Classification of Bearing Faults Combining Compressive Sampling, Laplacian Score, and Support Vector Machine

Citation for published version:
Ahmed, HOA, Wong, MLD & Nandi, AK 2017, Classification of Bearing Faults Combining Compressive Sampling, Laplacian Score, and Support Vector Machine. in *Proceedings of IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society.* IEEE, pp. 8053-8058, 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 29/10/17. https://doi.org/10.1109/IECON.2017.8217413

Digital Object Identifier (DOI):
10.1109/IECON.2017.8217413

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published In:
Proceedings of IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society

Classification of Bearing Faults Combining Compressive Sampling, Laplacian Score, and Support Vector Machine

*H.O.A. Ahmed[1], M. L. Dennis Wong[2], A. K. Nandi[1,3] **

[1]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, UB8 3PH, United Kingdom

[2]Heriot-Watt University Malaysia, Precinct 5, 62200 Putrajaya.

[3] The Key Laboratory of Embedded Systems and Service Computing, College of Electronic and Information Engineering, Tongji University, Shanghai, P.R. China

***Abstract*— Rolling element bearings have a pivotal role in rotating machine and their failures are the leading cause of more substantial failures in the machine. In response to their importance, there is a growing body of research looking at condition monitoring of rolling element bearings to avoid machine breakdowns. In this study, by taking advantages of Compressive Sampling (CS), Laplacian Score (LS) and Multi-class Support Vector Machine (MSVM), an intelligent method for rolling bearing fault classification is proposed. The CS is used to obtain compressed samples of the raw vibration signals, and the LS is used to rank the features of the obtained compressed samples with respect to their importance and correlations with the core fault characteristics. Then, based on LS ranking, we selected a small amount of the most significant compressed samples to produce the features vector. Finally, classification performance using MSVM shows high classification accuracy with a significantly reduced feature set.**

Keywords— Machine Condition Monitoring; Bearing Fault Classification; Compressive Sensing; Laplacian Score; Multi-Class Support Vector Machine.

I. Introduction

Rolling element bearings are one of the most important components in rotating machinery and their failures are the underlying causes for more substantial failures in the machine. Consequently, rolling element bearings Condition Monitoring [1 – 2] has received considerable attention of many researchers in the past years. Various methods have been proposed for bearing fault detection. Of these methods, vibration based machine condition monitoring (CM) is widely used in practice [3]. The analysis of bearing vibration signal can be done in three main domains of waveform data analysis in order to extract effective features, these include, time domain, frequency domain, and time-frequency domain [4 - 9]. The time-frequency domain has been used for non-stationary waveform signals that are very common when machinery fault happens. Therefore, numerous time-frequency analysis methods have been proposed and used in machinery fault diagnosis, e.g., Short Time Fourier Transform (SFFT), quadratic transforms (e.g., Wigner-Ville distribution), Wavelet Transform (WT), Empirical Mode Decomposition (EMD), Hilbert-Huang Transform (HHT) and Local Mean Decomposition [10 – 13].

Traditionally, vibration based CM systems acquire a big amount of vibration data from rotating machine, that need to be analysed and automatically classify the current bearing status. Learning from big data provide a means of large storage requirements and time for signal processing. To overcome these issues, many researchers have proposed various dimensionality reduction methods of feature extraction and feature selection that have been widely studied in machine fault diagnosis. The aim of these methods is to identify a lower-dimensional data that well represents the original data by keeping the important characteristics of the machine faults.

Of these methods, Principal Component Analysis (PCA) [14], Linear Discriminant Analysis (LDA) [15], Independent Component Analysis (ICA) [16] and Genetic algorithm (GA) [17] are amongst the most common techniques that have been effectively used in machine fault diagnosis. For example, Shuang et al. [18] proposed a method based on PCA and SVM and showed its efficiency in bearing fault diagnosis. Likewise, Wang et al. [19] suggested a PCA-based technique on defined time-frequency statistical features where the effectiveness of the proposed method in rolling bearing faults diagnosis was evaluated using a fuzzy C-means (FCM) model. In the same vein, Jiang et al. [20] proposed a method for condition monitoring for rolling element bearing based on PCA and Phase Space Reconstruction (PSR) that can effectively identify different conditions of rolling element bearings. Giabattoni et al. [21] setup a series of experiments using LDA-based method and show that the proposed algorithm improves the classification accuracy when the classes of motor bearing are overlapped. Widodo et al. [22] developed a method that combined ICA and SVM for fault diagnosis of induction motors. In a similar way, Chang et al. [23] also found a combination of Neural Network (NN) and ICA can achieve considerable classification accuracy of rotating machinery fault diagnosis. Ahmed et al. [24] conducted a series of trials in which Deep Neural Network (DNN) is employed to extract features from vibration signals in

*Corresponding author: Asoke.Nandi@brunel.ac.uk

order to classify bearing faults. A PCA-based approach to select the most representative features for the classification bearing faults was proposed by Malhi, et al. [25] and showed improvements in the classification accuracy for both NN and Radial Basis Function Network (RBFN). Overall, these studies highlight the need for feature extraction and feature selection to improve the efficiency of diagnosis methods and decrease the computational time.

In this study, by taking advantages of CS, LS and MSVM, an intelligent method for rolling bearing fault classification is proposed. CS is applied to generate compressed samples appropriate for bearing fault classification. In this way, we can avoid wasting much time on computing eigenvalues that included in most feature extraction approaches, such as PCA and LDA. Moreover, instead of using all the features in the compressed samples, we employed LS to select a small amount of the features in order to enhance the efficiency of the proposed method.

II. THEORETICAL BACKGROUND

A. Compressive Sampling

Compressive sampling also called compressed sensing or compressed sampling [26, 27] is an extension of sparse representation and special case of it, where only few measurements are available than the whole samples of the signal to be recovered. The basic idea of CS is that many real-world signals have sparse representations in some domain, e.g., Fourier Transform (FT), can be reconstructed from few measurements under specific conditions. Briefly, we present the basic CS framework as follows. Suppose we have the original signal vector (x) where $x \in R^{n\, x\, 1}$ .To get a set of sparse components of x we need to use sparsifying transform ψ such that

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

where s is $n\ x\ 1$ column vector with k nonzero coefficients and represent the sparse elements. According to CS sampling theory, the sparse signal s can be reconstructed from its compressed samples $y \in R^{m\, x\, 1}$ as shown in Fig. 1, where $m << n$ and y can be computed by the following equation

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

where ϕ is the measurement matrix that have to be incoherent with the sparsifying transform ψ and satisfy the data minimal information loss, i.e., satisfy Restricted Isometry Property (RIP).

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

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

Both Random matrix with i.i.d. Gaussian entries and Bernoulli (±1) matrix satisfy RIP. The size of the measurement matrix is ($m\ x\ n$) depends on the compressive sampling rate (α). To estimate the vector s, we need to solve the optimization problem using L1-norm; and the estimate $\hat{s}$ of s can be given by the equation:

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

where $\|\phi\psi s - y\|_{\ell 2}^2 \le \varepsilon$ for a chosen $\varepsilon > o$, and a particular regularization parameter $\gamma > 0$ that controls the relative importance applied to the sparseness ℓ1 and the ℓ2 error. Therefore, the original x can be reconstructed using the inverse of the sparsifying transform and $\hat{s}$ such that

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

This reconstruction indicate that the compressed samples y possess the quality of the original signal.

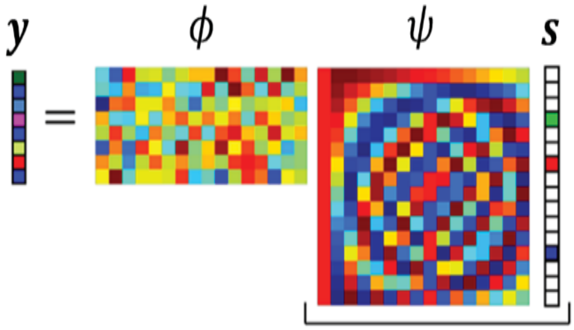


Fig. 1. Compressive sampling framework.

B. Feature Selection using Laplacian Score

Feature selection methods can be categorised into two groups. The wrapper techniques that evaluate the features using a learning algorithm, and the filter techniques examine essential characteristics of the data to evaluate the features. LS is a filter based technique that rank features based on their locality preserving power. In fact, LS is mainly based on Laplacian Eigenmaps and Locality Preserving Projection and can be briefly described as follows [28].

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

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & y_i = y_j \\ 0, & otherwise \end{cases} \tag{6}$$

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

$$L_r = \frac{\tilde{f_r}^T L \tilde{f_r}}{\tilde{f_r}^T D \tilde{f_r}} \tag{7}$$

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

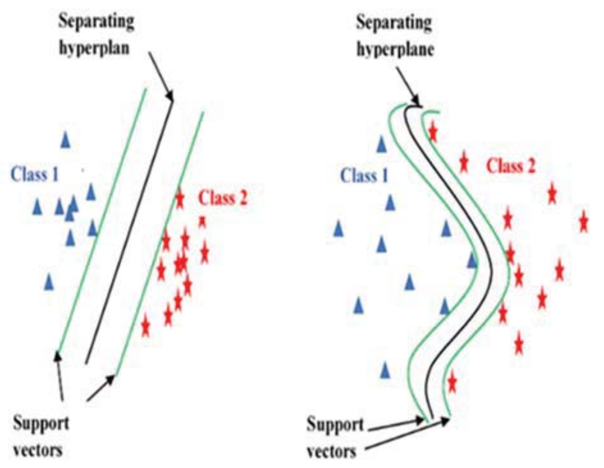
$$\tilde{f_r} = f_r - \frac{f_r^T D\mathbf{1}}{\mathbf{1}^T D\mathbf{1}} \tag{8}$$

More details of mathematical formulation of LS for feature selection can be found in [28].

C. Support Vector Machine

SVM is a supervised machine learning method that firstly proposed for binary classification problem [29]. The basic idea of SVM is that it can find the best hyper plane(s) to separate the two classes. Based on the features of the data, SVM can make linear or non-linear classifications by using different kernel functions, e.g., Radial basis function (RBF), Polynomial function (PF), and Sigmoid function (SF) [30]. This is illustrated in Fig. 2.

For the purpose of multiclass classification problems, i.e., three classes or more, SVM classifiers can be easily combined together to deal with multiclass case. For example, one-against-all and one-against-one methods based on binary classification are widely used in multiclass classification problems. A good comparison of methods for multiclass SVMs can be found in [31].



(a) (b)

Fig. 2. (a) Linear classifier, (b) Non-linear classifier

III. PROPOSED METHOD

Owing to the high dimensionality of vibration signals in bearing faults classification problems, computational efficiency and complexity reduction are very important issues. Based on the above methods, an intelligent bearing fault classification method is developed to classify the condition of bearing from raw vibration signal. As shown in Fig. 3, the proposed method is a combination of CS, LS, and MSVM classifier. The vibration signals collected from the machine are re-sampled using CS that produce compressed samples based on sampling rate (α). Then, the obtained compressed samples are ranked by utilizing LS. Finally, to construct the final feature vector we select the most k significant features that used as the input of the MSVM classifier.

IV. EMPIRICAL VALIDATION

To verify the capability of our proposed method for rolling bearing faults classification, various experiments were performed on bearing vibration data using the proposed method with different compressed data and different number of selected features.

A. Data Description

The vibration data were acquired from experiments on a small test rig which simulates running roller bearings environment. Six conditions of roller bearing status have been measured and tested. These include, two normal conditions, namely, a brand new condition (NO) and a worn but undamaged condition (NW); and four faulty conditions including, inner race (IR), an outer race (OR), rolling element (RE), and cage (CA) fault. Each condition has its corresponding unique characteristics. Data were recorded at 16 different speeds. Fig. 4 shows some typical time series plots for the six different above conditions.

Based on the fault conditions, the faults modulate the vibration signals with their characteristics, for example, the IR and OR conditions have a fairly periodic signals; while RE fault may or may not be periodic, depending on some factors including the level of damage and the loading of the bearing, CA fault generates random distortion that also depends on the degree of damage and the bearing loading.

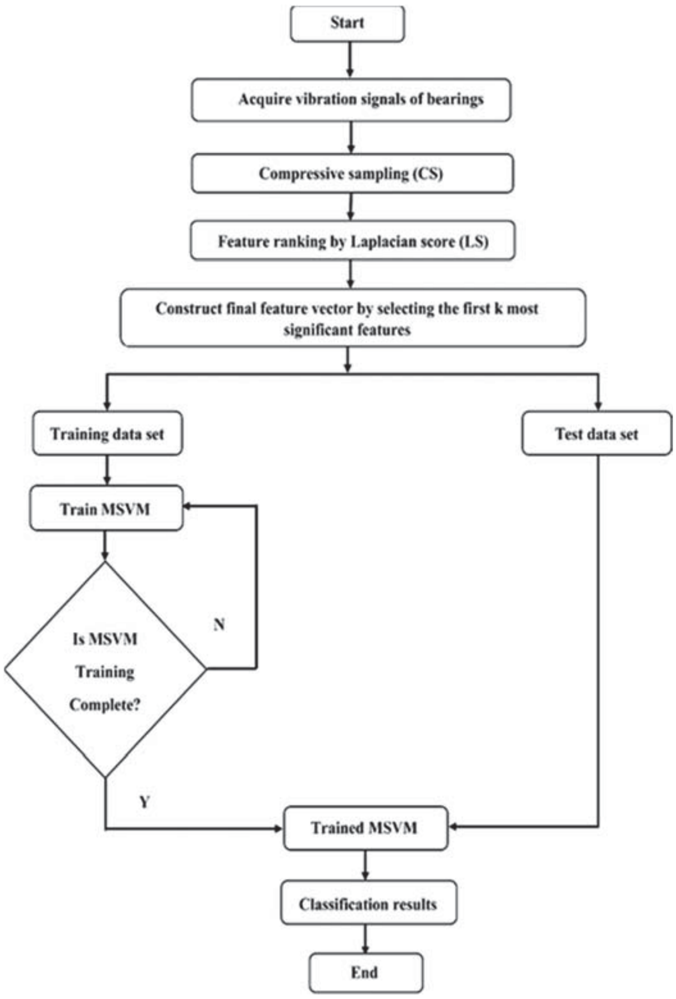


Fig. 3. Proposed method

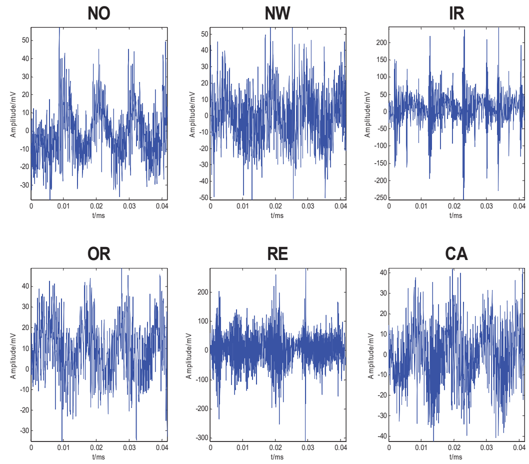


Fig. 4. Characteristic vibration signals for the six different conditions; amplitude (vertical) vs time (horizontal)

As shown in Fig. 5, the test rig used to collect the vibration data consists of a DC motor driving the shaft over a flexible coupling, with the shaft supported by two Plummer bearing blocks. A sequence of damage bearing were inserted in one of the Plummer blocks, and the resultant vibrations in the horizontal and vertical planes were measured using two accelerometers. The output from the accelerometers was fed back through a charge amplifier to a Loughborough Sound Image DSP32 ADC card (using a low-pass filter with a cut-off 18 kHz), and sampled at 48 kHz, giving a slight oversampling. The machine was run at a series of different speeds ranging between 25 and 75 rev/s, and ten time series were taken at each speed. This provided a total of 160 examples of each condition, and a total of 960 raw data files to work with.

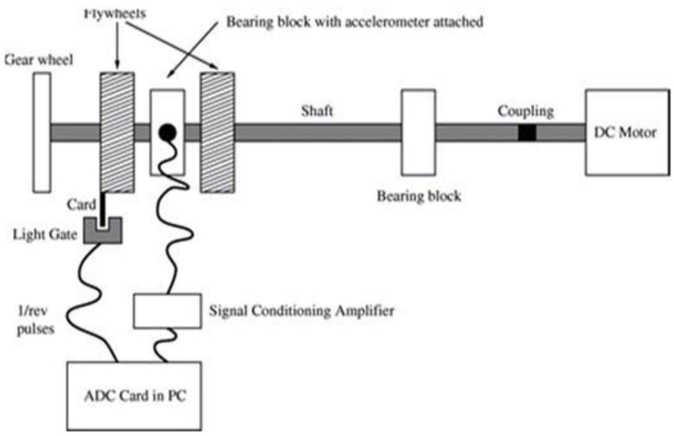


Fig. 5. The test rig used to collect the vibration data of bearings

B. *Experimental Setup*

We began by obtaining the compressed samples of the bearing vibration signal (i.e., "x = ψs" where "ψ" is the Fourier matrix). First, we used Fast Fourier Transform (FFT) to obtain the sparse components that needed by CS framework as shown in Fig. 6. Then we applied compressive sampling with different

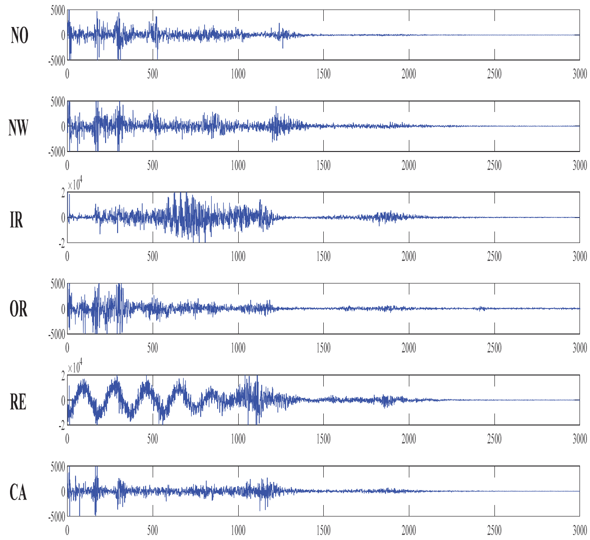


Fig. 6. Corresponding absolute values of Fourier coefficients for each condition signal

sampling rates (α) (0.1, 0.2 and 0.3) with 600, 1200, and 1800 compressed measurements of our original vibration signal using random Gaussian matrix with m*n size. The number of compressed signal elements m = α*n, where n is the length of the original vibration signal.

Based on CS theory described in section II, the obtained compressed samples have sufficient information of the original signal, i.e., possess the quality of the original signal. In our proposed method, these compressed samples are used to represent our original vibration signal. In order to further reduce the number of features, LS is suggested to select the k most significant features from a given compressed samples for the purpose of classification. We set k to 200 and 300. We randomly choose 576 samples for training set and the remaining 384 samples were used as testing set (i.e., 60% and 40% of the compressed samples with the selected features). Then we apply SVM with the fitcecoc function [32] using one-against-one (OAO) multiclass classification technique for every possible pair of classes and Gaussian RBF kernel function with kernel scale 17.

The computing platform for the experiment in this paper is a laptop, with an Intel® Core ™ i3-5010U CPU @ 2.1 GHZ processor, 8GB RAM, running on 64-bit Windows 10. MATLAB R 2016a is used as the main testing platform, with necessary classifier toolboxes.

C. *Experimental Results*

Various experiments were conducted to evaluate the classification ability of features generated by our proposed method. Table I provides a summary of the classification results obtained from several generated features based on different sampling rates (left column) and different amount of selected features (second left column). The classification accuracy rates are achieved by averaging the results of thirty trials for each classifier and for each experiment.

It is clear from the results that choice of small sampling rate (0.3) can lead to high classification accuracies in both cases of selected features (200 and 300), unlike choices of sampling rate (0.1 and 0.2) that suffer from a small degradation in the classification results in 200 selected features although they achieved better classification results for 300 selected features. Table II show some sample confusion matrices for the three sampling rates with 200 selected features.

Table I. Classification accuracies (%) and related Root Mean Square Error (RMSE) for generated features (k refers to number of selected features and α is the sampling rate).

α	k	Training accuracy (%)	Testing accuracy (%)
0.1 (600 samples)	200	99.6 ± 0.2	98.1 ± 0.7
	300	99.9 ± 0.4	99.7 ± 0.5
0.2 (1200 samples)	200	99.4 ± 2.7	98.5 ± 2.7
	300	99.9 ± 0.2	99.6 ± 0.3
0.3 (1800 samples)	200	99.8 ± 0.5	99.4 ± 0.9
	300	99.9 ± 0.1	99.8 ± 0.3

To verify the efficiency of the proposed method, Table III presents a comparison results of several methods using the same vibration datasets used in this paper. The first method in [33] where the same vibration data used and a classification result from raw vibration with entropic features is reported. The second method, we used the popular algorithm for ICA (Fast-ICA) in [34] to extract features from our raw vibration data with the same SVM classifier used to obtain results in Table I. Genetic Programming (GP) feature extraction method in [35] where the classification results with the same data reported. As a final point, PCA and LDA were used to extract features from compressed samples and the results are reported in [36]. It is apparent that results from our proposed method are very competitive. In particular, results from 300 selected features for all sampling rates, i.e., α = 0.1, 0.2 and 0.3, achieved better classification accuracies compared to other methods as can be seen in Table III.

Table II. Sample confusion matrix for α = 0.1, 0.2 and 0.3 with 200 selected features.

Bearing classes	True classes NO	NW	IR	OR	RE	CA	Class Prediction (%)
NO	64	0	0	0	0	0	100
NW	0	64	0	0	0	0	100
IR	0	2	59	0	3	0	92.2
OR	0	2	0	62	0	0	97
RE	0	0	2	0	62	0	97
CA	0	0	0	0	0	64	100

(a) Sampling rate α = 0.1

Bearing classes	True classes NO	NW	IR	OR	RE	CA	Class Prediction (%)
NO	64	0	0	0	0	0	100
NW	0	64	0	0	0	0	100
IR	1	1	60	0	2	0	96.8
OR	0	2	0	62	0	0	97
RE	0	0	2	0	62	0	97
CA	0	0	0	0	0	64	100

(b) Sampling rate α = 0.2

Bearing classes	True classes NO	NW	IR	OR	RE	CA	Class Prediction (%)
NO	64	0	0	0	0	0	100
NW	0	64	0	0	0	0	100
IR	0	2	62	0	0	0	97
OR	0	0	0	64	0	0	100
RE	0	2	0	0	62	0	97
CA	0	0	0	0	0	64	100

(c) Sampling rate α = 0.3

Table III. A comparison with the classification results from literature.

	Testing accuracy
Raw data (with entropy) [33]	98.9 ± 1.2
Fast-ICA (our results)	97.6 ± 2.4
GP extracted feature sets [35]	
SVM	97.1
ANN	96.5
[36] CS-PCA α = 0.1	98.5 ± 0.4
α = 0.2	98.7 ± 0.6
CS-LDA α = 0.1	89.6 ± 3.5
α = 0.2	100
Our proposed method (200 selected features)	
α = 0.1	98.1 ± 0.7
α = 0.2	98.5 ± 2.7
α = 0.3	99.4 ± 0.9
Our proposed method (300 selected features)	
α = 0.1	99.7 ± 0.5
α = 0.2	99.6 ± 0.3
α = 0.3	99.8 ± 0.3

V. Conclusion

In this investigation, the aim was to propose a new technique for bearing faults classification based on the combination of CS, LS and SVM, and to evaluate its efficiency using experimental bearing vibration data. From the experimental results, the proposed method has achieved a high classification accuracy with significantly reduced feature sets and its performance was compared with some existing methods. It appears that the proposed approach can be successfully used for a number of high-dimension pattern recognition applications by means of CS based samples compression, LS based feature selection and MSVM classifier based feature classification. On the basis of the promising results presented in this paper, the next stage of our research will be to explore different feature selection techniques with a range of sampling rates and small numbers of selected features.

Acknowledgment

Professor Nandi is a Distinguished Visiting Professor at Tongji University, Shanghai, China. This work was partly supported by the National Science Foundation of China grant number 61520106006 and the National Science Foundation of Shanghai grant number 16JC1401300.

References

[1] P. J. Tavner, "Review of condition monitoring of rotating electrical machines," *IET Electrical Power Applications*, Vol. 2, IET (2008), pp. 215-247.

[2] S. Nandi, H. A. Toliyat and X. Li, "Condition monitoring and fault diagnosis of electrical motors-a review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, 2005, pp. 719-729.

[3] A. K. Nandi, C. Liu, M. Wong, Intelligent Vibration Signal Processing for Condition Monitoring." *Proceedings of the International Conference Surveillance* 7, 2013.

[4] B. Sreejith, A. K. Verma and A. Srividya, "Fault diagnosis of rolling element bearing using time-domain features and neural networks," *IEEE Region 10 and the Third international Conference on Industrial and Information Systems*, Kharagpur, 2008, pp. 1-6.

[5] L. Zhang and A. K. Nandi, "Fault classification using genetic programming," *Mechanical Systems and Signal Processing,* vol. 21, no. 3, 2007, pp. 1273-1284.

[6] H. Guo, L. B. Jack and A. K. Nandi, "Feature generation using genetic programming with application to fault classification," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, vol. 35, no. 1, 2005, pp. 89-99.

[7] Z. Feng, M. Liang, F. Chu, Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples, *Mechanical Systems and Signal Processing* 38 (2013), pp.165–205.

[8] L. B. Jack and A. K. Nandi, "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 16, no. 2-3, 2002, pp. 373-390.

[9] R. Yan, R. Gao, and X. Chen. "Wavelets for fault diagnosis of rotary machines: a review with applications." *Signal Processing.* Vol. 96, Part A, 2014, pp. 1-15.

[10] D. H. Pandya, S. H. Upadhyay, S. P. Harsha, "Ann Based Fault Diagnosis of Rolling Element Bearing Using Time-Frequency Domain Feature," *International Journal of Engineering Science and Technology* 4, no. 6, 2012, pp. 2878–2886.

[11] Y. G. Lei, J. Lin, Z. J. He, et al., "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, 35 (1–2), 2013, pp. 108–126.

[12] B. A. Paya, I. I. Esat, M. N. M. Badi, "Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a pre-processor," *Mechanical Systems and Signal Processing,* 11 (1997) 751–765.

[13] W. J. Staszewski, K. Worden, G. R. Tomlinson, "Time-frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition", *Mechanical Systems and Signal Processing,* vol. 11, no. 5, 1997, pp. 673-692.

[14] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemom. Intell. Lab. Syst., vol. 2, no. 1–3, 1987, pp. 37–52.

[15] S. Balakrishnama, and A. Ganapathiraju. "Linear discriminant analysis-a brief tutorial," *Institute for Signal and information Processing* 18, 1998.

[16] P. Comon, "Independent component analysis, A new concept," *Signal Processing*, vol. 36, no. 3, 1994, pp. 287–314.

[17] L. B. Jack, and A. K. Nandi. "Genetic algorithms for feature selection in machine condition monitoring with vibration signals." *IEE Proceedings-Vision, Image and Signal Processing* 147.3 (2000): pp. 205-212.

[18] L. Shuang and L. Meng, "Bearing Fault Diagnosis Based on PCA and SVM," *International Conference on Mechatronics and Automation*, Harbin, 2007, pp. 3503-3507.

[19] F. Wang, J. Sun, D. Yan, S. Zhang, L. Cui and Y. Xu, "A Feature Extraction Method for Fault Classification of Rolling Bearing based on PCA," *Journal of Physics: Conference Series, IOP Publishing,* vol. 628, No. 1, 2015, pp. 012079.

[20] L. Jiang, Z. Liu, J. Cui, and Z. Li, "Condition monitoring of rolling element bearing based on Phase-PCA," 2013 *25th Chinese Control Decis. Conf. CCDC*, 2013, pp. 1082–1086.

[21] L. Ciabattoni, G. Cimini, F. Ferracuti, A. Freddi, G. Ippoliti and A. Monteriú, "A novel LDA-based approach for motor bearing fault detection," *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, Cambridge, 2015, pp. 771-776.

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

[23] Y. Chang and W. Jiao, "ICA-ANN method in fault diagnosis of rotating machinery," 2012 *IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, Zhangjiajie, 2012, pp. 236-240.

[24] H. O. A. Ahmed, M. L. Dennis Wong and A. K. Nandi, "Effects of deep neural network parameters on classification of bearing faults," IECON 2016 - *42nd Annual Conference of the IEEE Industrial Electronics Society*, Florence, 2016, pp. 6329-6334.

[25] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," in *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, 2004, pp. 1517-1525.

[26] D.L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory*. 52 (2006) 1289–1306.

[27] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," in *IEEE Signal Processing Magazine*, vol. 25, no. 2, 2008, pp. 21-30.

[28] X. He, D. Cai, and P. Niyogi. "Laplacian score for feature selection." NIPS. Vol. 186. 2005.

[29] C. Cortes, and V. Vapnik. "Support-vector networks." *Machine learning* 20.3, 1995, pp. 273-297.

[30] R. G. Brereton, G. R. Lloyd, "Support vector machines for classification and regression", Analyst, vol. 135, 2010, pp. 230-267.

[31] C. W. Hsu, C. J. Lin, "A comparison of methods for multiclass support vector machines", I*EEE Trans. Neural Networks*, vol. 13, 2002, pp. 415-425.

[32] "Fit multiclass models for support vector machines or other classifiers – MATLAB fitcecoc", Mathworks.com, 2017. [Online]. Available: http: //uk.mathworks.com/help/stats-/fitcecoc.html.

[33] M.L.D. Wong, M. Zhang, A.K. Nandi, "Effects of compressed sensing on classification of bearing faults with entropic features," *EUSIPCO proceeding, IEEE*, 2015, pp. 2296–2300.

[34] The FastICA MATLAB Package. Available from http://www.cis.hut.fi/ projects/ica/fastica.

[35] H. Guo, L. B. Jack, A. K. Nandi, Feature generation using genetic programming with application to fault classification, *IEEE Transactions on Systems, Man, and Cybernetics*, Part B: Cybernetics, Vol 35, No. 1. IEEE (2005), pp. 89-99.

[36] H.O.A. Ahmed, M.L.D. Wong, A.K. Nandi, "Compressive sensing strategy for classification of bearing faults". *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 2182-2186.