of the rotating machine condition using artificial neural networks," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vol. 211, no. C6, pp. 439–450, 1997.

[48] A. C. McCormick and A. K. Nandi, "Real-time classification of rotating shaft loading conditions using artificial neural networks," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 748–757, May 1997.

[49] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, Nov. 1993.

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

[51] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *Analyst*, vol. 135, no. 2, pp. 230–267, 2010.

[52] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

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

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

[55] H. O. A. Ahmed and A. K. Nandi, "Multiple measurement vector compressive sampling and Fisher score feature selection for fault classification of roller bearings," in *Proc. 22nd Int. Conf. Digit. Signal Process. (DSP)*, London, U.K., 2017, pp. 1–5.

[56] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Classification of bearing faults combining compressive sampling, Laplacian score, and support vector machine," in *Proc. 43rd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Beijing, China, Oct./Nov. 2017, pp. 8053–8058.

[57] M. Seera, M. L. D. Wong, and A. K. Nandi, "Classification of ball bearing faults using a hybrid intelligent model," *Appl. Soft Comput.*, vol. 57, pp. 427–435, 2017.

[58] 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.

[59] *Case Western Reserve University Bearing Data Center*. [Online]. Available: http://csegroup.case.edu/bearingdatacenter/home

HOSAMELDIN AHMED received the B.Sc. degree (Hons.) in engineering technology, specialization in electronic engineering, from the Faculty of Science and Technology, University of Gezira, Sudan, in 2000, and the M.Sc. degree in computer engineering and networks from the University of Gezira, in 2010. He is currently pursuing the Ph.D. degree in electronic and computer engineering with Brunel University London, U.K. Since 2014, he has been with Prof. A.. 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 a journal paper and several high quality international conferences. His research interests lie in the areas of 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. He held academic positions at several universities, including Oxford, Imperial College London, the University of Strathclyde, and Liverpool as well as the Finland Distinguished Professorship in Jyväskylä, Finland. In 2013, he moved to Brunel University London, to become the Chair and the Head of Electronic and Computer Engineering. He is currently a Visiting Professor with Tongji University, China, and an Adjunct Professor with the University of Calgary, Canada.

In 1983, he co-discovered the three fundamental particles known as W+, W-, and Z0 (by the UA1 team at CERN), providing the evidence for the unification of the electromagnetic and weak forces, for which the Nobel Committee for Physics in 1984 awarded the prize to his two team leaders for their decisive contributions. He has made many fundamental theoretical and algorithmic contributions to many aspects of signal processing and machine learning. He has much expertise in big data, dealing with heterogeneous data, and extracting information from multiple data sets obtained in different laboratories and different times. He has authored over 550 technical publications, including 220 journal papers and four books: *Automatic Modulation Classification: Principles, Algorithms and Applications* (Wiley, 2015); *Integrative Cluster Analysis in Bioinformatics* (Wiley, 2015), *Blind Estimation Using Higher-Order Statistics* (Springer, 1999), and *Automatic Modulation Recognition of Communications Signals* (Springer, 1996). His current research interests lie in the areas of signal processing and machine learning, with applications to communications, gene expression data, functional magnetic resonance data, machine condition monitoring, and biomedical data. The h-index of his publications is 67 (Google Scholar) and ERDOS number is 2.

Dr. Nandi is a fellow of the Royal Academy of Engineering, U.K., and of seven other institutions. Among the many awards, he received the Mountbatten Premium of the Institution of Electrical Engineers, U.K., in 1998, the Water Arbitration Prize of the Institution of Mechanical Engineers, U.K., in 1999, the Glory of Bengal Award for his outstanding achievements in scientific research in 2010, and the Institute of Electrical and Electronics Engineers (USA) Heinrich Hertz Award in 2012. He is an IEEE Distinguished Lecturer (2018–2019).