Intelligent Condition Monitoring for Rotating Machinery Using Compressively-Sampled Data and Sub-space Learning Techniques

H. O. A. Ahmed[1(✉)] and A. K. Nandi[1,2]

[1] Department of Electronic and Computer Engineering,
Brunel University London, Uxbridge UB8 3PH, UK
{hosameldin.ahmed, asoke.nandi}@brunel.ac.uk
[2] The Key Laboratory of Embedded Systems and Service Computing,
College of Electronic and Information Engineering, Tongji University,
Shanghai, People's Republic of China

Abstract. Rotating machines are widely used in industry. Unforeseen machine failures affect production schedules, product quality, and production costs. Therefore, condition monitoring of rotating machine can play an important role in machine availability. There is a growing number of methods for Machine Condition Monitoring (MCM). Yet, the performance of these methods is limited by the massive amounts of data need to be collected for MCM. This work proposes a computational method which can greatly reduce the high dimensional vibration dataset to a set of compressively-sampled measurements using Compressive Sampling (CS). Then, to learn fewer features from these compressively-sampled measurements we propose an effective multi-step feature learning algorithm that combines the advantages of Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Canonical Correlation Analysis (CCA). Finally, with these learned features, we use multi-class Support Vector Machine (SVM) to classify machine health conditions. Experiments on a roller element bearing fault classification task based on vibration signals are used to evaluate the efficiency of the proposed method. The most obvious finding to emerge from this study is that we are able to achieve high classification accuracy even from highly reduced vibration signal measurements. Moreover, the efficiency of our proposed method outperforms some recently published results. The proposed method offers better accuracy and has lower costs in time and storage requirements.

Keywords: Machine condition monitoring · Compressive sampling
Sub-space learning

1 Introduction

Manufacturing companies in today's global marketplace race, use their best efforts to cut costs and improve product quality to maintain their competitiveness. The health and availability of such rotating machines in industry have a direct effect on production schedules, production quality, and production costs. Unforeseen machine failures may

lead to unexpected machine downtime, accidents, and injuries. Condition Monitoring (CM) of rotating machine can play an important role in addressing these issues by reducing unplanned downtime of machinery, avoiding machine breakdown, and improving reliability and safety. Various components can be monitored in rotating machine, e.g., bearings, shafts, gearboxes, etc. Of these components, rolling element bearings are amongst the most important components to be monitored as their failures may lead to more significant failures in the machine. CM techniques in rotating machinery encompass the practice of monitoring measurable signals, e.g., vibration and acoustic signals that can be used uniquely or in mixture to identify changes in machine condition [1]. Consequently, this allows Condition Based Maintenance (CBM) to be arranged, or may be other actions to be taken to prevent machine breakdowns.

In general, the overall framework of rotating machine CM contains three main steps: (1) data acquisition; (2) signal processing; and (3) machine condition identification, as shown in Fig. 1. In the data acquisition step, vibration signals have been widely used since various characteristic features can be observed from vibration signals. In the second step, the collected signals are analysed by using signal processing techniques to extract features that represent the health condition of the monitored rotating machine. In the third step, an algorithm is used to identify the condition of rotating machine. The acquisition of vibration signals can be done through vibration sensors, e.g., velocity sensors and accelerometers [2].
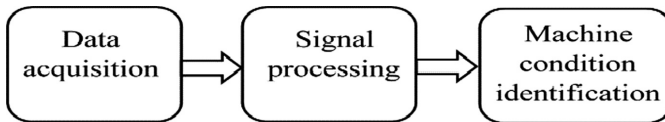

Fig. 1. The overall framework of machine condition monitoring.

The sampling theorems including Shannon-Nyquist theorem are in the core of the current sensing systems. However, Nyquist sampling rate which is at least twice the highest frequency enclosed in the signal is high for some recent emerging applications, e.g., industrial rotating machine, that provide a means of measuring a huge amount of data to be transmitted, stored, and processed. Likewise, some applications that include wideband it is often very costly to collect samples at the necessary rate.

In place of processing the original collected vibration signals, i.e., raw signals, the common methodology is to identify a lower – dimensional features space that can represent the acquired large amount of vibration signals while retaining the important information of the machine conditions. Normally, vibration signals can be analysed in three key sets including, time domain, frequency domain, and time-frequency domain [3, 4]. Time-domain based techniques extract features from the raw vibration signals using several statistical parameters, e.g., Impulse factor, skewness, kurtosis, Crest Factor, root mean square, peak-to-peak value, etc. [5]. The frequency domain based techniques, e.g., baseband auto-spectral density, linear frequency spectrum, and phase-averaged linear spectra, which can be produced by Fourier transforming time series, have the ability to observe substantial diagnosis information based on the frequency

features which are not easily to be observed from time domain series [6]. The time-frequency domain based techniques have been utilised for non-stationary waveform signals that are very common when machinery fault happens. In the literature, several time-frequency domain based techniques have been proposed and applied to vibration based machinery fault diagnosis, e.g., Wavelet Transform (WT), Hilbert-Huang transform (HHT) [7], Short Time Fourier Transform (STFT) [8], Empirical Mode Decomposition (EMD) [9] etc. Furthermore, the application of Spectral Kurtosis and Kurtogram have been investigated and effectively utilised for both vibration signal and acoustic emission in relation to bearing defect identification [10].

As far as the problem of the high dimensionality of the acquired vibration signals is concerned, dimensionality reduction techniques are employed as a data pre-processing stage or as part of the data analysis. Reducing the dimension of vibration signals is useful as it improves the computational efficiency and may develop the accuracy of the analysis. Normally, low-dimensional space of vibration signals can be generated by selecting a subgroup of the original features or by transforming it to a new reduced group of features. For instance, Sakthivel et al. [11] transformed statistical features extracted from vibration signals using several dimensionality reduction techniques including, Principal Component Analysis (PCA), Kernel PCA, Maximum Variance Unfolding (MVU), Local Tangent Space Analysis (LTSA), Diffusion Maps (DM), Laplacian Eigenmaps (LE), and Local Linear Embedding (LLE); then, they are classified using Decision Tree (DT), Bayes Net (BN), and Naïve Bayes (NB) classifiers. Wang et al. [12] proposed a PCA-based technique on defined time-frequency statistical features of rolling bearing vibration signals and faults diagnosis was evaluated using a fuzzy C-means (FCM) model. Widodo et al. [13] studied the application of Independent Component Analysis (ICA) and Support Vector Machines (SVMs) to detect and diagnosis induction motor faults using vibration signals. Ciabattoni et al. [14] presented a novel Linear Discriminant Analysis (LDA) based algorithm to deal with fault vibration data dimension reduction and fault detection problems.

Compressive Sampling (CS) [15] is a new technique that supports sampling below Nyquist rate and shows great possibilities to reconstruct the high dimensional signals from fewer measurements using various signal reconstruction techniques, e.g., Compressive Sampling Matching Pursuit (CoSaMP) [16]. In recent years, several publications have been appeared documenting the use of CS in machine fault diagnosis. For example, Wong et al. [17] demonstrated that it is possible to sample the vibration of roller element bearing at less than Nyquist rate using CS framework and recover the signal for fault classification. In the same way, Li et al. [18] shown the possibility of diagnosis the fault of train's rolling bearing from the reconstructed vibration signal based on CS. Another significant aspect of CS is the use of compressed measurements to diagnosis machine fault without reconstructing the original signal [19–22].

Motivated by the advantages of CS, PCA, LDA, and CCA, this paper proposes a new intelligent fault diagnosis method for rotating machinery. We demonstrate how to create a basis from CS based compressively-sampled data that contains highly correlated potential features of sufficient principal components and discriminative components for fault classification.

2   Compressive Sampling (CS)

The basic idea of CS is that many real-world signals that have sparse representations in some domain, e.g., Fourier Transform (FT), can be recovered from fewer measurements satisfying certain conditions. To obtain a compressed signal $y \in R^{m \times 1}$ from a given collected vibration signal $x \in R^{n \times 1}$ using the standard framework of CS can be explained as following [23]

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

where $s \in R^{n \times 1}$ is a column vector with k nonzero coefficients and represents the sparse components of x. $\psi \in R^{n \times n}$ is the sparsifying transform, e.g., Fast Fourier Transform (FFT) matrix. The compression process can be achieved through a measurement matrix $\phi \in R^{m \times n}$ and a compressive sampling rate (α) where $m = \alpha * n$ and $m \ll n$. The equation that describes the compressed signal $y \in R^{m \times 1}$ is as follows:

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

The measurement matrix ϕ should satisfy the Restricted Isometry Property (RIP), i.e., satisfy the data minimal information loss, to obtain compressed measurements that have the quality of the original signal. Random matrices, e.g., Gaussian random matrix, satisfy the RIP [24]. The possibility to reconstruct the original signal from these compressed samples indicate that the compressed samples possess the quality of the original signal.

Having explained CS framework, the following section will give brief descriptions of PCA, LDA, and CCA algorithms.

3   Subspace Learning Techniques

In terms of subspace learning techniques that used in this study, this section has presented brief descriptions of PCA, LDA, and CCA algorithms.

3.1   PCA

PCA is an orthogonal linear feature projection algorithm that aims to find all the components (eigenvectors) in descending order of significance. The procedure of PCA involves the following steps.

- Calculate the mean vector of the data.
- Compute the covariance matrix of the data.
- Obtain the eigenvalues and eigenvectors of the covariance matrix.

PCA can be employed to form a low-dimensional feature vector [25]. To reduce the dimensionality of the data by means of PCA, one ignores the least significant components from the PCA. Suppose that the input dataset $Y = [y_1, y_2, \ldots, y_L]$ has L

observations and m-dimensional space. PCA transforms Y to $\hat{Y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_L]$ in a new m1-dimensional space that can be represented by the following equation:

$$\hat{Y} = W^T Y \tag{3}$$

here W is projection matrix in which each column vector is composed of the corresponding eigenvectors of $m1$ largest eigenvalues ($m1 \ll m$) of the covariance matrix C that can be computed as follows

$$C = \frac{1}{L} \sum_{i=1}^{L} (y_i - \bar{y})(y_i - \bar{y})^T \tag{4}$$

where

$$\bar{y} = \frac{1}{L} \sum_{i=1}^{L} y_i \tag{5}$$

3.2   Fisher LDA

Unlike PCA, searching for the most significant components of samples, Fisher LDA [26] aims to discover discriminant components that distinguish different class samples. In fact, LDA collects the samples from the same class and expand the margin of samples from different classes. This method considers maximizing the Fisher criterion function $J(W)$, i.e., the ratio of the between the class scatter (S_B) to the within class scatter (S_w) such that

$$J(W) = \frac{|W^T S_B W|}{|W^T S_w W|} \tag{6}$$

$$S_B = \frac{1}{L} \sum_{i=1}^{c} l_i (\mu^i - \mu)(\mu^i - \mu)^T \tag{7}$$

$$S_w = \frac{1}{L} \sum_{i=1}^{c} \sum_{j=1}^{l_i} (y_j^i - \mu^i)(y_j^i - \mu^i)^T \tag{8}$$

where L is the total number of observations, c is the number of classes, $y \in R^{L \times m}$ is the training dataset, y_1^i represents the dataset belong to c-th class, l_i is the number of observations of the i-th class, μ^i is the mean vector of class i, and μ is the mean vector of all training dataset. LDA projects the space of the original data onto a $(c - 1) -$ dimension space by finding the optimal projection matrix W that maximizes the $J(W)$ in Eq. (6) such that

$$\hat{W} = \arg \max_{W} J(W) \tag{9}$$

here $\hat{W}$ is composed of the selected eigenvectors $(\hat{w}_1, \ldots, \hat{w}_{m2})$ with the first $m2$ largest eigenvalues $(m2 = c - 1)$.

3.3   CCA

Different from PCA and LDA that encompass only one dataset, CCA is a statistical method for finding linear combinations of two datasets by maximizing their correlation [27]. For example, let $(y_1, y_2) \in R^{m1}, R^{m2}$ be two vectors with covariance $(\sum_{11}, \sum_{22})$ and cross-covariance $(\sum_{12})$. CCA finds linear combinations of y_1 and y_2 vectors $(\acute{w}_1 y_1, \acute{w}_2 y_2)$ that are maximally correlated such that the following objective function is maximised

$$(w_1, w_2) = arg \max_{W_1, W_2} (\acute{w}_1 y_1, \acute{w}_2 y_2) \tag{10}$$

$$= arg \max_{W_1, W_2} \frac{\acute{w}_1 \sum_{12} w_2}{\sqrt{\acute{w}_1 \sum_{11} w_1 \acute{w}_2 \sum_{22} w_2}} \tag{11}$$

where $\sum_{11}, \sum_{22}$, and $\sum_{22}$, can be computed as $\sum_{11} = \frac{1}{m} y_1 \acute{y}_1, \sum_{22} = \frac{1}{m} y_2 \acute{y}_2$, $\sum_{12} = \frac{1}{m} y_1 \acute{y}_2$. Based on the theory of CCA, the objective function in (10) can be rewritten as follows:

$$arg \max_{W_1, W_2} \acute{w}_1 \sum_{12} w_2$$
$$\text{s.t} \quad \acute{w}_1 \sum_{11} w_1 = 1, \acute{w}_2 \sum_{22} w_2 = 1. \tag{12}$$

More details of the mathematical formulation of CCA can be found in [26].

4   The Proposed Method

To classify the condition of rotating machine from collected vibration signals, a novel condition monitoring method is proposed in this paper. The intention of this method is to reduce vibration signals dimensionality as much as possible while retaining appropriate fault information from which a basis of sufficient principal components (PCs) and enough discriminant components (DCs) will be created for fault classification. As shown in Fig. 2, our proposed method employed CS to obtain compressed measurements that possess the quality of the original vibration signals. Then, a multi-step methodology of PCA, LDA, and CCA is utilised to extract further reduced features from the obtained compressed measurements. Finally, with these extracted features SVM is used to classify rotating machine health condition.

As shown in Fig. 3, to obtain compressed measurements using CS framework in (2), we began by obtaining the sparse representation $S \in R^{nx1}$ from the raw vibration signal $x \in R^{nx1}$ using FFT. Then, the compression process is applied using Gaussian random matrix and an appropriate compressive sampling rate (α) to produce the compressed measurement $y \in R^{mx1}$. After the compression process, the compressively-sampled signals were used to create a further reduced features vector of sufficient principal components and enough discriminant components using a proposed multi-step technique.
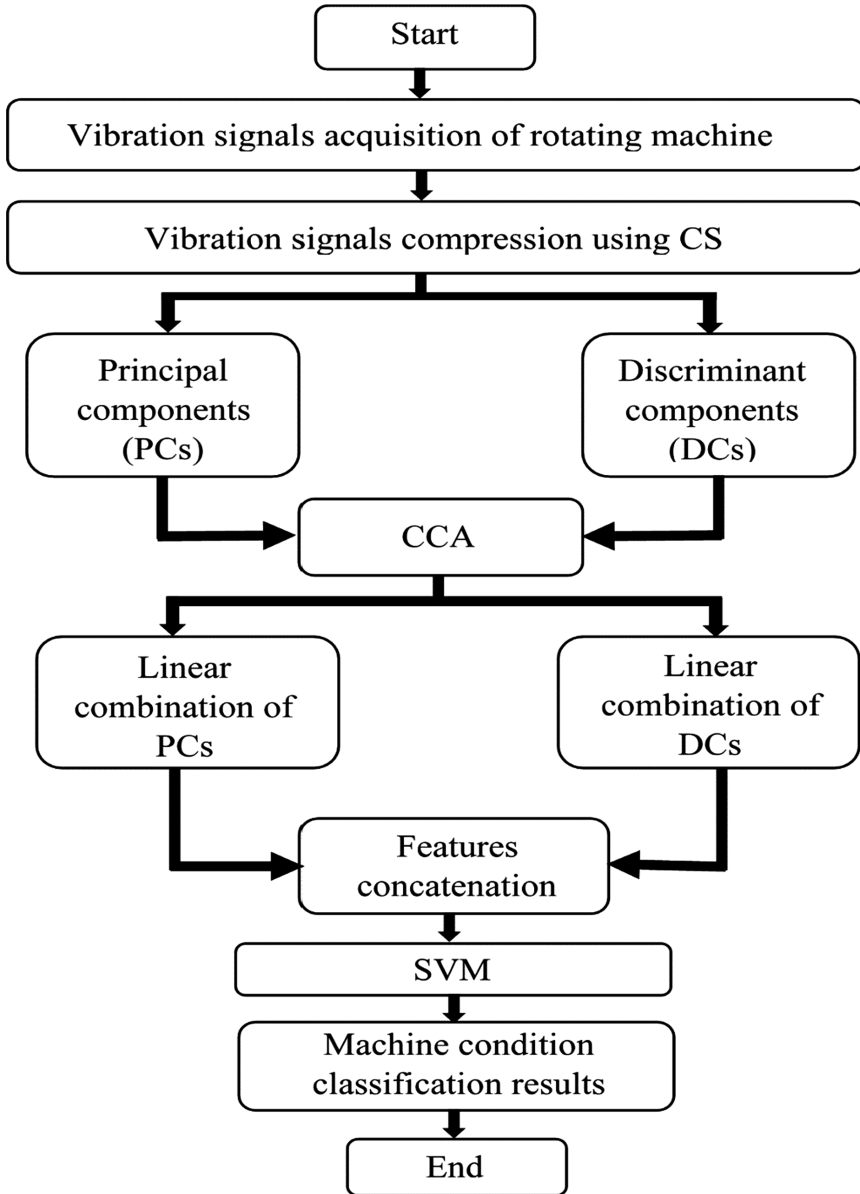
Fig. 2. The proposed method.

In this technique, PCA and LAD are first applied individually to extract subspaces of principal components $\hat{y}_1 \in R^{m1x1}$ and discriminant components $\hat{y}_2 \in R^{m2x1}$ from the compressed samples $y \in R^{mx1}$ where $m1$ is a chosen number of principal components and $m2 = c - 1$; here c is the number of classes associated to each machine condition. Once the PCA and LDA based subspaces were extracted, CCA is used to maximize the

correlation between $\hat{y}_1$ and $\hat{y}_2$ by finding the weighted linear composites w_1 and w_2. The size of w_1 and w_2 are defined by the smaller number of components of $\hat{y}_1$ and $\hat{y}_2$, i.e., $m1$ and $m2$ respectively. The linear combination of $\hat{y}_1 (\hat{y}_{1_{CCA}} = w_1 * \hat{y}_1)$ and $\hat{y}_2 (\hat{y}_{2_{CCA}} = w_2 * \hat{y}_2)$ will maximize their correlation.

Following this process, the learned features $\hat{y}_{1_{CCA}}$ and $\hat{y}_{2_{CCA}}$ are concatenated to obtain a vector that comprise high correlated representations of principal components and discriminative components. Finally, we used multi-class SVM classifier to classify machine condition.

Our proposed method is validated through computer experiments of a fault classification case study of rolling element bearings.
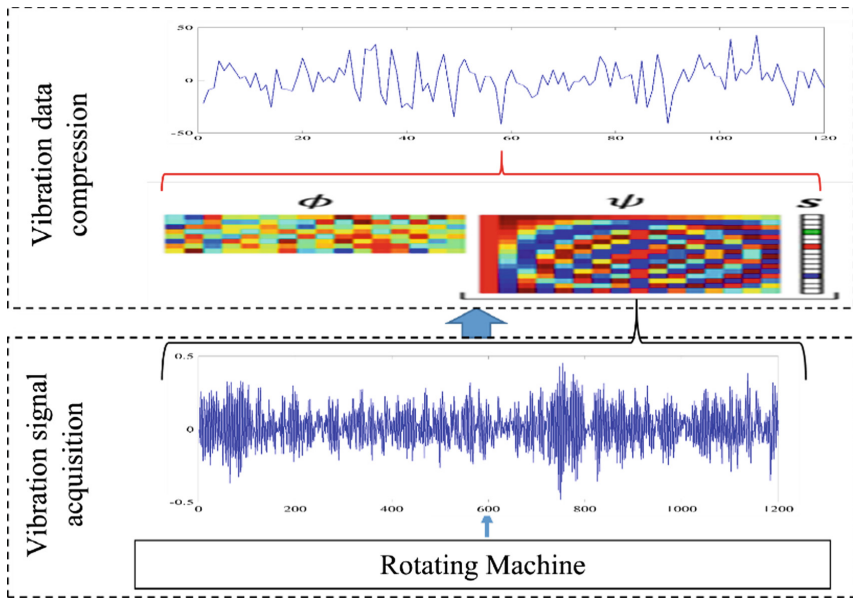

Fig. 3. The compression process of acquired vibration signal using CS framework.

5   Experimental Study

The bearing dataset used in this study is provided by the Case Western Reserve University (CWRU) [28]. The vibration signals were acquired from the drive end of a motor in a test rig (Fig. 4) under normal condition (NO), outer race (OR) fault, inner race (IR) fault, and rolling element (RE) fault. The dataset are further grouped by fault width (0.18, 0.36, and 0.53) under four motor loads (0, 1, 2, and 3 hp) with different shaft speeds (1797, 1772, 1750, 1730 rpm). The sampling rate used to sample the data was 12 kHz. In this study, the motor bearing dataset composed of these vibration signals with 10 bearing health condition and 100 signal examples for each health condition per fault width under four load conditions. Therefore, the total dataset

contains 400 examples for each health condition (4000 signal examples for all health conditions) with 1200 data points for each signal. Figure 5 depicts some typical time series plots for the ten different aforementioned health conditions. Moreover, the explanation of this dataset is presented in Table 1.
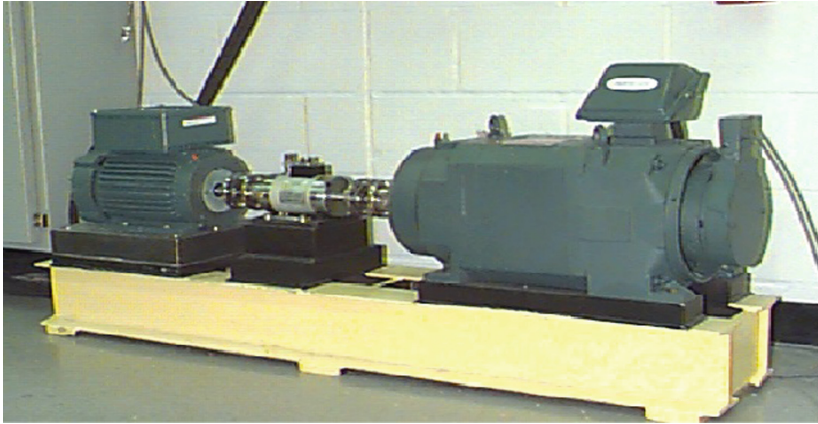

Fig. 4. CWRU bearing test rig [28].

Table 1. Description of bearing health conditions under four loads with different fault width.

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

To classify the motor bearings health condition in this case study, we applied our proposed method to this bearing dataset. Half of the bearing vibration signal examples were selected randomly for training, and the other half of the signal examples were used for testing. To apply our method, we began by obtaining the compressed measurements from the training set with 2000 examples and with 1200 time samples for every example using CS framework. We used FFT basis sparse representations of training signals and random Gaussian matrix as measurement matrix with different values of the compressive sampling rate (α) with 0.05, 0.07, and 0.1.

To learn features from the training set with compressed measurements, we applied our proposed multi-step approach using 9 components for LDA and 10 principal components for PCA for each of the compressed measurements described above. The concatenated features produced from this approach were used to train our classifier. For the classification problem, we employed SVM with "fitcecoc" function [29] on the learned features, i.e., the concatenated features. It uses $c(c-1)/2$ binary SVM models using one-versus-one coding design, where c is the number of unique class labels. This will return a fully trained error-correcting output codes (ECOC) multiclass model that cross-validated using 10-fold cross-validation. The overall classification results and their related root mean square errors of 20 experiments are shown in Table 2.
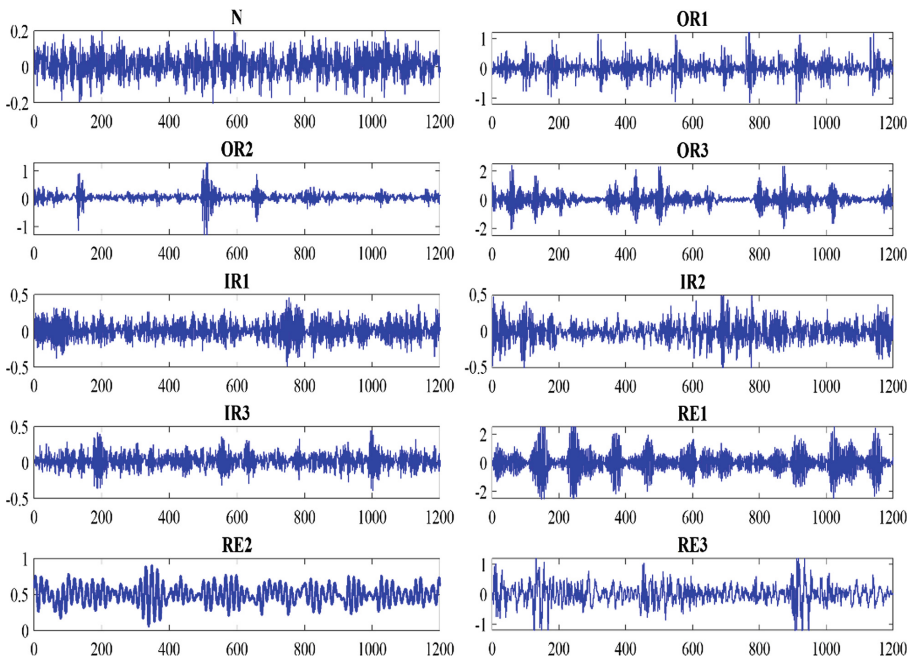

Fig. 5. Typical time series plots for the ten different health conditions.

Table 2. Classification results with their corresponding root mean square (RMSE) and computational time using various compressive sampling rates (α).

Sampling rate (α)	Training accuracy (%)	Training time (s)	Testing accuracy (%)	Testing time (s)
$\alpha = 0.1$ 120 samples	100 ± 0.0	4.83 ± 0.01	99.9 ± 0.1	1.62 ± 0.02
$\alpha = 0.07$ 84 samples	99.9 ± 0.1	4.61 ± 0.01	99.6 ± 0.2	1.41 ± 0.02
$\alpha = 0.05$ 60 samples	99.6 ± 0.2	4.36 ± 0.09	99.2 ± 0.3	1.24 ± 0.07

As Table 2 shows, our proposed method deliver high classification accuracy for each value of α with a small RMSE. Particularly, the testing classification accuracy for α = 0.1 is 99.9%, and the testing time required by our method is only 1.62 s. In general, the computational time increased slightly with the increase in α value. For example, the total time for training and testing with the smallest value of α, i.e., α = 0.05 (5.6 s) increased by less than 20% compared with the total time required by the largest value of α = 0.1 (6.45 s).

Table 3(a) and (b) shows some sample confusion matrices of ten types of a health condition in the classification results for values of α = 0.1 and 0.07 respectively. It can be clearly seen that with α = 0.1 in Table 3(a), only two signal examples of IR1 is likely to be estimated as IR3, i.e., the proposed method misclassified only 1% of testing examples of IR1 as IR3. With α = 0.7 in Table 3(b), only one of IR1 (0.5 of IR1 testing examples) is likely to be confused with IR3, two of IR2 (1% of IR2 testing examples) to be classified as IR1, and five of IR3 (2.5% of testing examples) is expected to be classified as IR1.

Table 3. Sample confusion matrix

(a) Compressively-sampled data α = 0.1

True classes	Predicted classes										Class Prediction (%)
	NO	OR1	OR2	OR3	IR1	IR2	IR3	RE1	RE2	RE3	
NO	200	0	0	0	0	0	0	0	0	0	100
OR1	0	200	0	0	0	0	0	0	0	0	100
OR2	0	0	200	0	0	0	0	0	0	0	100
OR3	0	0	0	200	0	0	0	0	0	0	100
IR1	0	0	0	0	198	0	2	0	0	0	99
IR2	0	0	0	0	0	200	0	0	0	0	100
IR3	0	0	0	0	0	0	200	0	0	0	100
RE1	0	0	0	0	0	0	0	200	0	0	100
RE2	0	0	0	0	0	0	0	0	200	0	100
RE3	0	0	0	0	0	0	0	0	0	200	100

True classes	Predicted classes										Class Prediction (%)
	NO	OR1	OR2	OR3	IR1	IR2	IR3	RE1	RE2	RE3	
NO	200	0	0	0	0	0	0	0	0	0	100
OR1	0	200	0	0	0	0	0	0	0	0	100
OR2	0	0	200	0	0	0	0	0	0	0	100
OR3	0	0	0	200	0	0	0	0	0	0	100
IR1	0	0	0	0	199	0	1	0	0	0	99.5
IR2	0	0	0	0	2	198	0	0	0	0	99
IR3	0	0	0	0	5	0	195	0	0	0	97.5
RE1	0	0	0	0	0	0	0	200	0	0	100
RE2	0	0	0	0	0	0	0	0	200	0	100
RE3	0	0	0	0	0	0	0	0	0	200	100

(b) Compressively-sampled data α = 0.7

For additional assessment of the efficiency of our proposed method, several experiments were conducted for $\alpha = 0.1$ with training size of 10% and 40%, and 20 trials for each experiment. The results of these experiments are compared to some recently published results [30–33]. Table 4 shows the comparisons. The first column denotes to the scenarios of the motor operation and load conditions in which the bearing samples acquired, these scenarios will be stated as fixed load and variable loads. The second column describes the methods utilised to classify bearing conditions. The third column defines the percentage of samples used for training. The fourth column refers to the related load of the data used with each method, and the fifth column presents testing accuracies obtained using these methods.

Table 4. A comparison with results from literature on CWRU vibration dataset of roller bearings

	Methods	Training size (%)	No of classes	Load (hp)	Testing Accuracy (%)
Fixed load	[30]	10	10	3	92.5
	[31]	75	10	0	88.9
	The proposed method	10	10	3	99.8 ± 0.2
	$\alpha = 0.1$			0	99.8 ± 0.1
Variable loads	[32]	40	11	0,1,2,3	97.91 ± 0.09
	[33]	10	10	0,1,2,3	99.66 ± 0.19
	The proposed method	10	10	0,1,2,3	99.8 ± 0.1
	$\alpha = 0.1$	40	10	0,1,2,3	99.9 ± 0.1

It can be clearly seen that compared with methods presented in Table 4, in both load conditions our method with the smallest percentage (10%) of samples of the original samples, i.e., compressed using $\alpha = 0.1$, achieved the highest classification accuracies. In particular, results of 10% training size of data collected under variable loads of 0, 1, 2, and 3 horsepower and separated loads of 0 and 3 horsepower with 99.8% classification accuracy. With 40% training size of data collected under variable loads, the proposed method achieved 99.9% classification accuracy. Taken together, these results show that the proposed method has the ability to classify the bearing conditions with high accuracy compared to other results reported in [30–33].

6   Conclusion

A new method for rotating machinery condition monitoring has been proposed. In this method, CS is used to generate compressively-sampled signals. Then, a multi-step feature learning approach joining the advantages of PCA, LDA, and CCA is proposed to learn further reduced features for fault classification. With these learned features, multi-class SVM classifier is employed to deal with the classification problem. From the experimental results, the proposed method has achieved a high classification accuracy with a significantly reduced dimension of the original signals and its

performance was benchmarked against some existing methods. Moreover, the high classification results achieved in both fixed and variable loads suggest that the proposed method is suitable in rotating machine condition monitoring where the properties of overloads or unexpected load changes may be a reason of rotating machine faults.

References

1. Tavner, P.J.: Review of condition monitoring of rotating electrical machines. IET Electr. Power Appl. **2**, 215–247 (2008)
2. Mechefske, C.K.: Machine condition monitoring and fault diagnostics. In: De Silva, C.W. (ed.) Vibration and Shock Handbook, chap. 25, pp. 25–35. CRC Press, Boca Raton (2005)
3. Nandi, A.K., Liu, C., Wong, M.: Intelligent vibration signal processing for condition monitoring. In: Proceedings of the International Conference Surveillance, vol. 7 (2013)
4. Feng, Z., Liang, M., Chu, F.: Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples. Mech. Syst. Signal Process. **38**(1), 165–205 (2013)
5. Samanta, B., Al-Balushi, K.R.: Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mech. Syst. Signal Process. **17**(2), 317–328 (2003)
6. Norton, M.P., Karczub, D.G.: Fundamentals of Noise and Vibration Analysis for Engineers. Cambridge University Press, Cambridge (2003)
7. Peng, Z.K., Peter, W.T., Chu, F.L.: A comparison study of improved Hilbert-Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. Mech. Syst. Signal Process. **19**(5), 974–988 (2005)
8. Al-Badour, F., Sunar, M., Cheded, L.: Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques. Mech. Syst. Signal Process. **25**(6), 2083–2101 (2011)
9. 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. **89**, 16–27 (2015)
10. Eftekharnejad, B., Carrasco, M.R., Charnley, B., Mba, D.: The application of spectral kurtosis on acoustic emission and vibrations from a defective bearing. Mech. Syst. Signal Process. **25**(1), 266–284 (2011)
11. Sakthivel, N.R., Nair, B.B., Elangovan, M., Sugumaran, V., Saravanmurugan, S.: Comparison of dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibration signals. Eng. Sci. Technol. Int. J. **17**(1), 30–38 (2014)
12. Wang, F., Sun, J., Yan, D., Zhang, S., Cui, L., Xu, Y.: A feature extraction method for fault classification of rolling bearing based on PCA. J. Phys. Conf. Ser. **628**(1), 12079–12085 (2015)
13. Widodo, A., Yang, B.S., Han, T.: Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. Expert Syst. Appl. **32**(2), 299–312 (2007)
14. Ciabattoni, L., Cimini, G., Ferracuti, F., Freddi, A., Ippoliti, G., Monteriú, A.: A novel LDA-based approach for motor bearing fault detection. In: IEEE 13th International Conference on Industrial Informatics (INDIN), pp. 771–776 (2015)
15. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
16. Needell, D., Tropp, J.A.: CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26**(3), 301–321 (2009)

17. Wong, M.L.D., Zhang, M., Nandi, A.K.: Effects of compressed sensing on classification of bearing faults with entropic features. In: IEEE 23rd European Signal Processing Conference (EUSIPCO), pp. 2256–2260 (2015)
18. Li, X.F., Fan, X.C., Jia, L.M.: Compressed sensing technology applied to fault diagnosis of train rolling bearing. Appl. Mech. Mater. **226**, 2056–2061 (2012)
19. Zhang, X., Hu, N., Hu, L., Chen, L., Cheng, Z.: A bearing fault diagnosis method based on the low-dimensional compressed vibration signal. Adv. Mech. Eng. **7**(7), 1–12 (2015)
20. Ahmed, H.O.A., Wong, M. L.D., Nandi, A.K.: Compressive sensing strategy for classification of bearing faults. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2182–2186 (2017)
21. Ahmed, H.O.A., Wong, M. L.D., Nandi, A.K.: Classification of bearing faults combining compressive sampling, Laplacian score, and support vector machine. In: 43rd Annual Conference of the IEEE Industrial Electronics Society, pp. 8053–8058 (2017)
22. 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. **99**, 459–477 (2018)
23. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
24. Baraniuk, R., Davenport, M., DeVore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. Approx. **28**(3), 253–263 (2008)
25. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. **2**(4), 433–459 (2010)
26. Sugiyama, M.: Local Fisher discriminant analysis for supervised dimensionality reduction. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 905–912 (2006)
27. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: An overview with application to learning methods. Neural Comput. **16**(12), 2639–2664 (2004)
28. Case Western Reserve University Bearing Data Center. http://csegroups.case.edu/bearingdatacenter/pages/download-data-file. Accessed 21 Dec 2017
29. Fit multiclass models for support vector machines or other classifiers – MATLAB fitcecoc. Mathworks.com (2017). http://uk.mathworks.com/help/stats-/fitcecoc.html. Accessed 23 Dec 2017
30. Jin, X., Zhao, M., Chow, T.W., Pecht, M.: Motor bearing fault diagnosis using trace ratio linear discriminant analysis. IEEE Trans. Ind. Electron. **61**(5), 2441–2451 (2014)
31. Du, W., Tao, J., Li, Y., Liu, C.: Wavelet leaders multifractal features based fault diagnosis of rotating mechanism. Mech. Syst. Signal Process. **43**(1), 57–75 (2014)
32. Zhang, X., Liang, Y., Zhou, J.: A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. Measurement **69**, 164–179 (2015)
33. Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S.X.: An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans. Ind. Electron. **63**(5), 3137–3147 (2016)