Article

Intrinsic Dimension Estimation-Based Feature Selection and Multinomial Logistic Regression for Classification of Bearing Faults Using Compressively Sampled Vibration Signals

Hosameldin O. A. Ahmed [1] **and Asoke K. Nandi** [2,3,*]

1 Department of Mechanical and Aerospace Engineering, Brunel University London, Uxbridge UB8 3PH, UK; hosameldin.ahmed3@brunel.ac.uk
2 Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, UK
3 School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
* Correspondence: asoke.nandi@brunel.ac.uk

Abstract: As failures of rolling bearings lead to major failures in rotating machines, recent vibration-based rolling bearing fault diagnosis techniques are focused on obtaining useful fault features from the huge collection of raw data. However, too many features reduce the classification accuracy and increase the computation time. This paper proposes an effective feature selection technique based on intrinsic dimension estimation of compressively sampled vibration signals. First, compressive sampling (CS) is used to get compressed measurements from the collected raw vibration signals. Then, a global dimension estimator, the geodesic minimal spanning tree (GMST), is employed to compute the minimal number of features needed to represent efficiently the compressively sampled signals. Finally, a feature selection process, combining the stochastic proximity embedding (SPE) and the neighbourhood component analysis (NCA), is used to select fewer features for bearing fault diagnosis. With regression analysis-based predictive modelling technique and the multinomial logistic regression (MLR) classifier, the selected features are assessed in two case studies of rolling bearings vibration signals under different working loads. The experimental results demonstrate that the proposed method can successfully select fewer features, with which the MLR-based trained model achieves high classification accuracy and significantly reduced computation times compared to published research.

Keywords: vibration-based condition monitoring; rolling bearing fault diagnosis; compressive sampling (CS); feature selection; multinomial logistic regression (MLR)

1. Introduction

Rolling bearings are critical components of the entire system of rotating machines and play a crucial role in retaining motion between motionless and moving parts. Failure of rolling bearing is one of the key problems in rotating machines that may lead to major catastrophes in machines [1]. It has previously been observed that approximately 40–90% of rotating machines failures are related to bearing faults [2]. Therefore, in most manufacturing procedures, rolling bearings need to be monitored to avoid machine failures. Numerous techniques can be used for machine condition monitoring such as vibration monitoring, electric motor current monitoring, acoustic emission monitoring, etc. Of these, vibration-based condition monitoring has been extensively utilized and has become a widely approved method [3]. As presented in Figure 1, a roller bearing is comprised of some components containing the inner race, the outer race, the rolling elements, and the cage [4]. Bearings faults can happen for several reasons such as fatigue, incorrect lubrication, contamination, corrosion, etc. [5]. In practice, faults in rolling bearings produce a series of impulses that repeat periodically at a rate named the bearing fundamental defect

frequency (BFDF), which usually depends on the site of the faults, the geometry of the bearing, and the shaft speed, as displayed in Figure 2 [6].
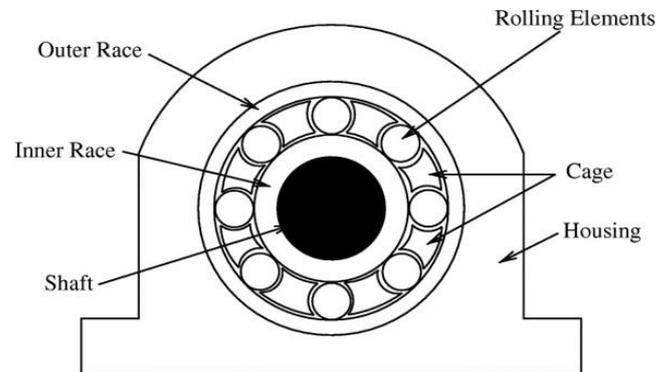
Figure 1. A typical roller bearing [4].
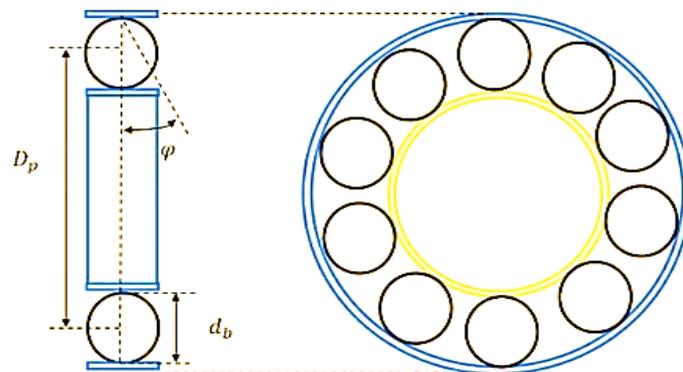
Figure 2. Rolling element bearing geometry [6].

Based on the damaged component, the BFDFs are categorized into four groups: (i) bearing pass frequency of the inner race (BPFI), (ii) bearing pass frequency of the outer race (BPFO), (iii) ball spin frequency (BSF), and (iv) fundamental train frequency (FTF), which are connected to the defect at the outer race, the inner race, the rolling element, and the cage [7]. These frequencies could be described using the following equations:

$$\text{BPFI} = \frac{N_b S_{sh}}{2}\left(1 + \frac{d_b}{D_p}cos\varphi\right) \tag{1}$$

$$\text{BPFO} = \frac{N_b S_{sh}}{2}\left(1 - \frac{d_b}{D_p}cos\varphi\right) \tag{2}$$

$$\text{BSF} = \frac{D_p}{2d_b}\left(1 - \left(\frac{d_b}{D_p}cos\varphi\right)^2\right) \tag{3}$$

$$\text{FTF} = \frac{S_{sh}}{2}\left(1 - \frac{d_b}{D_p}cos\varphi\right) \tag{4}$$

Here, N_b represents the number of rolling elements, S_{sh} represents the shaft speed, d_b is the rolling element diameter, D_p is the pitch diameter, and φ represents the angle of the load from the radial plane. It has previously been observed that the frequency of the obtained vibration signal specifies the cause of the fault, and the amplitude shows the fault severity.

In vibration-based machine fault diagnosis practice, we handle a huge gathering of vibration signals obtained from several sources in the machines and some background noises. Consequently, it is challenging to use the raw vibration signals directly for fault diagnosis. Much of the current literature on vibration-based fault diagnosis pays particular attention to introducing methods capable of obtaining useful information, usually called features, from the raw vibration signals, which can be successfully used to classify

the health condition of the machine. Automatic vibration-based condition monitoring employs machine learning classifiers to classify the vibration signal with its correct health condition type using the obtained features as inputs. Various techniques have been introduced for vibration signal analysis that can be utilized to obtain useful features from the raw vibration data. These include time-domain methods, frequency-domain methods, and time-frequency domain methods.

Numerous studies have examined many statistical techniques and some other cutting-edge methods to extract features from vibration signals in the time domain. For instance, McCormick and Nandi conducted several investigations to classify the condition of a small rotating machine using many statistical parameters such as mean and variance estimated from the time series of vibration signals, which are then applied as inputs to multi-layer perceptron and radial basis function neural networks [8]. Jack and Nandi introduced a genetic algorithm (GA) to select the most significant input features to artificial neural networks (ANN) from a large group of statistical estimates in machine condition monitoring situations [9]. In the same vein, Jack and Nandi attempted to improve the overall generalisation performance of support vector machines (SVMs) and ANNs techniques in two-class fault/no-fault recognition by applying a GA-based feature selection process [10]. In [11] a neural network technique for automated fault diagnosis of rolling element bearing using vibration data was proposed. First, the authors examined ten time-domain features including peak value (Pv), root mean square (RMS), standard deviation (SD), Kurtosis value (Kv), Crest Factor (CrF), Clearance Factor (ClF), Impulse Factor (IF), Shape Factor (ShF), Weibull negative log-likelihood (Wnl), and normal negative likelihood as inputs to ANN. Then only Kurtosis and normal negative likelihood are used as input features to ANN. The results showed that the ANN with Kurtosis and negative normal likelihood performed fault diagnosis with the same accuracy as the ANN with the ten-time domain features that were examined first. Furthermore, Prieto et al. presented a technique for bearing fault detection using statistical time-domain features and ANN. In this technique, several time-domain-based features including, mean, maximum value, RMS, SD, variance, ShF, CrF, latitude Factor, IF, Kv, and normalized fifth and sixth moments were computed from the acquired vibration signals [12].

Previous research has established that frequency domain methods can divulge information based on frequency characteristics that are not certain to be observed in the time domain. Various frequency domain methods have been extensively used for vibration signals analysis in the context of bearing fault classification. For example, McCormick and Nandi examined the application of ANN for rotating shaft's fault diagnosis using moments of vibration time series as input features. These features were compared with features computed using the fast Fourier transform (FFT) as a suitable choice for real-time implementation [13]. Li et al. introduced a method for motor-bearing fault detection using frequency domain vibration signals and ANN [14]. In this method, the acquired vibration signals in the time domain were converted into the frequency domain using the FFT method. Then, the converted vibration signals in the frequency domain are used as inputs to train the ANN. Zeng and Wang proposed a framework for machine fault classification that comprises data acquisition, data processing, feature extraction, fault clustering, and fault assignment [15]. In this framework, the acquired vibration was transformed into the frequency domain using the FFT. Dhamande and Chaudhari proposed a method for bearing fault diagnosis based on statistical feature extraction in the time and frequency domain and ANN [16]. In this method, many statistical parameters of the vibration data were computed in the time domain including mean, SD, variance, RMS, the absolute maximum of the vibration signal, skewness, kurtosis, CrF, and a combination of them. Additionally, in the frequency domain, several statistical parameters were estimated including the mean, the variance, the third moment, the fourth moment, the grand mean, SD concerning the grand moment, as well as the third and fourth moments concerning the grand mean [16]. Similarly, Helmi and Forouzantabar proposed a technique for rolling bearing fault detection of electric motor applying time domain and frequency domain features with the

adaptive neuro-fuzzy interface system (ANFIS) network [17]. In this technique, 15 time-domain features such as mean SD, IF, and skewness were computed. Then, the frequency spectrums were obtained using FFT, and 13 frequency-domain features such as mean, frequency centre, and kurtosis were computed [17].

Moreover, data from several studies suggest that the time-frequency domain methods have been introduced to deal with nonstationary waveform signals, which are very common when machine failures happen. The literature on machines fault diagnosis has highlighted several time frequency methods that are employed to transform the vibration signals in the time domain into the time-frequency domain. For instance, Wang and Chen investigated the sensitivity of three time-frequency domain methods, namely, short-time Fourier transform (STFT), wavelet analysis (WA), and pseudo Wigner–Ville distribution (PWVD) for a rotating machine's fault diagnosis [18]. In [19], the authors presented a feature extraction methodology that is based on empirical mode decomposition (EMD) energy entropy for rolling element bearings fault diagnosis. In this methodology, a mathematical analysis process to select the most significant intrinsic mode functions (IMFs) was introduced. Then, the selected features were applied as inputs to train an ANN-based model, which is used to classify bearings faults. Furthermore, Djebala et al. presented a denoising technique based on discrete wavelet analysis of the acquired vibration signals for bearing fault detection [20]. In [21], a deep learning-based approach for bearing fault diagnosis is proposed. In this approach, the acquired signals were preprocessed using STFT to generate a spectrum matrix. Then sub-patterns were generated from the spectrum matrix and used to obtain the optimized deep learning structure, the large memory storage retrieval (LAMSTAR) neural network for bearing fault diagnosis. Furthermore, Immovilli et al. introduced a technique for the detection of generalized-roughness bearing fault using the spectral-kurtosis energy of vibration or current signals [22]. Lei and colleagues presented an improved kurtogram method for fault diagnosis of rolling element bearings. In this method, the wavelet packet transform (WPT) was used as the filter of the kurtogram method to overcome the limitations of the original kurtogram [23]. Recently, Hongwei et al. proposed a method for rolling element bearing fault diagnosis based on Fuzzy C-means (FCM) clustering of vibration images that were obtained using EMD-PWVD [24]. In this method, first, the acquired vibration signals with different fault degrees were converted into contour time-frequency images by using the EMD-PWVD technique. Then, the obtained vibration images were divided into sections and their energy distributions values were used as image features. Furthermore, in [25], the authors presented a method for rotating machinery fault diagnosis using time frequency domain features and CNN knowledge transfer.

On the other hand, despite the various techniques described above, which are used to process and examine vibration signals in the time domain, the frequency domain, and the time frequency domain, various studies have proposed other methods that can reduce the computational complexity and enhance fault classification accuracy. Of these, numerous methods have been introduced for learning subspace features from the raw vibration signals in rotating machine fault diagnosis. For instance, in [26], a method is proposed for incipient failures in large-size low-speed rolling bearings using the multiscale principal component analysis (MSPCA) and the ensemble empirical mode decomposition (EEMD). Guo et al. introduced a feature extraction approach for rolling element bearing fault diagnosis using the envelope extraction and the independent component analysis (ICA) technique [27]. In [28], the authors evaluated the use of the principal component analysis (PCA) technique and NN performance for bearing fault diagnosis. In their experiments, the vibration signals were preprocessed using detrended-fluctuation analysis (DFA) and rescaled-range analysis (RSA) techniques. Additionally, Dong et al. introduced a technique for bearing fault diagnosis using kernel PCA (KPCA) and an optimized k-nearest neighbour model [29]. In this technique, first, the original vibration signals were decomposed using local mean decomposition (LMD). Then, the entropy values of product functions that represent the input features were computed utilizing the Shannon method. The

KPCA was used to reduce the dimension of the original features needed to train the k-nearest neighbour model [29].

Additionally, data from several studies demonstrate that the use of feature selection techniques will reduce the computational cost and might remove irrelevant and redundant features and accordingly may improve learning performance [1]. Feature selection techniques can be grouped into filter models (e.g., Fisher score, Relief and Relief-F algorithms), wrapped models (e.g., Genetic algorithm), and embedded models (e.g., LASSO and elastic net). Numerous studies have investigated the application of feature selection methods in the context of vibration-based bearings fault diagnosis. For example, Haroun et al. introduced a feature selection method for bearing fault detection and diagnosis using a self-organizing map (SOM) [30]. In this method, the authors, employed multiple methods from the time domain, frequency domain, and time-frequency domains to extract features. Then, Relief-F and minimum redundancy maximum relevance (mRMR) were used to select the optimal features from the extracted features. With these selected features, the SOM was applied to classify the bearing's health condition [30]. Furthermore, a method for machinery fault diagnosis using redefined dimensionless indicators (RDIs) and mRMR was introduced [31]. In this method, first, the original vibration signals were pre-processed using variation mode decomposition to construct multiple RDIs. Then, the mRMR technique was employed to select several important RDIs. Finally, with these selected RDIs, a grid support vector machine (SVM) was used to carry out the identification of machinery faults. In [32], the authors proposed a methodology for bearing fault diagnosis of induction motors using a genetic algorithm (GA) and machine learning classifiers. In this methodology, first, some statistical features were obtained from the raw signals. Then, the GA was employed to select the most important features. Finally, with the selected features, three different classification algorithms namely, k-nearest neighbour (KNN), decision tree (DT), and random forest (RF) were trained to accomplish the classification task.

Furthermore, recent advances in dimensionality reduction methods have aided the investigation of the compressive sampling (CS) technique [33]. Several researchers have used CS to reduce the dimensionality of the original vibration signals for rolling bearing fault classification. For instance, Wong et al. examined the effects of CS on bearing fault classification [34]. In this investigation, the authors resampled the originally collected vibration signals using a random Bernoulli matrix to match the compressive sampling process. Then, sample entropy-based features were obtained from both the normalized raw vibration signals and the reconstructed signals. Finally, the SVM was trained using the obtained features to accomplish the fault classification task [34]. Tang et al. proposed a CS framework of characteristics harmonics to detect bearing faults [34]. In this method, the characteristics harmonics were obtained from sparse measurements via a compressive matching pursuit technique during the procedure of incomplete reconstruction [35]. Moreover, Xinpeng et al. introduced a bearing fault detection technique using CS and matching pursuit (MP) reconstruction algorithm [36]. In [4], a method for bearing fault classification from highly compressed measurements is proposed. In this method, CS was used to produce highly compressed measurements of the original bearing vibration signals. Then, a deep neural network (DNN) based on sparse autoencoder (SAE) was utilized to learn overcomplete sparse representations of the compressed measurements, which were used for the classification of bearing fault using the Softmax layer. Moreover, Ahmed and Nandi proposed a three-stage method for rolling bearings fault diagnosis using CS and subspace learning techniques [37]. In this method, the CS technique was employed to obtain compressively sampled vibration signals from the original vibration signals. Then, a multistep feature learning algorithm using PCA, linear discriminant analysis (LDA), and canonical correlation analysis (CCA) was used to obtain fewer features from the compressively sampled signals [37]. In [38], a framework for bearing fault classification using CS and feature ranking is proposed. Then, the authors used the CS process to produce com-

pressively sampled signals from the raw vibration signals using two compressible representations of vibration signals, namely, Fourier transform-based coefficients and threshold Wavelet transform-based coefficients. Then, various feature ranking procedures were used to select fewer features from the compressively sampled signals. Finally, three classifiers were evaluated for the classification of bearing faults using these selected features.

The issue of collecting a large amount of vibration data for machine fault diagnosis has attracted considerable attention as it requires large storage to be stored and time to be processed. Existing research has highlighted various techniques for vibration signal analysis that can be applied to obtain useful features from the originally collected vibration data. However, the number of obtained features could be a contributing factor to the performance of these techniques in terms of classification accuracy and computational time, which are particularly important in the real implementation of fault diagnosis techniques. This study aims to contribute to this growing area of research by investigating the following: 1. A method that can reduce the high dimensionality of the raw vibration data to a fewer number of features capable of achieving high fault classification accuracy and highly reduced computational time, 2. The CS is an appropriate mechanism to compress the original high dimensional vibration data and then further reduce the dimension of the compressed vibration data to a far fewer number of features that satisfactorily represent the health condition of rolling bearings, and 3. The multinomial logistic regression (MLR) algorithm is a possible classifier to deal with the bearing's fault classification task using the fewer selected features.

To accomplish high classification accuracy and highly reduced computation time, this paper proposes a new methodology for bearing fault classification based on intrinsic dimension estimation-based feature selection and multinomial logistic regression using compressively sampled vibration signals. In the methodology, the input vibration signals are resampled using CS to reduce the high-dimensional samples of the originally collected vibration data. Then, to further reduce the number of features of the compressively sampled vibration signals to far fewer features that can sufficiently represent the health condition of rolling bearings consequently achieve high classification accuracy and highly reduced computation time, a feature selection procedure based on intrinsic dimension estimation, stochastic proximity embedding (SPE), and neighbourhood component analysis (NCA) is applied. Finally, to perform the bearing's fault classification task, the fewer selected features are applied as inputs to a multinomial logistic regression (MLR) classification algorithm. The contributions of this paper are as follows:

1. The proposed method produces far fewer features that can represent the health condition of bearings. This study accomplishes high classification accuracies and highly reduced computational time with regression analysis-based predictive modelling technique, namely the multinomial logistic regression (MLR) classifier using the fewer selected features as inputs.
2. A dimensionality reduction process has been proposed, which comprises (1) data compression using CS and (2) intrinsic dimension estimation-based feature selection process, which includes (a) SPE-based feature selection that utilizes a self-organizing iterative scheme to embed the compressed data dimension into a further lower dimension, and (b) the non-parametric NCA-based feature selection that maximizes the stochastic variant of the leave-one-out nearest neighbour score to achieve the best classification accuracy on the training set. This ensures selecting fewer features from the high dimensional data capable to achieve high classification accuracy and a reduced computational time.
3. We studied the impact of values of two parameters within the data compression and feature selection process used in the proposed method, namely the compressive sampling rates and the NCA tolerance values on the number of the selected features and the fault classification accuracy.
4. Two fault classification case studies of rolling element bearings vibration signals under different working loads are used to evaluate the proposed method.

5. Compared to recently published classification results from the literature on the same vibration-bearing datasets used in this study, our proposed method achieves high classification accuracy and highly reduced computation time, which suggests that our proposed methodology could be used in actual applications of vibration-based machine fault diagnosis.

The remainder of this paper is organized as follows. Section 2 describes the proposed framework. Section 3 is devoted to descriptions of the experimental study used to validate the proposed framework and presents comparison results. Finally, Section 4 offers some conclusions.

2. The Proposed Method

The methodological approach taken in this study is a mixed methodology based on vibration data compression, feature selection based on intrinsic dimension estimation, and regression analysis-based predictive modelling techniques. The proposed method automatically learns and selects far fewer features from compressively sampled vibration signals, which can be used as inputs to a classifier for bearing fault detection and classification. The key objective of the proposed method is to achieve high fault classification accuracy while highly reducing the computation time. The flow chart of the proposed method is presented in Figure 3. First, the compressive sampling (CS) mechanism was employed to compress the acquired vibration data. Then, a feature selection procedure based on intrinsic dimension estimation, stochastic proximity embedding (SPE), and neighbourhood component analysis (NCA) was utilized to estimate and further reduce the dimensionality of the compressively sampled data. Finally, with these reduced features, a classifier was used to classify the bearing's health condition, namely, the multinomial logistic regression (MLR) algorithm was employed to perform the classification task. The following subsections discuss the proposed method in more detail.
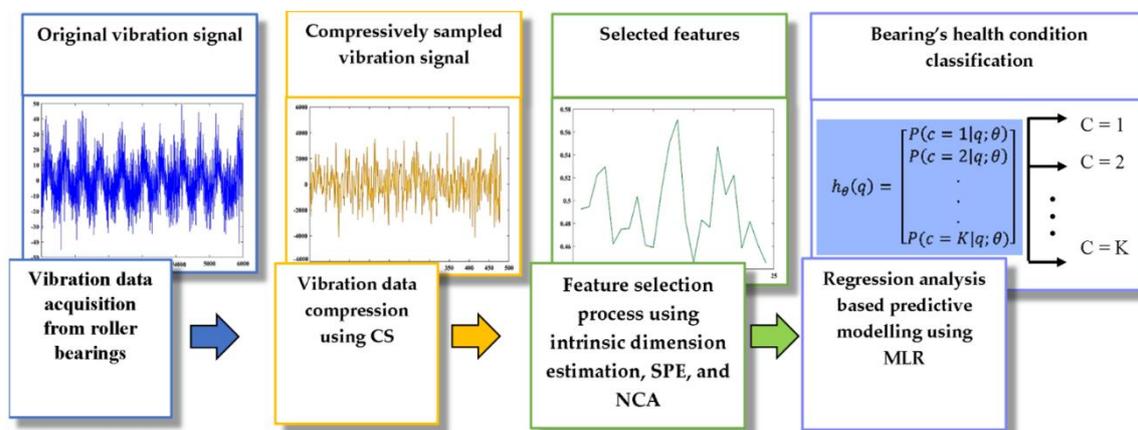
Figure 3. The proposed method.

2.1. Vibration Data Compression Using CS

To reduce the high dimensional of the collected vibration data, the proposed method uses the CS mechanism to obtain compressively sampled vibration signals from the original signals. The central principle of the CS is that a finite-dimensional signal having sparse or compressible representations can be reconstructed from a small number of linear measurements much lower than measurements based on the Nyquist sampling rate {xe "Nyquist sampling rate"}. Machine vibration signal has compressible representations in several domains such as in the frequency domain using FFT {xe "frequency domain"}. Therefore, in recent times, there has been a growing interest in the application of CS in machine fault diagnosis {xe "*fault diagnosis*"}. There are many benefits of CS in vibration-based bearing fault diagnosis, e.g., reducing the high dimension of the acquired vibration

data, reducing computation time required to analyze the collected data, reducing data transmission cost in the cases where it is essential to send the collected data from remote places, e.g., fault diagnosis of offshore wind turbines.

The successful implementation of the CS mechanism is based on two fundamental concepts: (i) the sparsity of the targeted signal, and (ii) the measurements matrix that fulfils the minimal data information loss, which is usually called the restricted isometry property {xe "restricted isometry property"} (RIP) [39]. Briefly, we describe the CS mechanism as follows.

Let $x \in R^{n \times 1}$ be the originally collected time-indexed signal. With an identified sparsifying transform {xe "sparsifying transform"} matrix $\psi \in R^{n \times n}$ where the columns represent the basis elements $\{\psi_i\}_{i=1}^n$, the signal x can be described as follows,

$$x = \sum_{i=1}^{n} \psi_i \, s_i \tag{5}$$

or,

$$x = \psi s \tag{6}$$

Here, s represents a $n * 1$ column vector of coefficients. In case the basis ψ generates q-sparse representations of the signal x then x of length n can be signified with $q \ll n$ nonzero coefficients. Consequently, Equation (5) can be rewritten as follows,

$$x = \sum_{i=1}^{q} \psi_{ni} \, s_{ni} \tag{7}$$

Here, ni represents the index of the basic elements and the coefficients corresponding to the q nonzero elements. Accordingly, $s \in R^{n \times 1}$ represents a vector column with q nonzero elements and characterizes the sparse representation {xe "sparse representation"} vector of the signal x. Consistent with the CS mechanism, with $m \ll n$ projections of the vector x, measurement vectors $\{\emptyset_j\}_{j=1}^m$, and the sparse representations s, the compressed measurements of the signal x can be obtained using the following equation,

$$y = \emptyset \psi s = \theta s \tag{8}$$

Here, y is a $m * 1$ column vector of the compressed measurements and $\theta = \emptyset \psi$ represents the measurement matrix. Figure 4 shows an illustration of the CS framework that can be used to produce the single measurement vector of the compressed measurements y. According to the CS theory, the original signal x can be reconstructed from the compressed measurements y by applying a recovery algorithm. This can be completed by first recovering the sparse representation vector s and then employing the inverse of the sparsifying transform {xe "sparsifying transform"} ψ to recover x. One of the solutions to be used to recover the sparse representations $s \in R^n$ from its compressed measurement vector $y \in R^m$ is the l_0 minimization technique, which searches for a sparse vector consistent with the measured data $y = \theta s$ such that,

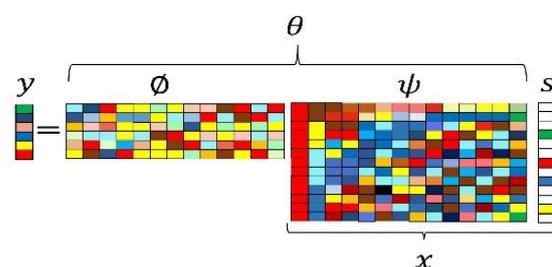$$\hat{s} = \arg \min_{z} \|s\|_0 \text{ such that } \theta s = y \tag{9}$$

Figure 4. Single measurement vector compressive sampling framework [37].

Moreover, the convex optimization $\|.\|_1$ can also be employed in place of $\|.\|_0$, such that,

$$\hat{s} = \arg\min_z \|s\|_1 \text{ such that } \theta s = y \tag{10}$$

In case the measurement matrix θ satisfies the {xe "Restricted Isometry Property"} RIP, the sparse representation s can be reconstructed by solving the convex program in Equation (10). The matrix θ satisfies the *rth* restricted isometry property {xe "Restricted Isometry Property"} (RIP) if there exists a $\delta_r \ll 1$, such that

$$(1 - \delta_r)\|s\|_{l_2}^2 \leq \|\theta s\|_{l_2}^2 \leq (1 + \delta_r)\|s\|_{l_2}^2 \tag{11}$$

In our case, the collected vibration data is a collection of signals, which can be represented as a matrix of sparse vectors Y, such that,

$$Y = \Theta S \tag{12}$$

where $Y \in R^{m \times L}$, L is the number of observations and m is the number of compressed measurements, $\Theta \in R^{m \times n}$ represents a dictionary, and $S \in R^{n \times L}$ is a sparse representation {xe "sparse representation"} matrix. Therefore, multiple measurement vector compressive sampling {xe "Compressive Sampling"} is used in our proposed method.

Our proposed method is intended to learn directly from the compressed vibration signals. To obtain the compressively sampled signals from the collected vibration dataset $X = \{x_1, x_2, \dots, x_L\} \in R^n$, first, the {xe "Fast Fourier Transform"} FFT, which commonly provides a sparse basis for vibration signals, is employed to produce the sparse representation ($S \in R^{nxL}$) that comprises only a small number $q \ll n$ of nonzero coefficients. The FFT algorithm calculates the n-point complex discrete Fourier transform (DFT) of the signal X. In this study, we utilise the magnitude of the DFT to get S. Then, a random matrix with i.i.d Gaussian entries matrix, which satisfies the RIP, is used as the measurement matrix $\Theta \in R^{mxn}$ [40]. Finally, a compressed sampling rate (α) is used to produce the compressively sampled signals $Y \in R^{mxL}$, where m represents the number of compressed signal elements and given by $m = \alpha * n$. This compression process is summarized in Algorithm 1 below:

Algorithm 1 Vibration data compression using CS

---

1. Input: vibration dataset $X \in R^{n \times L}$; measurement matrix $\Theta \in R^{mxn}$; and compressive sampling rate α
2. Output: compressively sampled vibration signals $Y \in R^{m \times L}$
3. Produce the sparse representations S of X: $abs(FFT(X)) \rightarrow S \in R^{n \times L}$
4. Project S into Θ with compressed sampling rate α to obtain compressively sampled signals $Y \in R^{m \times L}$

2.2. Feature Selection Process

Based on CS theory, the compressively sampled signals (Y) have sufficient information to reconstruct successfully the originally collected vibration signals. Nevertheless, the dimensions of these compressively sampled vibration signals might be further reduced to attain more reduction in the computational cost while achieving high classification accuracy. Accordingly, our proposed method offers a feature selection process to learn and select fewer features from the compressively sampled signals (Y) to achieve superior classification accuracy and reduced computation costs. The feature selection process starts by identifying the minimal number of features required to represent the compressively sampled vibration signals Y, using a global dimension estimator, namely the geodesic minimal spanning tree (GMST). The GMST calculates the geodesic graph G from which the intrinsic dimension (d) is projected by calculating multiple minima spanning trees (MSTs) in which each data sample x_i is linked to its k nearest neighbours [41], such that,

$$d(Y) = min \sum_{e \in T} D_{Eucl} \tag{13}$$

Here, T signifies the set of all the subtrees of G, e is an edge in T, and D_{Eucl} is the Euclidean distance of e. Then, with the computed minimal number of features d, where $d < m$, the SPE technique is employed to convert the compressively sampled data (Y) into a reduced-dimensionality space of significant representation ($Z \in R^{d \times L}$).

The SPE is a nonlinear approach that has many benefits such as being simple to implement, very fast, scales linearly with the size of the data in both time and memory, and is relatively insensitive to missing data [42]. Thus, it was decided that SPE is an appropriate method to use for this investigation. The SPE utilizes a self-organizing iterative scheme to embed m-dimensional data into d dimensions, such that the geodesic distances in the original m dimensions are preserved in the embedded d dimension. Briefly, we describe the simplified SPE procedure as follows [43]:

1. Initialize the coordinates y_i. Select an initial learning rate β.
2. Select a pair of points, i, and j, at random, and calculate their distance: $d_{ij} = \|y_i - y_j\|$. If $d_{ij} \neq r_{ij}$ (r_{ij} is the distance of the corresponding proximity), update the coordinates y_i and y_j using the following equations,

$$y_i \leftarrow y_i + \beta \frac{1}{2} \frac{r_{ij} - d_{ij}}{d_{ij} + v} (y_i - y_j) \tag{14}$$

and

$$y_j \leftarrow y_j + \beta \frac{1}{2} \frac{r_{ij} - d_{ij}}{d_{ij} + v} (y_j - y_i) \tag{15}$$

Here, v is a small number to avoid division by zero. For a given number of iterations, this step will be repeated for several steps and β will be reduced by a recommended decrement $\delta\beta$. Finally, to obtain far fewer selected features in the feature selection process step, our proposed method uses the NCA technique to automatically selects a subset from the SPE-based learned features by converting $Z \in R^{d \times L}$ into $Q \in R^{f \times L}$ where $f < d$. Briefly, we describe the NCA feature selection as follows:

Let $Z = \{(z_1, c_1), ..., (z_i, c_i), ..., (z_L, c_L\}$ be a training set samples with d-dimension and $c_i \in \{1, ..., C\}$ is the matching class label. The NCA searches for a weighting vector w that uses to select a feature subset. In this method, first, the weighted distance between two samples z_i and z_j can be computed using the following equation,

$$D_w(z_i, z_j) = \sum_{r=1}^{d} w_r^2 |z_{il} - z_{jl}| \tag{16}$$

Here, w_r is a weight-related to the *r-th* feature. Then, the strategy is to maximize the leave-one-out classification accuracy on the training set. The reference point is defined by a probability distribution. In our case, the probability of z_i chooses z_j as its reference point such that,

$$p_{ij} = \begin{cases} \dfrac{k(D_w(z_i, z_j))}{\sum_{k \neq i} k(D_w(z_i, z_j))}, \text{if } i \neq j \\ 0, \text{if } i = j \end{cases} \tag{17}$$

Here, $k\left(D_w(z_i, z_j)\right) = \exp\left(-(D_w(z_i, z_j))/\sigma\right)$ is a kernel function and σ is an input parameter that represents the kernel width. The probability of correct classification of y_i can be computed using the following equation,

$$p_i = \sum_j c_{ij} p_{ij} \tag{18}$$

Here,

$$c_{ij} = \begin{cases} 1 \text{ if } c_i = c_j \\ 0 \text{ otherwise} \end{cases} \tag{19}$$

This process of NCA is summarized in Algorithm 2 below [44]:

Algorithm 2 NCA Feature Selection

---

1.   Input:

$Z \in R^{d \times L}$; γ: initial step length; σ: kernel width; λ: regularisation parameter; and η: small positive constant.

2.   Initialization: $w^{[0]} = (1, 1, \dots, 1)$, $\epsilon^{(0)} = -\infty$, $t = 0$.

3.   Repeat

4. $for\ i = 1, \dots, L\ \ do$

5.   Compute p_{ij} and p_i using $w^{(t)}$ according to (2) and (3)

6. $for\ r = 1, \dots, d\ \ do$

7. $\nabla\_r = 2(1/\sigma(p_i \sum_{j \neq i} p_{ij}|z_{ir} - z_{jr}| - \sum_j c_{ij}|z_{ir} - z_{jr}|) - \lambda)w_r^{(t)}$

8. $t = t + 1$

9. $w^{(t)} = w^{(t-1)} + \gamma\Delta$

10. $\epsilon^{(t)} = \xi(w^{(t-1)})$

11.   if $\epsilon^{(t)} > \epsilon^{(t-1)}$ then γ = 1.01 γ

12.   else

13.   γ = 0.4γ

14.   until $|\epsilon^{(t)} - \epsilon^{(t-1)}| < \eta$

15. $w = w^t$

16.   Return w

To select the most important features based on features weights, the criteria are based on a threshold value (*Thr*), which can be computed as follows:

$$Thr = \tau \max(w) \tag{20}$$

Here τ is the tolerance value.

2.3. Regression Analysis-Based Predictive Modelling Using MLR

Regression analysis can be used as a predictive modelling method as it defines the relationship between a dependent variable and one or more independent descriptive variables. There are many types of regression analysis methods such as linear regression, polynomial regression, logistic regression, etc. Of these, logistic regression (LR) is one of the most used methods in many machine learning applications. The LR is usually utilized for binary classification, i.e., the class labels c has only two values, e.g., (Fault, Normal). Briefly, we describe the LR as follows:

Let a training data $Q = \{q_1, q_2, \dots, q_L\}$ with f-dimension produced in the feature selection step of our proposed method. The LR is a probabilistic discriminative model that learns $P(c|Q)$ directly from the training data where $c_i \in \{0, 1\}$, such that,

$$P(c = 1|q) = h_1(q) = g(-\theta^T h) = 1/(1 + e^{-\theta^T q}) \tag{21}$$

Here $g(-\theta^T h)$ is the logistic function, which is also called the sigmoid function. Since $\sum P(C) = 1$, we can compute $P(c = 0|Q)$ as follows:

$$P(c = 0|q) = h_0(q) = 1 - P(c = 1|q) = 1 - (1/(1 + e^{-\theta^T q})) \tag{22}$$

The likelihood of the parameters of L training examples can be computed using the following equation,

$$L(\theta) = \prod_{i=1}^{L} ((g(\theta^T q^i))^{c(i)}(1 - g(\theta^T q^i))^{1-c(i)} \tag{23}$$

Here $\theta = [\theta_0, \theta_1, \ldots, \theta_i]$ represents the parameters of the model. Nevertheless, the log-likelihood is widely utilised and, consequently, Equation (23) and can be updated as Equation (24), such that

$$\log L(\theta) = \sum_{i=1}^{L} \log \left(\left(\left(g(\theta^T q^i) \right)^{c(i)} (1 - g(\theta^T q^i))^{1-c(i)} \right) \right) \tag{24}$$

To avoid overfitting a regularisation term, λ is added to the log-likelihood function, such that

$$\log L(\theta) = \sum_{i=1}^{L} \log \left(P(c^i = c_k | q^i; \theta) \right) - \frac{\lambda}{2} \|\theta\|^2 \tag{25}$$

The MLR classifier, which also goes with the name SoftMax regression in ANN, generalizes the LR to a multi-class classification problem with multi-labels $c^{(i)} \in \{1, \ldots, K\}$, such that

$$h_\theta(q) = = \begin{bmatrix} P(c = 1 | q; \theta) \\ P(c = 2 | q; \theta) \\ \cdot \\ \cdot \\ \cdot \\ P(c = K | q; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} \exp(\theta^{(j)T}q)} \begin{bmatrix} e^{\theta^{(1)T}q} \\ e^{\theta^{(2)T}q} \\ \cdot \\ \cdot \\ \cdot \\ e^{\theta^{(K)T}q} \end{bmatrix} \tag{26}$$

Here, $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(K)}$ represent the parameters of the multinomial logistic regression model. In this study, we are dealing with a multiclassification problem, so MLR is employed to perform the classification task in our proposed method.

3. Experimental Study

Two fault classification case studies of rolling element bearings using vibration signals are presented to evaluate the proposed method.

3.1. First Case Study

The vibration dataset used in this case study is acquired from experiments on a test rig that simulates running roller bearings' environment. In these experiments, several interchangeable faulty roller bearings are inserted in the test rig to symbolize the type of faults that can normally happen in roller bearings. As shown in Figure 5, the test rig to collect the vibration dataset of bearings contains a 12V DC electric motor driving the shaft via a flexible coupling. The shaft was supported by two blocks of Plummer bearing, where several damaged bearings were inserted. Two accelerometers were used to measure the vibration signals in both the horizontal and vertical planes. The output from the accelerometers was fed back over a charge amplifier to a Loughborough Sound Images DSP32 ADC card with a low-pass filter using a cut-off of 18 kHz. The sampling rate was 48 kHz. Six health conditions of roller bearings have been recorded with two normal conditions {XE "normal conditions"}, i.e., brand new condition (NO) and worn yet undamaged condition (NW), and four faulty conditions {XE "faulty condition"} with, inner race fault {XE "inner race fault"} (IR), an outer race fault {XE "outer race fault"} (OR), rolling element fault {XE "rolling element fault"} (RE), and cage fault (CA). Table 1 explains the corresponding characteristics of these bearing health conditions [4].
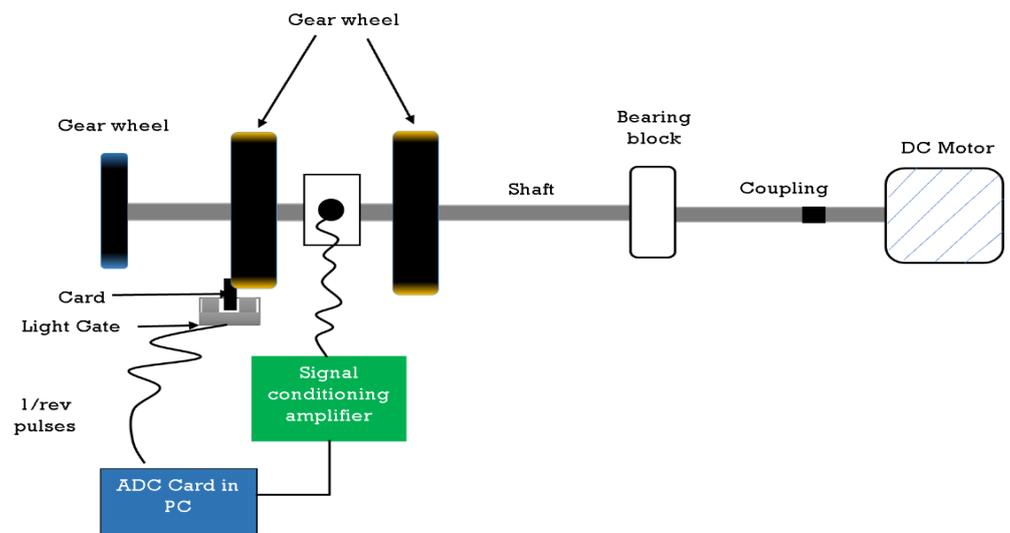
Figure 5. The test rig used to collect the vibration data of bearings of the first case study [4].

The data was recorded using 16 different speeds within 25–75 rev/s. In each speed, ten time series were recorded for each condition, i.e., 160 examples per condition. This resulted in a total of 960 examples with 6000 data points each. Figure 6 illustrates some typical time series plots for the six different conditions.
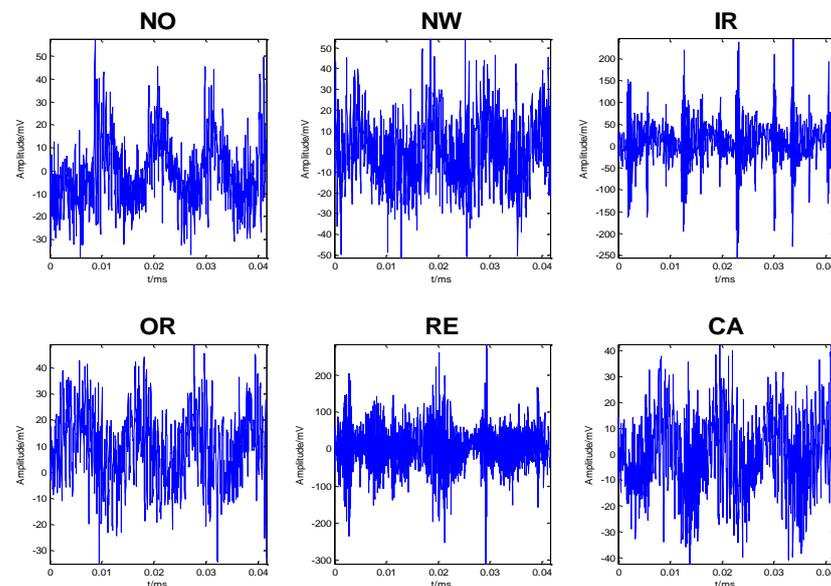
Figure 6. Typical time domain vibration signals for the six different conditions [4].

Table 1. The characteristics of bearings' health conditions in the obtained bearing dataset.

Condition	Characteristic
NO	The bearing was brand new and in perfect condition.
NW	The bearing was in service for some time but in good condition.
IR	Inner race fault. This fault was created by cutting a small groove in the raceway of the inner race.
OR	Outer race fault. This fault was created by cutting a small groove in the raceway of the outer race.
RE	Roller element fault. This fault was created by using an electrical etcher to mark the surface of the balls, simulating corrosion.

| CA	Cage fault. This fault was 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 in a regular space and had freedom to move.

To apply our proposed method in this case study, first, we randomly selected 50% of the total observations for training, and the remaining 50% is employed for testing the trained model. Then, we computed the compressively sampled vibration signal from the high dimensional data X which has 6000 time samples for each of its 960 observations. As described in algorithm 1, the FFT basis was used as the sparse representation of each signal in X. After, the CS framework with different sampling rates (α) (0.1, 0.2, and 0.3) using a random Gaussian matrix was used to produce the compressed measurements.

To estimate and reduce further the dimensionality of the compressively sampled signals, first, we computed the intrinsic dimension (d) using the GMST technique. Then, the process combining the SPE and NCA techniques was performed to learn and select far fewer features from the compressively sampled vibration signals. The compressively sampled vibration signals were transformed into further reduced dimensions using the defined intrinsic dimension and the SPE method. Then, a regularized NCA based on feature weights and a relative threshold was employed to select far fewer features with f-dimension from the SPE-based selected features with d-dimension where $f < d$. We computed the best regularization parameter (λ) value that corresponds to the minimum average loss to be used in fitting the NCA model on all the reduced dimension data. The final selected features were computed using the feature weights of the NCA model and a relative threshold. The stochastic gradient descent (SGD) solver was used for estimating feature weights. Two tolerance values (0.01 and 0.02) were tested in this case study for computing the threshold values used in the feature selection process. Figure 7 shows an example of the average loss values versus λ values. Figure 8 presents an example of the selected features and their corresponding weights. Moreover, Table 2 shows examples of the computed values of the average intrinsic dimension, the dimension of the NCA-based selected features, least loss, and best λ values taken from 10 trials.

The first benefit of the proposed method is to obtain far fewer features from the acquired vibration signals to be successfully used for rolling bearing fault diagnosis and consequently reduce the computational cost. Therefore, the first set of analyses examined the impact of the tolerance values and the compressive sampling rates on the number of the selected features using our proposed method. As shown in Table 2, the average least losses were slightly reduced from 0.014 to 0.013; 0.013 to 0.009; and from 0.013 to 0.010 using the tolerance value of 0.01 in place of 0.02. Furthermore, with a 0.02 tolerance value, the computed average best lambda value for NCA remained the same for all the compressive sampling rates (with $\lambda = 0.003$), while for the tolerance value of 0.01, the value $\lambda = 0.004$ achieved for the sampling rate of $\alpha = 0.3$, and $\lambda = 0.004$ was obtained for both $\alpha = 0.1$ and $\alpha = 0.2$.

Moreover, as Table 2 shows, all the computed average least lost values are very small—in the range of 0.009–0.014—although the feature dimension was reduced from 6000 (in the original raw vibration signals) to 600 (the dimension of the compressively sampled signals with $\alpha = 0.1$), which reduced to 28 (*d-dimension*) and then further reduced to 8 (*f-dimension wit 0.01 tolerance value*). This suggests that the feature selection is a good idea to be performed in our proposed method.
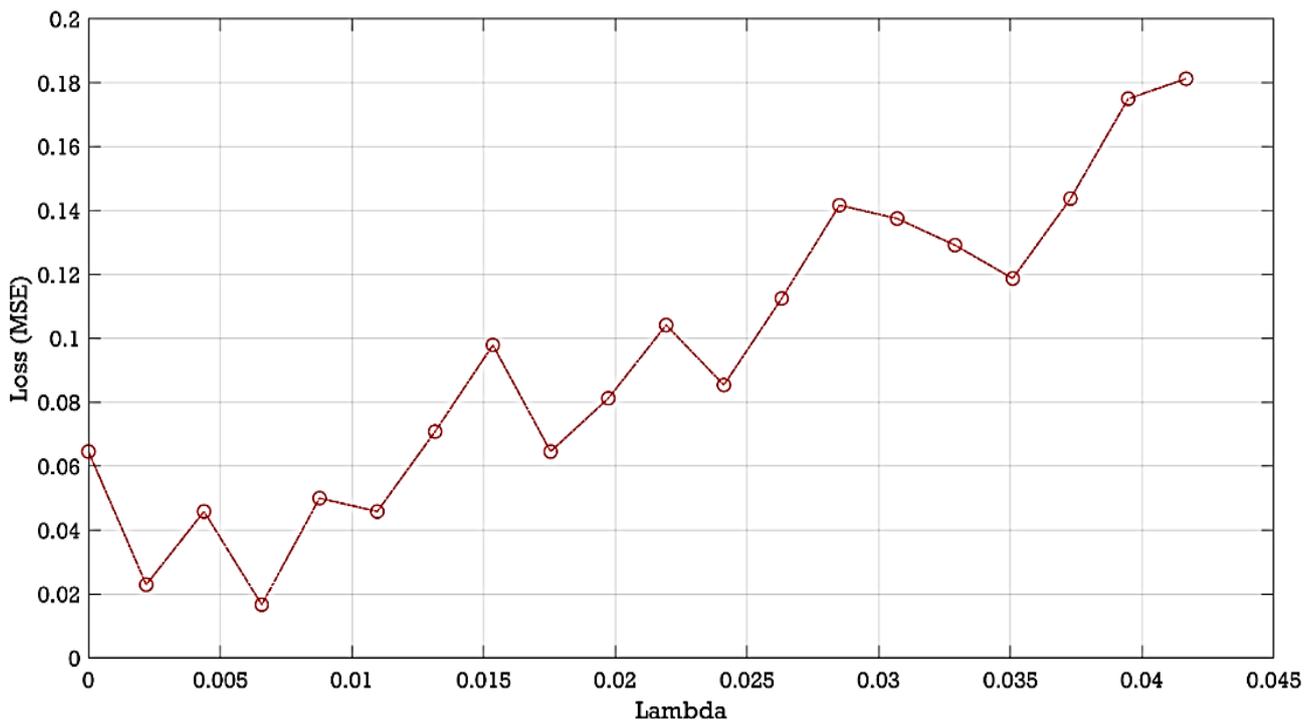
Figure 7. Example of the average loss values versus λ values computed from the reduced dimension of compressively sampled data with $\alpha = 0.2$.
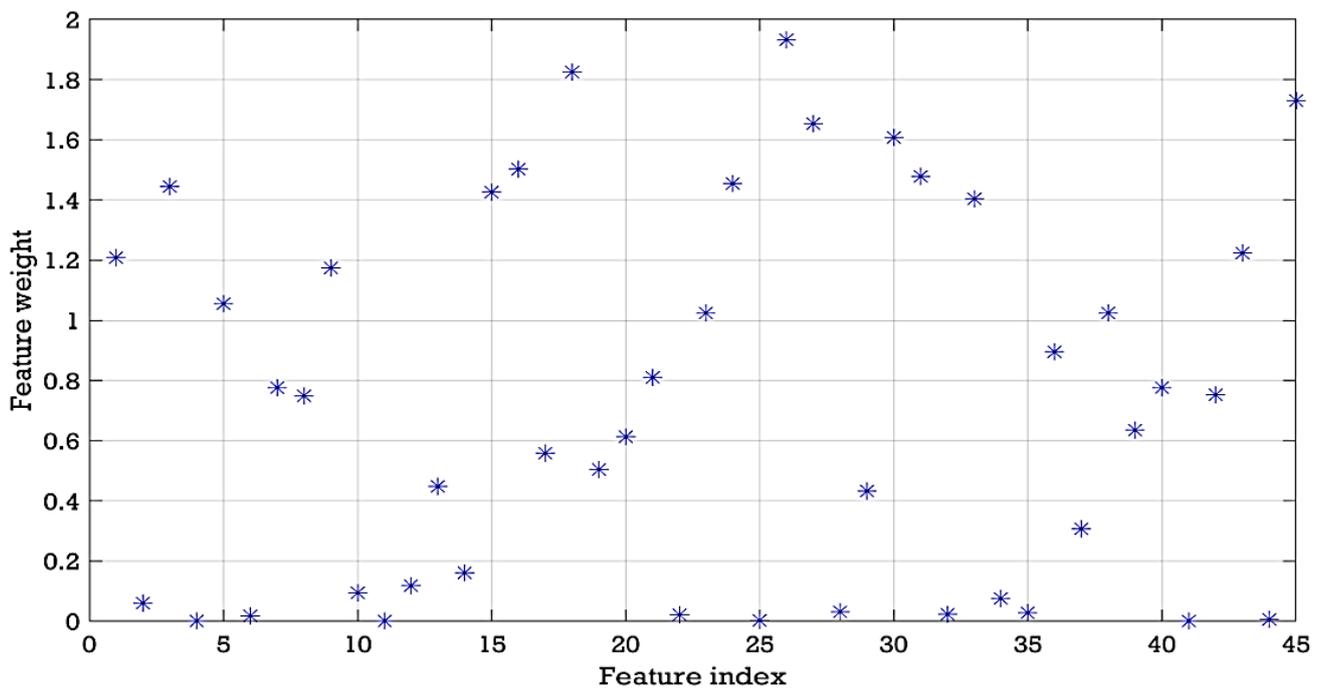
Figure 8. Example of the selected features and their corresponding weights using $\alpha = 0.2$ and NCA tolerance value = 0.02.

Furthermore, the NCA tolerance values have a clear effect on the computed intrinsic dimension (*d*) and the dimension of the NCA-based selected features. For example, with a 0.02 tolerance value, we obtained intrinsic dimension (*d*) of 62, 55, 33, which reduced to 28, 40, and 26 with 0.01 tolerance value for $\alpha = 0.1$, $\alpha = 0.2$, and $\alpha = 0.3$, respectively. Furthermore, with the tolerance value of 0.02, we obtained the final feature dimension (*f*) of 18, 14, and 11, which reduced to 8, 10, 8 with 0.01 tolerance for $\alpha = 0.1$, $\alpha = 0.2$, and

$\alpha = 0.3$ respectively. Taken together when the tolerance value decreases the dimension of both selected features, i.e., d and f, decreases.

Table 2. Examples of the computed values of the average intrinsic dimension, the dimension of the NCA-based selected features, least loss, and best λ values taken from 10 trials.

NCA Tolerance Value	CS Sampling Rate (α)	Average Intrinsic Dimension (d)	The Average Dimension of NCA-Based Selected Features (f)	Average Least Loss	Average Best Lambda for NCA
0.01	0.1	28	8	0.013	0.004
	0.2	40	10	0.009	0.004
	0.3	26	8	0.010	0.003
0.02	0.1	62	18	0.014	0.003
	0.2	55	14	0.013	0.003
	0.3	33	11	0.013	0.003

The final NCA-based selected features were used to train the classification algorithm, i.e., the MLR, to classify among c classes of roller bearing health conditions. The overall results are shown in Table 3, where the classification accuracy is the average of 10 trials for each experiment and the time obtained by averaging the training time and testing time of these 10 trials. It is apparent from this Table that our proposed method achieved high classification accuracies (all above 99%) for all the compressive sampling rates and tolerance values with less than 35% of the acquired vibration data samples. Classification accuracies from our proposed method are 99.9%, 99.7%, and 99.5% for only 30%, 20%, and 10% of the whole collected data with 8, 10, 8 selected features (with tolerance value = 0.01) respectively, used to train the MLR classifier. Additionally, with tolerance value = 0.02, $\alpha = 0.3$, and 11 selected features, the proposed method achieved 100% classification accuracy for every single run in our experiments. Moreover, the trained classification model of our proposed method requires less than 0.016 s to complete the classification task.

Table 3. Classification results with their corresponding RMSE and computational time for the automatically selected features (d refers to the intrinsic dimension, and f is the dimension of the NCA-based selected feature) using two values of NCA tolerance and three compressive sampling rates.

NCA Tolerance Value	CS Sampling Rate (α)	d	f	MLR Classifier			
				Training Accuracy (%)	Training Time (s)	Testing Accuracy (%)	Testing Time (s)
0.01	0.1	28	8	99.8 ± 0.3	5.34 ± 1.7	99.5 ± 0.6	0.015 ± 0.002
	0.2	40	10	99.9 ± 0.1	4.6 ± 2.3	99.7 ± 0.3	0.003 ± 0.00
	0.3	26	8	100 ± 0.0	3.3 ± 0.5	99.9 ± 0.1	0.003 ± 0.001
0.02	0.1	62	18	99.9 ± 0.2	3.37 ± 0.8	99.7 ± 0.3	0.003 ± 0.001
	0.2	55	14	99.9 ± 0.1	3.55 ± 0.9	99.8 ± 0.2	0.003 ± 0.001
	0.3	33	11	100 ± 0.0	4.6 ± 2.0	100 ± 0.0	0.004 ± 0.003

Table 4 provides sample confusion matrices of the classification results of MLR classifier using selected features with tolerance value = 0.01 and a sampling rate of (a) $\alpha = 0.1, 0.2$, and $\alpha = 0.3$. As can be seen from Table 4c, the recognition of the bearing health conditions with $\alpha = 0.3$ is 100%. In Table 4a, with 10% testing data (with $\alpha = 0.1$), our method misclassified one of the testing examples of condition 5, i.e., RE, as condition three, i.e., IR. In Table 4b, with 20% testing data (with $\alpha = 0.2$), our method misclassified only two of the testing examples of condition 2 (NW) as condition 6 (CA).

Table 4. Sample confusion matrices of the classification results of MLR classifier using selected features with tolerance value = 0.01 and a sampling rate of (a) $\alpha = 0.1$, (b) $\alpha = 0.2$, and (c) $\alpha = 0.3$.

	NO	NW	IR	OR	RE	CA
NO	80	0	0	0	0	0
NW	0	80	0	0	0	0
IR	0	0	80	0	0	0
OR	0	0	0	80	0	0
RE	0	0	1	0	79	0
CA	0	0	0	0	0	80

(a)

	NO	NW	IR	OR	RE	CA
NO	80	0	0	0	0	0
NW	0	78	0	0	0	2
IR	0	0	80	0	0	0
OR	0	0	0	80	0	0
RE	0	0	0	0	80	0
CA	0	0	0	0	0	80

(b)

	NO	NW	IR	OR	RE	CA
NO	80	0	0	0	0	0
NW	0	80	0	0	0	0
IR	0	0	80	0	0	0
OR	0	0	0	80	0	0
RE	0	0	0	0	80	0
CA	0	0	0	0	0	80

(c)

Comparisons of Results

In this subsection, a comparison of various methods using the same vibration dataset of rolling bearings used in the first case study (see Table 5). In [34], three methods were used for bearing fault diagnosis using SVM. The first method used the whole collected vibration data. The second method used compressively sampled datasets of $\alpha = 0.25$ and $\alpha = 0.5$, while the third method used the corresponding reconstructed signals of these compressively sampled data. In [45], a method using a genetic programming (GP) algorithm for feature extraction was used, and then ANN and SVM were employed to classify bearing health conditions. In [46], a hybrid model comprising the fuzzy min–max (FMM) neural network and random forest (RF) with sample entropy (SampEn) and power spectrum (PS) features was utilized to classify bearing health conditions. In [37], a three-stage hybrid method consisting of CS, PCA, LDA, and canonical correlation analysis (CCA) was used for bearing fault classification from (1) the whole 6000 samples from the frequency domain, and (2) compressively sampled data with $\alpha = 0.1$ and $\alpha = 0.2$. In [38], a framework combining CS and feature ranking techniques including fisher score, Laplacian score, Relief-F, Pearson correlation coefficients, and Chi-square (Chi-2) were used for bearing fault classification from compressively sampled vibration data with 0.1, feature dimension of 120. Then, with these features, the MLR classifier was used to classify bearing faults. In [47], a three-stage method combining CS (with $\alpha = 0.1, 0.2$, and 0.3), a feature selection procedure, and SVM was used for bearing fault classification.

Table 5. A comparison with the classification results from the literature on the vibration bearing dataset of the first case study.

Ref	Method		Testing Accuracy	Testing Time
[33]	Raw vibration with entropic features + SVM		98.9 ± 1.2	
	Compressed sampled with $\alpha = 0.5$ followed by signal reconstruction + SVM		92.4 ± 0.5	–
	Compressed sampled with $\alpha = 0.25$ followed by signal reconstruction + SVM		84.6 ± 0.41	
[44]	GP generated feature sets (un-normalised data)			
		ANN	96.5	
		SVM	97.1	
[45]	FMM-RF:	SamEn	99.7 ± 0.02	
		PS	99.7 ± 0.50	–
		SamEn + PS	99.8 ± 0.41	
[36]	CPDC (with 6000 inputs from FFT)		99.4 ± 0.5	64.9
	CS-CPDC $\alpha = 0.1$		99.8 ± 0.2	6.7
	$\alpha = 0.2$		99.9 ± 0.1	7.8
[37]	With FFT, α=0.1, feature dimension = 120, and LRC classifier)			
	CS-FS		99.7 ± 0.4	
	CS-LS		99.5 ± 0.3	
	CS-Relief-F		99.8 ± 0.2	–
	CS-PCC		99.8 ± 0.3	
	CS-Chi-2		99.5 ± 0.5	
[46]	Feature selection (with λ=0.004, tolerance value = 0.02) from compressively sampled data and SVM for fault classification:			
	$\alpha = 0.1$ and feature dimension = 14		98.8 ± 2.4	
	$\alpha = 0.2$ and feature dimension = 13		99.9 ± 0.2	–
	$\alpha = 0.3$ and feature dimension = 26		99.9 ± 0.1	
	Our proposed method with $\lambda = 0.003$, NCA tolerance value = 0.01, $\alpha = 0.1$, and feature dimension = 8:			
		MLR classifier	99.5 ± 0.6	0.015
		SVM classifier	99.5 ± 0.5	0.060
	Our proposed method with $\lambda = 0.003$, NCA tolerance value = 0.01, $\alpha = 0.2$, and feature dimension = 10:			
		MLR classifier	99.7 ± 0.3	0.003
		SVM classifier	99.8 ± 0.2	0.040
	Our proposed method with $\lambda = 0.003$, NCA tolerance value = 0.01, $\alpha = 0.3$, feature dimension = 8:			
		MLR classifier	100 ± 0.0	0.003
		SVM classifier	100 ± 0.0	0.030

As Table 5 shows, the classification results of our proposed method are better than those reported in [34,45]. Moreover, our results are the same as, if not better than the classification results described in [37,38,46,47]. Our method is extremely fast and needs only 0.015 s to complete the fault classification task compared to the method in [37], which needs 6.7 s using a classification model trained with 10% of the whole data. Furthermore, results from our proposed method remain as good as, if not well improved than the results stated in [38], although we used limited features (only 8 features), which are not met by the method in [38], which used 120 features.

For further verification of the efficacy of the proposed method, we conducted three experiments using our proposed method using the same settings used to perform our experiments with $\alpha = 0.1, 0.2,$ and 0.3. Then, we employed SVM in place of the MLR classifier in our proposed method to classify bearing health conditions to examine the speed

and accuracy performance of our proposed method compared to the method used in [47]. The results are presented in the last row of Table 5. The classification results of our method with MLR classifier are as good as, if not better than the results of our method with SVM. Interestingly, the results demonstrate that our method with MLR classifier is faster and requires only 25%, 13.3%, and 10% of the time of our method with SVM to complete the classification task using classification models trained with $\alpha = 0.1$, 0.2, and 0.3, respectively.

3.2. Second Case Study

The bearing datasets used in this case are provided by Case Western Reserve University (https://engineering.case.edu/bearingdatacenter/download-data-file, accessed on 2 April 2022). The bearing datasets were obtained from a motor driving mechanical system where the faults were planted into the drive-end bearing of the motor under four different speeds and several health conditions, namely, normal condition (NO), with an IR fault (IR), with a roller element fault (RE), and with an OR fault (OR). Then, the datasets were further categorized by the width of the fault (0.18–0.71 mm) and the load of the motor (0–3 hp). The sampling rates {xe "sampling rates"} utilized were 12 kHz for some of the sampled data and 48 kHz for the rest of the sampled data. At each speed, 100 time series were recorded for each condition per load. For the IR, OR, and RE conditions, vibration signals for four different fault widths (0.18, 0.36, 0.53, and 0.71 mm) were separately recorded. In this study, of these acquired vibration signals, two groups of datasets were prepared for the evaluation of our proposed method.

The first group of datasets is selected from the data files of the vibration signals sampled at 48 kHz with fault width (0.18, 0.36, and 0.53 mm) and fixed loads including 1, 2, and 3 hp, and the number of examples chosen is 200 examples per condition. This offered three different datasets A, B, and C with 2000 total examples and 2400 data points for each signal. The second type of bearing dataset was chosen from the data files of vibration signals sampled at 12 kHz with fault size (0.18, 0.36, 0.53, and 0.71 mm) and load 2 hp, and the number of examples chosen was 60 examples per condition. This offered a dataset D with 720 total examples and 2000 data points for each signal. The description of the used bearing vibration dataset is presented in Table 6.

To classify the bearing's health conditions from datasets A, B, C, and D described above, the same steps as in the first case study were followed to apply our proposed method. First, 50% of the total observations of datasets A, B, C, and D were randomly chosen for training and the other 50% is employed for testing the trained model. Then, we obtained the compressively sampled vibration signal from the high dimensional datasets A, B, and C, with 2400 time samples for each of the 2000 observations and dataset D with 2000 time samples for each of the 720 observations. The FFT coefficients were employed as sparse representations of all the datasets used in the second case study, i.e., A, B, C, and D. Then, we adopted the CS mechanism with different sampling rates (α) of 0.1, 0.2, and 0.3, using a random Gaussian matrix as described in the first case study to obtain the compressively sampled signals for each dataset.

Table 6. Description of the bearing health conditions of the bearing vibration dataset used in the second case study.

Health Condition	Fault 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

To estimate and reduce further the dimensionality of the compressively sampled signals, we applied the same steps of the feature selection process step of our proposed method as described in the first case study. Two tolerance values (0.01 and 0.02) were investigated for the feature selection process. Table 7 presents **e**xamples of the computed values of the average intrinsic dimension, the dimension of the NCA-based selected features, least loss, and best λ values taken from 10 for each dataset. To test the impact of the tolerance values and the compressive sampling rates on the number of the selected features using our proposed method. Table 7 shows the computed values of the average intrinsic dimension *(d)*, the dimension of the NCA-based selected features *(f)*, the best least loss, and best λ values taken from 10 trials for datasets A, B, C, and D. Moreover, it can be seen from the data in Table 7 that all the computed average least lost values are exceedingly small in the range of 0.000–0.003 although the feature dimension was reduced to a fewer number of features, e.g., for dataset A the feature dimension were reduced from 2400 (in the original raw vibration signals) to 240 (the dimension of the compressively sampled signals with $\alpha = 0.1$), which reduced to 13 (*d-dimension*) and then further reduced to 6 (*f-dimension with 0.01 tolerance value*).

Moreover, the computed average best lambda values for the NCA algorithm are in the range of 0.0002–0.0041. The NCA tolerance values have a clear effect on the computed intrinsic dimension (*d*) and the dimension of the NCA-based selected features. For example, for dataset A with a 0.02 tolerance value, we obtained an intrinsic dimension (*d*) of 18, 21, and 25, which reduced to 13, 15, and 15 with 0.01 tolerance values for $\alpha = 0.1$, $\alpha = 0.2$, and $\alpha = 0.3$, respectively. Furthermore, for dataset A with the tolerance value of 0.02, we obtained the final feature dimension (*f*) of 10, 12, and 14, which reduced to 6, 7, 9 with 0.01 tolerance for $\alpha = 0.1$, $\alpha = 0.2$, and $\alpha = 0.3$, respectively. It can therefore be assumed that when the tolerance value decreases the dimension of both selected features, i.e., *d* and *f*, decreases. Figure 9 shows an example of the average loss values versus λ values. Figure 10 presents an example of the selected features and their corresponding weights.

Table 7. Examples of the computed values of the average intrinsic dimension, the dimension of the NCA-based selected features, least loss, and best λ values taken from 10 trials for datasets A, B, C, and D.

Dataset	NCA Tolerance Value	CS Sampling Rate (α)	*d*	*f*	Average Least Loss	Average Best Lambda for NCA
A	0.01	0.1	13	6	0.001	0.0011
		0.2	15	7	0.000	0.0009
		0.3	15	9	0.001	0.0006
	0.02	0.1	18	10	0.000	0.0009
		0.2	21	12	0.000	0.0004
		0.3	25	14	0.000	0.0003
B	0.01	0.1	17	7	0.000	0.0007
		0.2	19	9	0.001	0.0006
		0.3	24	9	0.000	0.0002
	0.02	0.1	28	11	0.000	0.0009
		0.2	23	10	0.000	0.0005
		0.3	26	12	0.000	0.0003
C	0.01	0.1	16	5	0.001	0.0010

		0.2	17	7	0.000	0.0006
		0.3	23	8	0.000	0.0006
	0.02	0.1	21	11	0.000	0.0009
		0.2	22	12	0.000	0.0004
		0.3	27	14	0.000	0.0003
	0.01	0.1	15	4	0.003	0.0041
		0.2	18	5	0.001	0.0041
D		0.3	20	7	0.001	0.0015
	0.02	0.1	17	9	0.001	0.003
		0.2	21	10	0.001	0.003
		0.3	23	9	0.001	0.003

The final NCA-based selected features from datasets A, B, C, and D were used to train the MLR classifier to obtain a trained classification model for each dataset to classify among c classes of roller bearing health conditions. Table 8 shows an overview of the testing results for each bearing vibration dataset where the classification accuracy is the average of 10 trials for each experiment, and the time obtained by averaging the testing time of these 10 trials. One of the more significant findings to emerge from the results in Table 8 is that the classification results with $\alpha = 0.2$ and 0.3 and tolerance values = 0.01 and 0.02 for all datasets A, B, C, and D are all over 99%. For datasets B and C with $\alpha = 0.3$ and tolerance values of 0.01 and 0.02, our proposed method attained 100% classification accuracy.
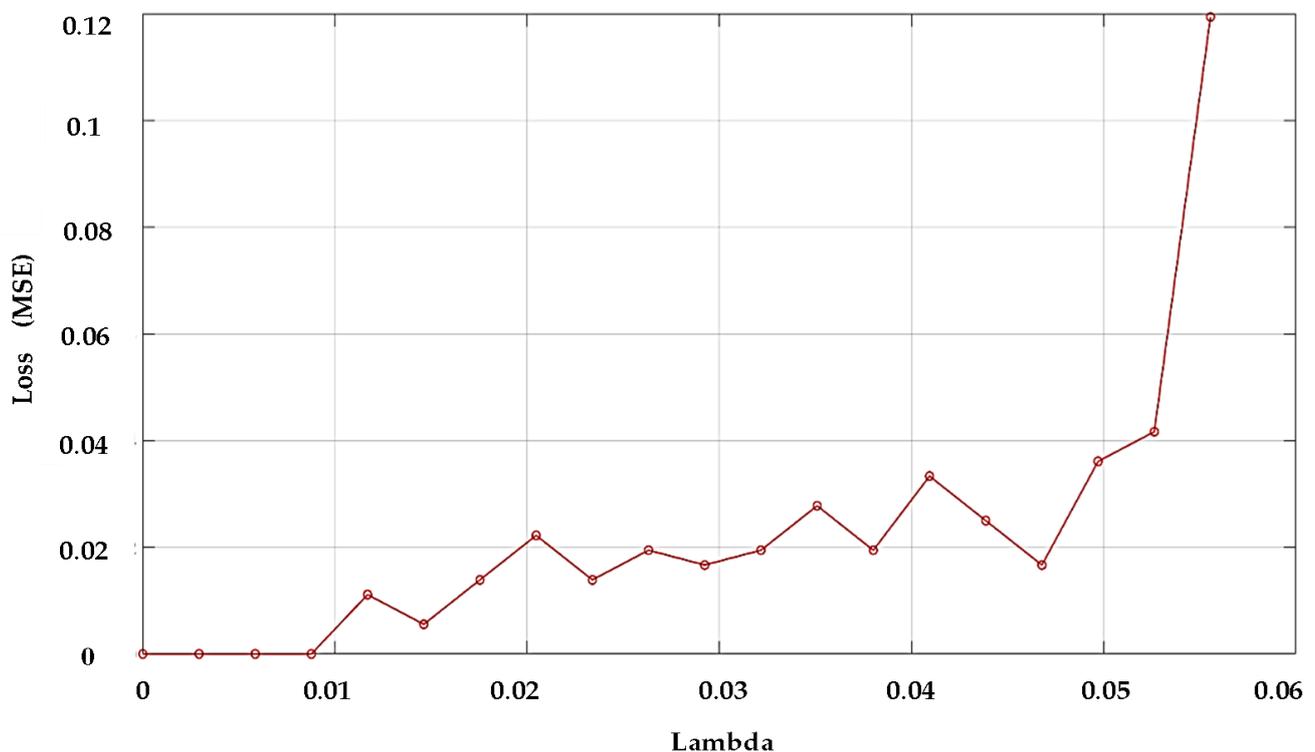
Figure 9. Example of the average loss values versus λ values computed from the reduced dimension of compressively sampled data with $\alpha = 0.2$ and dataset A.
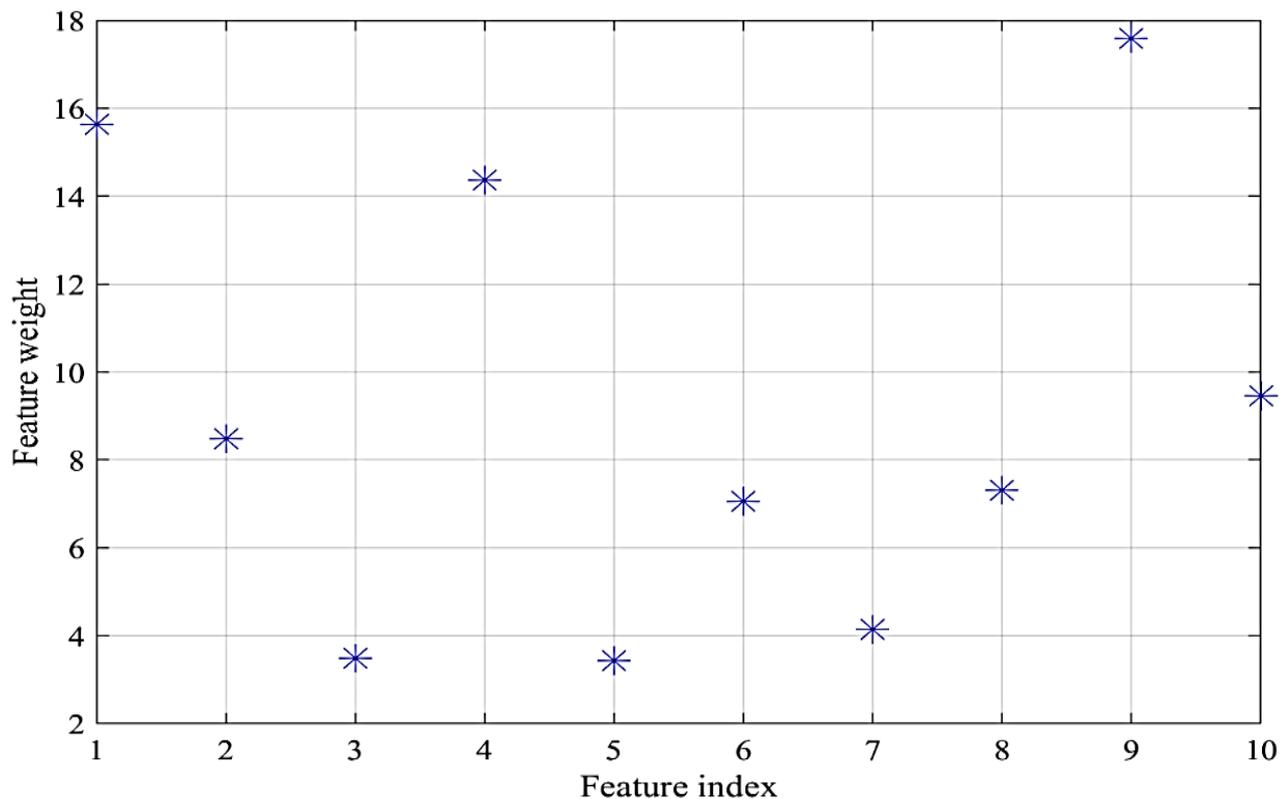
Figure 10. Example of the selected features from dataset A and their corresponding weights using $\alpha = 0.2$ and NCA tolerance value = 0.02.

Similarly, for dataset A with $\alpha = 0.3$ and tolerance value of 0.02, our proposed method achieved 100% classification accuracy for every single run in our experiments. Furthermore, the trained classification models of our proposed method for all bearing vibration datasets A, B, C, D require less than 0.01 s to complete the classification task. These findings suggest that our proposed method is fast and offers high classification accuracies for rolling bearings from vibration datasets under different load levels as in A, B, and C.

Table 8. Classification results with their corresponding RMSE and computational time for the automatically selected features (*d* refers to the intrinsic dimension, and *f* is the dimension of the NCA-based selected feature) using two values of the NCA tolerance and compressive sampling rates for datasets A, B, C, and D (all classification accuracies of 100% are in bold).

Dataset	NCA Tolerance Value	CS Sampling Rate (α)	d	f	MLR Classifier	
					Testing Accuracy (%)	Testing Time (s)
A	0.01	0.1	13	6	98.5 ± 0.7	0.002
		0.2	15	7	99.9 ± 0.1	0.003
		0.3	15	9	99.9 ± 0.1	0.002
	0.02	0.1	18	10	99.6 ± 0.2	0.003
		0.2	21	12	99.8 ± 0.2	0.006
		0.3	25	14	**100 ± 0.0**	0.003
B	0.01	0.1	17	7	99.2 ± 0.7	0.002
		0.2	19	9	99.5 ± 0.5	0.003
		0.3	24	9	**100 ± 0.0**	0.009
	0.02	0.1	28	11	99.9 ± 0.1	0.003
		0.2	23	10	99.9 ± 0.1	0.006

		0.3	26	12	**100 ± 0.0**	0.003
		0.1	16	5	99.7 ± 0.2	0.002
	0.01	0.2	17	7	99.9 ± 0.1	0.003
C		0.3	23	8	**100 ± 0.0**	0.002
		0.1	22	11	99.9 ± 0.1	0.003
	0.02	0.2	27	12	99.9 ± 0.1	0.006
		0.3	15	14	**100 ± 0.0**	0.003
		0.1	18	4	92.7 ± 2.9	0.002
	0.01	0.2	20	5	99.1 ± 0.8	0.002
D		0.3	17	7	99.9 ± 0.1	0.002
		0.1	21	9	99.9 ± 0.1	0.002
	0.02	0.2	23	10	99.9 ± 0.1	0.002
		0.3	13	9	99.9 ± 0.1	0.002

Table 9 presents sample confusion matrices of the classification results of MLR classifier using selected features with tolerance value = 0.01 and a sampling rate of (a) α = 0.1, 0.2, and 0.3 with dataset A. As can be seen from Table 9a with 10% testing data our method misclassified three of the testing examples of condition 6, i.e., IR2, as condition 1, i.e., NO, condition 3, i.e., RE2, and condition nine, i.e., OR2, respectively. In Table 9b, with 20% testing data, our method misclassified three of the testing examples of condition 6, i.e., IR2, as condition 7, i.e., IR3. Moreover, in Table 9c the recognition of all bearing health conditions is 100%.

Table 9. Sample confusion matrices of the classification results of MLR classifier using selected features with tolerance value = 0.01 and a sampling rate of (a) α = 0.1, (b) α = 0.2, and (c) α = 0.3 with the dataset A.

	NO	RE1	RE2	RE3	IR1	IR2	IR3	OR1	OR2	OR3
NO	100	0	0	0	0	0	0	0	0	0
RE1	0	100	0	0	0	0	0	0	0	0
RE2	0	0	100	0	0	0	0	0	0	0
RE3	0	0	0	100	0	0	0	0	0	0
IR1	0	0	0	0	100	0	0	0	0	0
IR2	1	0	1	0	0	97	0	0	1	0
IR3	0	0	0	0	0	0	100	0	0	0
OR1	0	0	0	0	0	0	0	100	0	0
OR2	0	0	0	0	0	0	0	0	100	0
OR3	0	0	0	0	0	0	1	0	0	99

(a)

	NO	RE1	RE2	RE3	IR1	IR2	IR3	OR1	OR2	OR3
NO	100	0	0	0	0	0	0	0	0	0
RE1	0	100	0	0	0	0	0	0	0	0
RE2	0	0	100	0	0	0	0	0	0	0
RE3	0	0	0	100	0	0	0	0	0	0
IR1	0	0	0	0	100	0	0	0	0	0
IR2	0	0	0	0	0	97	3	0	0	0
IR3	0	0	0	0	0	0	100	0	0	0
OR1	0	0	0	0	0	0	0	100	0	0
OR2	0	0	0	0	0	0	0	0	100	0
OR3	0	0	0	0	0	0	0	0	0	100

(b)

	NO	RE1	RE2	RE3	IR1	IR2	IR3	OR1	OR2	OR3
NO	100	0	0	0	0	0	0	0	0	0
RE1	0	100	0	0	0	0	0	0	0	0
RE2	0	0	100	0	0	0	0	0	0	0
RE3	0	0	0	100	0	0	0	0	0	0
IR1	0	0	0	0	100	0	0	0	0	0
IR2	0	0	0	0	0	100	0	0	0	0
IR3	0	0	0	0	0	0	100	0	0	0
OR1	0	0	0	0	0	0	0	100	0	0
OR2	0	0	0	0	0	0	0	0	100	0
OR3	0	0	0	0	0	0	0	0	0	100

(c)

Comparisons of Results

For additional assessment of the efficiency of the proposed method, Table 10 shows comparisons with some recently published results using the same bearing vibration datasets used in the second case study. In [4], a CS-DNN technique that involves a deep neural network method (DNN) with two hidden layers method combined with the Haar Wavelet-based CS technique was used to classify rolling bearing from the same rolling bearing datasets A, B, and C with $\alpha = 0.1$. In [38], a framework combining CS and feature ranking techniques including Fisher score, Laplacian score, Relief-F, Pearson correlation coefficients, and Chi-square (Chi-2) were used for bearing fault classification from the dataset D with $\alpha = 0.1$ and a feature dimension of 120. With these features, the MLR classifier was used to classify bearing faults. In [48], several methods were used to classify bearing faults with the same roller bearing the dataset D used in the second study. One of the methods used feature selection by adjunct rand index and standard deviation ratio (FSAR) the original feature set (OFS). Some of the other techniques utilized PCA, LDA, LFDA, and support margin LFDA (SM-LFDA). The selected features of these methods were used to train SVM to be used for bearing fault classification. Moreover, classification results for bearing fault classification using two methods with datasets A, B, and C are reported in [49]. The first method is based on a deep neural network (DNN) and the second method is a method based on a backpropagation neural network (BPNN). Additionally, classification results using a generic multi-layer perceptron (MLP) method with datasets A, B, and C are reported in [50].

As shown in Table 10, the results from our proposed method are better than those reported in [38,48,50]. Additionally, our classification results are better than the results produced using BPNN in [49]. Moreover, our results are the same as, if not better than the results reported in [4] and the results obtained using DNN in [49]. Remarkably, the results show that our method is faster than the CS-DNN technique that used in [4] as our method requires less than 0.005 s while the CS-DNN requires at least 5.7 s to complete the classification task.

In summary, the high reduction in the computation time originates from two reasons—(i) using CS that allow us to use a smaller sampling rate as in $\alpha = 0.1, 0.2,$ and 0.3, and (ii) selecting far fewer features to be used for training the classification algorithm and for classifying rolling bearing health conditions using the trained classification model. Finally, our proposed method achieves classification results for all the rolling bearing vibration datasets A, B, C, and D that are the same as, if not more improved than, fault classification results from the literature on the same vibration bearing datasets.

Table 10. A comparison with the classification results from the literature on the vibration bearing datasets A, B, C, and D of the second case study.

Ref	Dataset	Method	Testing Accuracy (%)	Testing Time (s)
[4]	A	CS-DNN with α=0.1	99.3 ± 0.6	5.7
	B		99.7 ± 0.5	5.9
	C		100 ± 0.0	5.7
[37]	D	With FFT, α=0.1, feature dimension=120, and LRC		–
		CS-FS	98.4 ± 1.6	
		CS-LS	99.1 ± 0.8	
		CS-Relief-F	99.3 ± 0.6	
		CS-PCC	99.2 ± 0.8	
		CS-Chi-2	97.5 ± 2.6	
[47]		OFS-FSAR-SVM		–
		Selected features = 25	91.46	
		Selected features = 50	69.58	
		OFS-FSAR-PCA-SVM		
		Selected features = 25	91.67	
		Selected features = 50	69.79	
		OFS-FSAR-LDA-SVM		
		Selected features = 25	86.25	
		Selected features = 50	92.70	
		OFS-FSAR-LFDA-SVM		
		Selected features = 25	93.75	
		Selected features = 50	94.38	
		OFS-FSAR-(SM-LFDA)-SVM		
		Selected features = 25	94.58	
	D	Selected features = 50	95.63	
[48]	A	DNN	99.95 ± 0.06	–
	B		99.61 ± 0.21	
	C		99.74 ± 0.16	
	A	BPNN	62.20 ± 18.09	
	B		61.95 ± 22.09	
	C		69.82 ± 17.67	
[49]	A	MLP		–
	B			
	C			
The proposed method		Our proposed method with λ=0.003, NCA tolerance value = 0.02, and α=0.1.		
	A	feature dimension= 10	99.6 ± 0.2	0.003
	B	feature dimension= 11	99.9 ± 0.1	0.003
	C	feature dimension= 11	99.9 ± 0.1	0.003
	D	feature dimension= 9	99.9 ± 0.1	0.002

4. Conclusions

The purpose of the present research was to examine the classification of bearing health conditions with far fewer selected features of compressively sampled vibration signals to achieve highly reduced computation time and yet to achieve high classification accuracy. The proposed method comprises a CS-based technique, which was used to obtain compressed vibration signals, followed by an intrinsic dimension estimation-based

feature selection process that includes SPE-based feature selection with a self-organizing iterative scheme to embed the compressed data dimension into a further lower dimension, and a non-parametric NCA-based feature selection that maximize the stochastic variant of the leave-one-out nearest neighbour score to achieve the best classification accuracy on the training set. This ensures selecting fewer features from the high dimensional data capable of achieving high classification accuracy and a reduced computation time.

The multinomial logistic regression (MLR) algorithm was used to classify bearing faults. Two fault classification cases of rolling bearings vibration signals under different working loads were used to test the proposed method. The first set of analyses inspected the impact of the compressive sampling rate and the tolerance values on the number of the selected features The experimental results of bearing faults classification demonstrated that the proposed method could obtain higher classification accuracy and higher reduction in the computational time. The higher reduction in the computation time originates from two causes, (i) using CS that allows us to use a smaller sampling rate as in α = 0.1, 0.2, and 0.3, and (ii) selecting far fewer features to be used for training the classification algorithm and for classifying rolling bearing health conditions using the trained classification model. Finally, our proposed method achieves classification results for all the rolling bearing vibration datasets A, B, C, and D that are as good as, if not better than, classification results from the literature on the same vibration bearing datasets.

References

1. Ahmed, H.; Nandi, A.K. *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*; John Wiley & Sons: Hoboken, NJ, USA, 2020.
2. Immovilli, F.; Bellini, A.; Rubini, R.; Tassoni, C. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Trans. Ind. Appl.* **2010**, *46*, 1350–1359. https://doi.org/10.1109/TIA.2010.2049623.
3. Randall, R.B. *Vibration-Based Condition Monitoring: Industrial, Aerospace, and Automotive Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
4. Ahmed, H.O.A.; Wong, M.L.D.; Nandi, A.K. Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features. *Mech. Syst. Signal Process.* **2018**, *99*, 459–477. https://doi.org/10.1016/j.ymssp.2017.06.027.
5. Nandi, S.; Toliyat, H.A.; Li, X. Condition monitoring and fault diagnosis of electrical motors—A review. *IEEE Trans. Energy Convers.* **2005**, *20*, 719–729. https://doi.org/10.1109/TEC.2005.847955.
6. Ahmed, H. Intelligent Methods for Condition Monitoring of Rolling Bearings Using Vibration Data. Ph.D. Thesis, Brunel University London, London, UK, 2019.
7. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. https://doi.org/10.1016/j.triboint.2015.12.037.
8. McCormick, A.C.; Nandi, A.K. A comparison of artificial neural networks and other statistical methods for rotating machine condition classification. In Proceedings of the IEE Colloquium on Modeling and Signal Processing for Fault Diagnosis (Digest No: 1996/260), Leicester, UK, 18 September 1996; Volume 260, pp. 2/1–2/6. https://doi.org/10.1049/ic:19961372.
9. Jack, L.B.; Nandi, A.K. Genetic algorithms for feature selection in machine condition monitoring with vibration signals. *IEE Proc.-Vis. Image Signal Process.* **2000**, *147*, 205–212.

10. Jack, L.B.; Nandi, A.K. Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mech. Syst. Signal Process.* **2002**, *16*, 373–390.

11. Sreejith, B.; Verma, A.K.; Srividya, A. Fault diagnosis of rolling element bearing using time-domain features and neural networks. In Proceedings of the 2008 IEEE Region 10 and the Third international Conference on Industrial and Information Systems, Kharagpur, India, 8–10 December 2008; pp. 1–6. https://doi.org/10.1109/ICIINFS.2008.4798444.

12. Prieto, M.D.; Cirrincione, G.; Espinosa, A.G.; Ortega, J.A.; Henao, H. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *IEEE Trans. Ind. Electron.* **2012**, *60*, 3398–3407. https://doi.org/10.1109/TIE.2012.2219838.

13. McCormick, A.C.; Nandi, A.K. Real-time classification of rotating shaft loading conditions using artificial neural networks. *IEEE Trans. Neural Netw.* **1997**, *8*, 748–757. https://doi.org/10.1109/72.572110.

14. Li, B.; Goddu, G.; Chow, M.Y. Detection of common motor bearing faults using frequency-domain vibration signals and a neural network-based approach. In Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No.98CH36207), Philadelphia, PA, USA, 26–26 June 1998; Volume 4, pp. 2032–2036. https://doi.org/10.1109/ACC.1998.702983.

15. Zeng, L.; Wang, H.P. Machine-fault classification: A fuzzy-set approach. *Int. J. Adv. Manuf. Technol.* **1991**, *6*, 83–93.

16. Dhamande, L.S.; Chaudhari, M.B. Bearing fault diagnosis based on statistical feature extraction in time and frequency domain and neural network. *Int. J. Veh. Struct. Syst.* **2017**, *8*, 229–240 https://doi.org/10.4273/ijvss.8.4.09.

17. Helmi, H.; Forouzantabar, A. Rolling bearing fault detection of an electric motor using time domain and frequency domain feature extraction and ANFIS. *IET Electr. Power Appl.* **2019**, *13*, 662–669.

18. Wang, H.; Chen, P. Fuzzy diagnosis method for rotating machinery in variable rotating speed. *IEEE Sens. J.* **2010**, *11*, 23–34. https://doi.org/10.1109/JSEN.2010.2049103.

19. Ali, J.B.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27. https://doi.org/10.1016/j.apacoust.2014.08.016.

20. Djebala, A.; Ouelaa, N.; Hamzaoui, N. Detection of rolling bearing defects using discrete wavelet analysis. *Meccanica* **2008**, *43*, 339–348.

21. He, M.; He, D. Deep learning-based approach for bearing fault diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065. https://doi.org/10.1109/TIA.2017.2661250.

22. Immovilli, F.; Cocconcelli, M.; Bellini, A.; Rubini, R. Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals. *IEEE Trans. Ind. Electron.* **2009**, *56*, 4710–4717. https://doi.org/10.1109/TIE.2009.2025288.

23. Lei, Y.; Lin, J.; He, Z.; Zi, Y. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2011**, *25*, 1738–1749. https://doi.org/10.1016/j.ymssp.2010.12.011.

24. Fan, H.; Shao, S.; Zhang, X.; Wan, X.; Cao, X.; Ma, H. Intelligent fault diagnosis of rolling bearing using FCM clustering of EMD-PWVD vibration images. *IEEE Access* **2020**, *8*, 145194–145206. https://doi.org/10.1109/ACCESS.2020.3012559.

25. Ye, L.; Ma, X.; Wen, C. Rotating Machinery Fault Diagnosis Method by Combining Time-Frequency Domain Features and CNN Knowledge Transfer. *Sensors* **2021**, *21*, 8168. https://doi.org/10.3390/s21248168.

26. Žvokelj, M.; Zupan, S.; Prebil, I. Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis. *Mech. Syst. Signal Process.* **2010**, *24*, 1049–1067. https://doi.org/10.1016/j.ymssp.2009.09.002.

27. Guo, Y.; Na, J.; Li, B.; Fung, R.F. Envelope extraction-based dimension reduction for independent component analysis in fault diagnosis of rolling element bearing. *J. Sound Vib.* **2014**, *333*, 2983–2994. https://doi.org/10.1016/j.jsv.2014.02.038.

28. De Moura, E.P.; Souto, C.R.; Silva, A.A.; Irmao, M.A.S. Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. *Mech. Syst. Signal Process.* **2011**, *25*, 1765–1772. https://doi.org/10.1016/j.ymssp.2010.11.021.

29. Dong, S.; Luo, T.; Zhong, L.; Chen, L.; Xu, X. Fault diagnosis of bearing based on the kernel principal component analysis and optimized k-nearest neighbour model. *J. Low-Freq. Noise Vib. Act. Control* **2017**, *36*, 354–365. https://doi.org/10.1177/1461348417744302.

30. Haroun, S.; Seghir, A.N.; Touati, S. Feature selection for enhancement of bearing fault detection and diagnosis based on a self-organizing map. In *International Conference on Electrical Engineering and Control Applications*; Springer: Cham, Switzerland, 2016; pp. 233–246.

31. Hu, Q.; Si, X.S.; Qin, A.S.; Lv, Y.R.; Zhang, Q.H. Machinery fault diagnosis scheme using redefined dimensionless indicators and mRMR feature selection. *IEEE Access* **2020**, *8*, 40313–40326. https://doi.org/10.1109/ACCESS.2020.2976832.

32. Toma, R.N.; Prosvirin, A.E.; Kim, J.M. Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. *Sensors* **2020**, *20*, 1884. https://doi.org/10.3390/s20071884.

33. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306.

34. Wong, M.L.D.; Zhang, M.; Nandi, A.K. Effects of compressed sensing on the classification of bearing faults with entropic features. In Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015; pp. 2256–2260. https://doi.org/10.1109/EUSIPCO.2015.7362786.

35. Tang, G.; Hou, W.; Wang, H.; Luo, G.; Ma, J. Compressive sensing of roller bearing faults via harmonic detection from under-sampled vibration signals. *Sensors* **2015**, *15*, 25648–25662. https://doi.org/10.3390/s151025648.

36. Xinpeng, Z.; Niaoqing, H.; Zhe, C. A bearing fault detection method base on compressed sensing. In *Engineering Asset Management-Systems, Professional Practices and Certification*; Springer: Cham, Switzerland, 2015; pp. 789–798.

37. Ahmed, H.O.; Nandi, A.K. Three-stage hybrid fault diagnosis for rolling bearings with compressively sampled data and subspace learning techniques. *IEEE Trans. Ind. Electron.* **2018**, *66*, 5516–5524. https://doi.org/10.1109/TIE.2018.2868259.

38. Ahmed, H.O.; Nandi, A.K. Compressive sampling and feature ranking framework for bearing fault classification with vibration signals. *IEEE Access* **2018**, *6*, 44731–44746. https://doi.org/10.1109/ACCESS.2018.2865116.

39. Candes, E.J.; Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies. *IEEE Trans. Inf. Theory* **2006**, *52*, 5406–5425. https://doi.org/10.1109/TIT.2006.885507.

40. Baraniuk, R.G.; Wakin, M.B. Random projections of smooth manifolds. *Found. Comput. Math.* **2009**, *9*, 51–77.

41. Costa, J.A.; Hero, A.O. Geodesic entropic graphs for dimension and entropy estimation in manifold learning. *IEEE Trans. Signal Process.* **2004**, *52*, 2210–2221. https://doi.org/10.1109/TSP.2004.831130.

42. Agrafiotis, D.K.; Xu, H.; Zhu, F.; Bandyopadhyay, D.; Liu, P. Stochastic proximity embedding: Methods and applications. *Mol. Inform.* **2010**, *29*, 758–770.

43. Agrafiotis, D.K. Stochastic proximity embedding. *J. Comput. Chem.* **2003**, *24*, 1215–1221.

44. Yang, W.; Wang, K.; Zuo, W. Neighborhood component feature selection for high-dimensional data. *J. Comput.* **2012**, *7*, 161–168.

45. Guo, H.; Jack, L.B.; Nandi, A.K. Feature generation using genetic programming with application to fault classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2005**, *35*, 2256–2260. https://doi.org/10.1109/TSMCB.2004.841426.

46. Seera, M.; Wong, M.D.; Nandi, A.K. Classification of ball bearing faults using a hybrid intelligent model. *Appl. Soft Comput.* **2017**, *57*, 427–435. https://doi.org/10.1016/j.asoc.2017.04.034.

47. Ahmed, H.; Nandi, A. Three-stage method for rotating machine health condition monitoring using vibration signals. In Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, China, 26–28 October 2018; Volume 26, pp. 285–291. https://doi.org/10.1109/PHM-Chongqing.2018.00055.

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

49. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72*, 303–315.

50. de Almeida, L.F.; Bizarria, J.W.; Bizarria, F.C.; Mathias, M.H. Condition-based monitoring system for rolling element bearing using a generic multi-layer perceptron. *J. Vib. Control* **2015**, *21*, 3456–3464.