

Three-stage Method for Rotating Machine Health Condition Monitoring Using Vibration Signals

H.O.A. Ahmed¹ and A. K. Nandi^{1,2*}

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

²The Key Laboratory of Embedded Systems and Service Computing, College of Electronic and Information Engineering, Tongji University, Shanghai, P.R. China

Abstract— This paper proposes a new three-stage method for rotating machines health condition monitoring. In the first stage of the proposed method, Multiple Measurement Vectors Compressive Sampling (MMV-CS) is used to obtain compressively-sampled signals from the acquired raw vibration signals. In the second stage, a process combining Geodesic Minimal Spanning Tree (GMST), Stochastic Proximity Embedding (SPE), and Neighbourhood Component Analysis (NCA) is used to estimate and further reduce the dimensionality of the compressively-sampled signals. In the third stage, with these reduced features, multi-class Support Vector Machine (SVM) classifier is used to classify machine health conditions. Experiments on a roller element bearing fault detection and classification task based on vibration signals are used to verify the efficiency of the proposed method. Results show that the proposed method with fewer features achieved high classification accuracy of bearings health conditions and outperformed recently published results.

Index Terms—Machine Condition Monitoring, Compressive Sampling, Dimensionality Reduction, Multi-class Support Vector Machine.

I. INTRODUCTION

Rotating machines are essential for most engineering process in industry and their breakdowns can directly affect production timetables, production quality, and production costs. Therefore, it is very important to monitor machine health condition to avoid machine breakdowns. In fact, the need for an effective condition monitoring and machinery maintenance program exists wherever complex and expensive machinery are used to deliver critical function of businesses. Machine Condition Monitoring (MCM) has been applied in various sensitive applications of rotating machines, such as power generation, oil and gas, helicopter, etc. [1 – 3].

Based on types of sensor data acquired from rotating machines, MCM can be categorised into the following: vibration monitoring, acoustic emission monitoring, electric current monitoring, temperature monitoring, and chemical monitoring [4]. Of these methods, vibration-based condition monitoring has been extensively studied and has become a well-accepted technique of many planned maintenance managements [5, 6]. Vibration signals can be collected from rotating machine through vibration sensors, e.g., displacement sensors, and

accelerometers. In practice, the acquired vibration signal is usually sampled satisfying Nyquist rate, in which the sampling rate must be at least double the maximum frequency present in the signal. Thus, the collected vibration data represents a large amount of time series data that need to be transmitted, stored, and processed. It is clear acquiring a large amount of data requires a large storage and time for signal processing. In addition, this might limit the number of machines to be monitored remotely through Wireless Sensor Networks (WSNs) due to bandwidth and power restrictions.

The aim of vibration-based MCM is to early detect and identify machine faults by analysing the physical features of the collected vibration signals from which one is able to classify the vibration signal into the corresponding machine health condition using a classifier. Then, the decision of maintenance is made based on the current machine health condition. However, the large amount of collected vibration signals make it difficult to achieve accurate and early detection before machine breakdowns. To address the challenges of analysing a large amount of vibration data, the common used approach is to extract features from the time series vibration data that can be accomplished using various techniques. For instance, extracting certain features from the time series raw vibration data, which can be done directly in the time-domain using some statistical parameters, e.g., root mean square, peak-to-peak value, crest Factor, kurtosis, skewness, etc. [7]; or by transforming the time series raw vibration data to other domains, such as frequency domain by using Fast Fourier Transform (FFT) and time-frequency domain using various techniques, e.g., Short Time Fourier Transform (STFT), Wavelet Transform (WT), Hilbert-Huang Transform (HHT), Empirical Mode Decomposition (EMD), etc. [8].

As far as the high dimensionality of the extracted features is concerned, various dimensionality reduction techniques are used to improve the computational efficiency and the classification accuracy. In fact, dimensionality reduction can be performed by transforming the original features into a new reduced space of features or by selecting a subspace of the original features. For example, Harmouche et al. [9] proposed a method that combines the envelop analysis, the sliding FFT, and Principal Component Analysis (PCA) for the classification of bearing faults. Here PCA is employed to get a lower dimensional principal space. In

[10] comparison of several linear and non-linear dimensionality reduction techniques, e.g., PCA, Kernel PCA, Maximum Variance Unfolding (MVU), Local Tangent Space Analysis (LTSA), Diffusion Maps (DM), Laplacian Eigenmaps (LE), are used to reduce various statistical features extracted from the raw vibration signal. With this reduced features, a decision tree is used to classify faults of mono block centrifugal pump. A new method combining the time-frequency distribution and the nonlinear manifold to extract effective quantitative representations of machinery health pattern is proposed in [11]. Li et al. [12] proposed a method based on WT-Autogressive (AR) model and PCA for gear multi-fault diagnosis. Ciabattini et al. [13] presented a novel Linear Discriminant Analysis (LDA) based algorithm to reduce the fault data dimension for motor bearing fault detection.

Various studies have assessed the efficacy of selecting a subspace of the original features. For instance, Jack et al. [14] examined the use of a genetic algorithm (GA) to select the most important features from different form of features extracted from the raw vibration sampled data for MCM. Van et al. [15] employed a feature selection technique combining Distance Evaluation Optimization (DET) and Particle Swarm Optimization (PSO) to select the superior feature subset of a set of features extracted using non-local-means denoising and EMD for bearing fault diagnosis.

New advances in transform coding techniques have facilitated investigation of Compressive Sampling (CS) framework that relies on linear dimensionality reduction. Recently, several studies in vibration-based MCM have based their criteria for reducing the amount of the collected data on CS. For example, Wong et al. [16] studied the effects of CS on the classification of bearing faults and found slight performance degradation after reconstructing the original signal from the compressed measurements. Ahmed et al. [17] proposed a compressive sampling strategy for classification of bearing faults based on learning directly from the compressed measurements. Similarly, an intelligent condition monitoring for bearing faults based on learning from highly compressed measurements is proposed in [18]. In an effort to select fewer features from the CS-based compressed measurements, a method combining CS, Laplacian Score (LS), and SVM to classify rolling bearing faults is proposed in [19].

In this paper, an automatic three-stage method for rotating machine health condition monitoring using vibration signals is presented. In this method CS is used to obtain compressively sampled data from the acquired vibration data. Then, to further reduce the dimensionality of these compressively sampled data, first, GMST is used to find the minimal number of features necessary to represent the original data, second, with these minimal number of features a combination of SPE and NCA is used to automatically select far fewer features of the compressively-sampled data. With these features SVM is used to classify machine health condition.

II. COMPRESSIVE SAMPLING

Compressive sampling (CS) [20] is an extension of sparse representations and special case of it. The basic idea of CS is that many real-world signals have sparse representations in some domain, e.g., frequency domain, can be recovered from

fewer measurements satisfying certain conditions. CS framework relies on two principles: (1) sparsity of the signal of interest, and (2) the measurements matrix that fulfills the data minimal information loss, i.e., satisfies Restricted Isometry Property (RIP). Briefly, we define the single measurement vector compressive sampling (SMV-CS) as follows:

Let $x \in R^{n \times 1}$ be an original time indexed signal. Given a sparsifying transform matrix $\psi \in R^{n \times n}$ whose columns are the basis elements $\{\psi_i\}_{i=1}^n$. Thus, x can be represented as follows:

$$x = \sum_{i=1}^n \psi_i s_i \quad (1)$$

or

$$x = \psi s \quad (2)$$

Here s , is $n \times 1$ column vector of coefficients. Provided that the basis ψ generates q -sparse representations of x , then equation (1) can be rewritten as follows:

$$x = \sum_{i=1}^q \psi_{ni} s_{ni} \quad (3)$$

Here ni is the index of the basis elements and the coefficients corresponding to the q non-zero elements. Thus, $s \in R^{n \times 1}$ is a vector column with only q non-zero elements and represents the sparse representation vector of x .

CS framework can produce $m \ll n$ projections of the vector x with a group of measurement vectors $\{\phi_j\}_{j=1}^m$ and the sparse representations s of x such that

$$y = \phi \psi s = \theta s \quad (4)$$

Here y is $m \times 1$ column vector of the compressed measurements and $\theta = \phi \psi$ is the measurement matrix. To generate good compressed measurements that can be used to reconstruct the original signals, the measurement matrix θ has to fulfill the data minimal information loss, i.e., satisfy the RIP.

Definition 1.1: The measurement matrix $\theta \in R^{m \times n}$ satisfies the Restricted Isometry Property (RIP) if there exists a parameter $\delta \in (0,1)$ such that

$$(1 - \delta) \|s\|_{l_2}^2 \leq \|\theta s\|_{l_2}^2 \leq (1 + \delta) \|s\|_{l_2}^2 \quad (5)$$

The size of θ ($m \times n$) relies on the compressive sampling rate (α) (i.e., $m = \alpha \times n$). Fig. 1 shows an illustration of the CS framework. According to CS theory, given the compressed measurements y we can recover s by solving the following optimization problem

$$\hat{s} = \min_{s \in R^n} \frac{1}{2} \|\theta s - y\|_{l_2}^2 + \gamma \|s\|_{l_1} \quad (6)$$

Here $\|\theta s - y\|_{l_2}^2 \leq \varepsilon$ for a selected $\varepsilon > 0$, and a specific regularization parameter $\gamma > 0$ that controls the relative importance applied to the sparseness l_1 and the l_2 norms. Thus, with \hat{s} and the inverse ψ we can recover the original vector x such that

$$\hat{x} = \psi^{-1} \hat{s} \quad (7)$$

Multiple measurement vectors CS (MMV-CS) is considered for signals that are represented as a matrix such that

$$Y = \Theta S \quad (8)$$

where $Y \in R^{m \times L}$, L is number of observations and $S \in R^{n \times L}$ is a sparse representation matrix.

In this paper, MMV-CS has been utilised since the dataset consists of a matrix of multiple measurements. Furthermore, since it is possible to recover the original signal (x) from the compressed data (y) this indicates that y possesses the quality of the original signal x . Consequently, in this study, we use the compressed measurements directly.

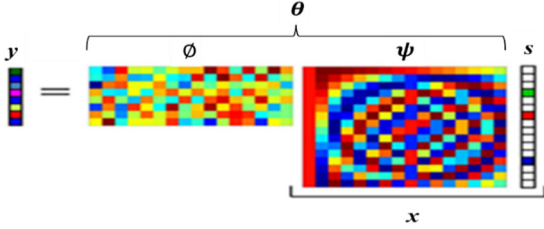


Fig. 1. Compressive sampling framework

III. THE PROPOSED METHOD

This section describes the proposed method for machine health condition monitoring. As shown in Fig. 2, the process to monitor the health condition of a rotating machine using the proposed method consist of three stages, data acquisition and compression, feature learning and selection, and machine health condition classification.

A. The first stage: Data acquisition and compression

In this stage, the proposed method collects vibration data from a rotating machine of interest and then compress it using MMV-CS framework in equation (8). Here FFT is chosen to obtain the sparse representations ($S \in R^{n \times L}$) from the collected raw vibration data ($X \in R^{n \times L}$) and the measurement matrix ($\Theta \in R^{m \times n}$) is selected to be a random Gaussian matrix, which satisfies the RIP [21] and the number of the compressed measurements m is controlled using a compressed sampling rate (α) where $m = \alpha \cdot n$. The process of data compression is summarised in the following Algorithm. 1.

Algorithm 1: Compressive sampling process

Input: $X \in R^{n \times L}$, $\Theta \in R^{m \times n}$, α

Output: $Y \in R^{m \times L}$

- 1: $\text{FFT}(X) \rightarrow S \in R^{n \times L}$
- 2: Project S into Θ with compressed sampling rate α to obtain compressively-sampled signal $Y \in R^{m \times L}$

B. The second stage: Feature learning and selection

Despite the fact that the compressively-sampled signals obtained in the first stage are ideal for recovering the original signals, they might not be the best for signal classification. In addition, they may contain an over-representations of the data. Thus, in this stage, we motivate that further dimensionality reduction is useful in early fault detection and may improve classification results. So, the proposed method employ a procedure combining GMST, SPE, and NCA to reduce the

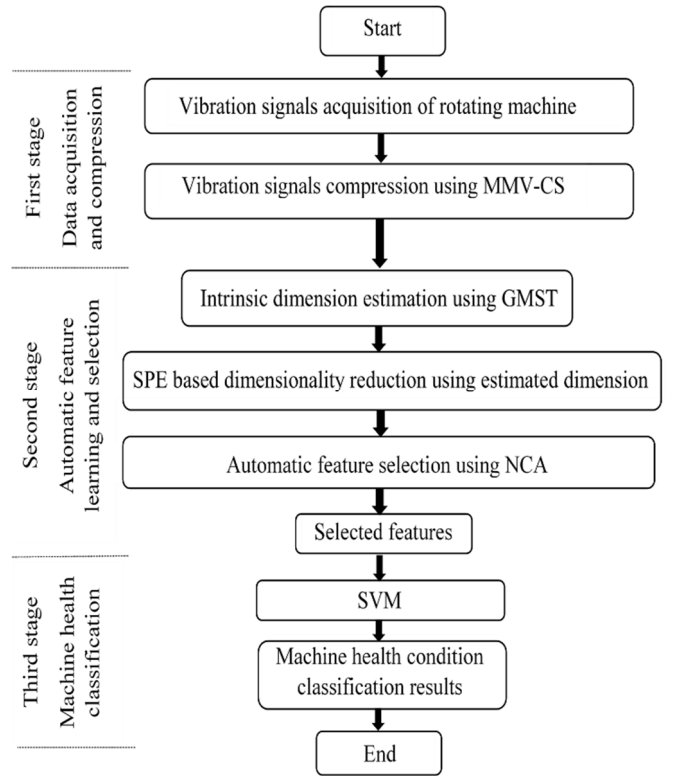


Fig. 2. The proposed method

dimensionality of the compressively-sampled signal. First, the global dimension estimator GMST is used to define the intrinsic dimensionality, i.e., define the minimal number of features required to represent the compressively-sampled data [22, 23]. GMST compute the geodesic graph G from which the intrinsic dimension (p) is estimated by computing multiple Minimum Spanning Tree (MST) in which each data sample x_i is linked to its k nearest neighbors such that

$$p(Y) = \min \sum_{e \in T} D_{Eucl} \quad (9)$$

Here T represent the set of all the sub-trees of G , e is an edge in T , and D_{Eucl} is the Euclidean distance of e .

Second, having defined the minimal number of features p ($p < m$), the compressively-sampled data can be transformed into a reduced dimensionality space of significant representation. Various linear and nonlinear techniques have been proposed and used to reduce data dimensionality. In this stage of our proposed method, we used a combination of SPE [24] and NCA [25] to automatically select far fewer features of the compressively-sampled data. SPE is a nonlinear technique that has several attractive features: (1) simple to implement, (2) it is very fast, (3) it scales linearly with the size of the data in both time and memory, and (4) it is relatively insensitive to missing data [26]. So, it was decided that SPE is a suitable technique to adopt for our investigation.

SPE uses a self-organizing iterative scheme to embed an m dimensional data into p dimensions, such that the geodesic distances in the original m dimension are preserved in the embedded d dimension. To compute a reduced dimension from

the compressively-sampled signals using SPE, the following steps are performed.

- Initialize the coordinates y_i . Select an initial learning rate β .
- Select a pair of points, i and j , at random and compute 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 by:

$$y_i \leftarrow y_i + \beta \frac{1}{2} \frac{r_{ij} - d_{ij}}{d_{ij} + v} (y_i - y_j) \quad (10)$$

and

$$y_j \leftarrow y_j + \beta \frac{1}{2} \frac{r_{ij} - d_{ij}}{d_{ij} + v} (y_j - y_i) \quad (11)$$

Here v is a small number to avoid division by zero. For a given number of iterations, this step will be repeated for a prescribed number of steps and β will be decreased by a suggested decrement $\delta\beta$.

Finally, after mapping the compressively-sampled data to the reduced dimension space using SPE, each feature vector will be further reduced by selecting k features using NCA to improve the performance of the proposed method. In our case, we applied “fscna” function [27] to perform the feature selection. It achieves feature selection by regularising the feature weights and select features with feature weights that is greater than a relative threshold (τ).

C. The third stage: Machine health condition classification

To classify machine health condition, we employed Multi-class Support Vector Machine (SVM) classifier. The basic idea of SVM is that it can find the best hyperplane(s) to separate two classes. Based on the features of the data, SVM can make linear or non-linear classifications by different kernel functions, e.g., Radial Basis Function (RBF), Polynomial Function (PF), and Sigmoid Function (SF) [28]. Multi-class SVM includes multiple two-class sub-problems, i.e., SVM classifiers that can be easily combined together using one-versus-one and one-versus-all coding design.

In our case, we applied “fitcecoc” function [29] on the learned features from the second stage. It uses $c(c-1)/2$ binary SVM models using one-versus-one coding design, where c is the number of unique class labels. This will return a fully trained error-correcting output codes (ECOC) multiclass model that cross-validated using 10-fold cross-validation.

IV. EMPIRICAL VALIDATION

In order to validate the proposed method, we now perform several computer experiments on a collected vibration dataset of roller bearing using our proposed method.

A. Data description

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

case (CA) fault. Fig. 3 displays some typical time series plots for the six conditions described above.

As shown in Fig. 4, the test rig used to collect the vibration data involves a DC motor driving the shaft through a flexible coupling, with the shaft supported by two Plummer bearing blocks. A series of damaged bearing was inserted in one of the Plummer blocks, and the resultant vibrations in the horizontal and vertical planes were measured using two accelerometers. The output from the accelerometers was fed back through a charge amplifier to a Loughborough Sound Images DSP32 ADC card (using a low-pass filter with a cut-off 18 kHz), and sampled at 48 kHz. The machine was run at a series of 16 different speeds ranging between 25 and 75 rev/s, and ten time series were taken at each speed. This gave a total of 160 examples of each condition, and a total of 960 raw data files to work with. The description of the dataset is presented in Table. 1.

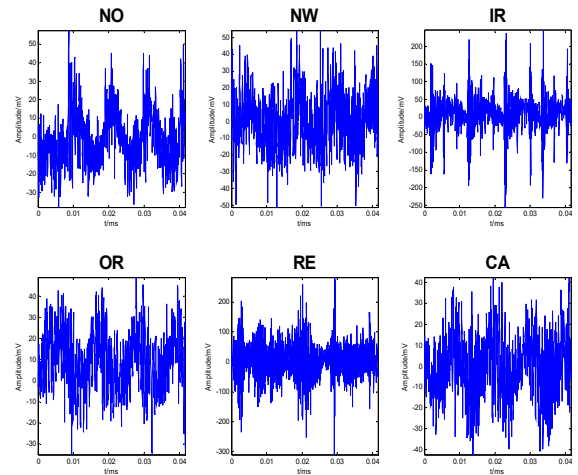


Fig. 3. Typical time – domain vibration signals for the six different conditions

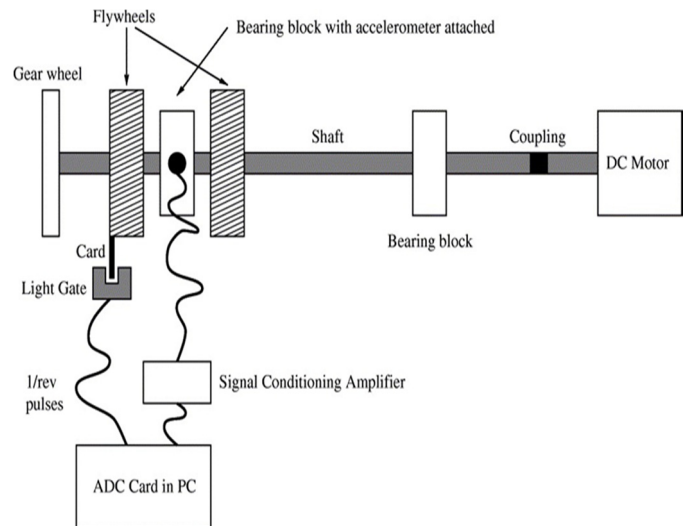


Fig. 4. The test rig used to collect the vibration data of roller bearings.

TABLE I. DESCRIPTION OF BEARING DATASET, L REFERS TO NUMBER OF OBSERVATIONS AND N REFERS TO NUMBER OF SAMPLES.

Bearing conditions		L	N
Normal	NO	160	6000
	NW	160	6000
Fault	IR	160	6000
	OR	160	6000
	RE	160	6000
	CA	160	6000

B. Experimental results

To apply our proposed method to this bearing dataset, fifty percent of the total observations were randomly selected for training and the other fifty percent for testing. We started by obtaining the compressively-sampled vibration data by means of the MMV-CS model with FFT basis as sparse representation of the raw vibration data, a random Gaussian matrix as a measurement matrix, and different values of α (0.1, 0.2, 0.3, and 0.4).

To guarantee that our CS model produces enough samples for the purpose of bearing fault classification, we used the produced compressively-sampled signals in the first stage to reconstruct the original signal X by applying the Compressive Sampling Matching Pursuit (CoSaMP) algorithm [30]. For example, with $\alpha = 0.1$ the average percentage reconstruction errors for the six conditions of bearings are 1.8% (NO), 0.9% (NW), 3.3% (IR), 1.6% (OR), 0.7% (RE), and 2.6% (CA), which indicate good signal reconstruction.

To learn features from the compressively-sampled signals, we used the second stage of our proposed method described in Fig. 2. First, we used GMST to define the intrinsic dimensionality, i.e., define the minimal number of features required to represent the compressively-sampled data for each value of α . Second, with this intrinsic dimension, we used SPE to find a reduced dimensional embedding of a compressively-sampled data that preserve the distances between neighboring points, i.e., keep the distance between two data points in the reduced dimensional data of Y in equation (8) to be identical with their corresponding points distance in Y . The parameters of SPE algorithm were set as follows: number of iterations were set to 100, number of steps was set to 100, β was set to 1, and $\delta\beta$ was set to 0.01.

In the final step of the second stage of our proposed method we performed feature selection using “fscna” [27]. This function returns a model that contains information about the training data, model, and optimization. NCA was fitted with stochastic gradient descent (SGD) and regularisation parameter (λ). To find the best value of λ that produces the minimum classification loss, we tuned λ by training the NCA model using different values of λ on the features that produced from Y with $\alpha = 0.1$ in the second step of the second stage of our proposed method. Fig. 5 shows the loss values versus the λ values. The parameters of NCA were set as follows: τ was set to be 0.02 and the best value of λ with minimum loss ($\lambda = 0.004$) was selected.

Figure 6 shows an example of the selected features using our proposed method with $\alpha = 0.1$ (i.e., $m = 600$) and intrinsic dimension $p = 130$. As can be seen in Fig. 5 only 19 features will be selected from the features produced using the intrinsic

dimension $p = 130$ of the 600 data points of the compressively

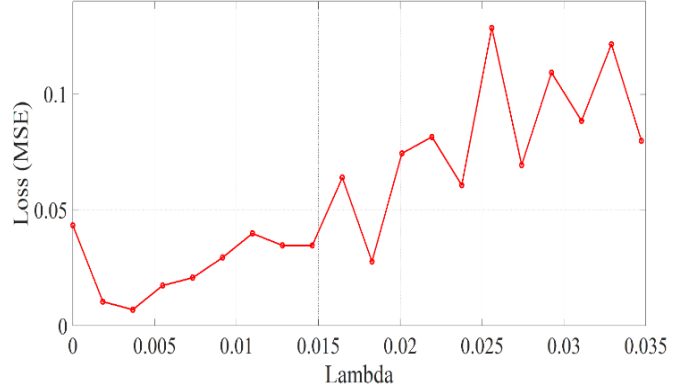


Fig. 5. The loss values versus the regularization parameter (λ) values.

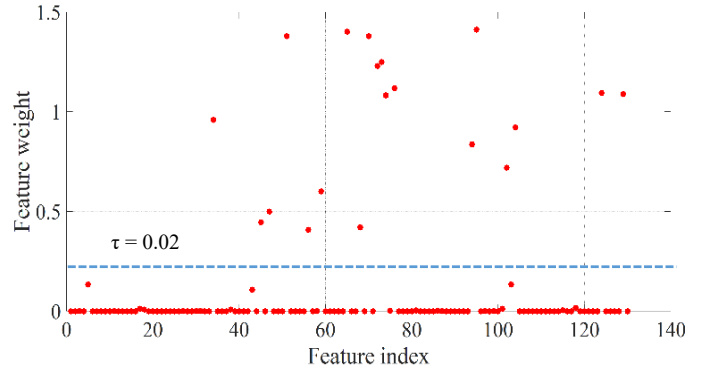


Fig. 6. Example of selected feature using our proposed method.

sampled signal.

The selected features in the second stage were used to train the multi-class SVM. For better evaluation of the trained multi-class SVM model, we applied 10-fold cross-validation where the training dataset is randomly subdivided into ten subsets. Several experiments were conducted to evaluate the classification performance of features selected using our proposed method. Table II provides a summary of the classification results achieved from various compressively-sampled datasets based on different values of α as can be seen in the left column of Table II.

TABLE II. CLASSIFICATION ACCURACIES (%) AND RELATED ROOT MEAN SQUARE (RMS) FOR SELECTED FEATURES (p REFERS TO THE INTRINSIC ESTIMATED DIMENSION, AND k IS THE NUMBER OF THE AUTOMATICALLY SELECTED FEATURES).

α	p	k	Testing Classification Accuracy (%)
0.1 (600 compressed samples)	39	14	98.8 ± 2.4
0.2 (1200 compressed samples)	46	15	99.4 ± 0.5
0.3 (1800 compressed samples)	64	26	99.9 ± 0.2
0.4 (2400 compressed samples)	59	28	99.9 ± 0.1

The classification accuracy rates are obtained by averaging the results of ten trials for each experiments. It is apparent from Table II that high level of classification accuracy were achieved using less than 40% of the original data samples. Particularly, classification accuracy from the proposed method are 99.9%, for 40% and 30% of the original data with 28 and 26 selected features respectively. Table III presents some sample confusion matrix for $\alpha = 0.1, 0.2,$ and 0.3 .

TABLE III. SAMPLE CONFUSION MATRIX

Bearing classes	True classes						Class Prediction (%)
	NO	NW	IR	OR	RE	CA	
NO	79	1	0	0	0	0	98.75
NW	0	80	0	0	0	0	100
IR	0	0	78	2	0	0	97.5
OR	0	0	3	77	0	0	96.25
RE	0	0	0	0	80	0	100
CA	0	0	0	0	0	80	100

(a) Sampling rate $\alpha = 0.1$

Bearing classes	True classes						Class Prediction (%)
	NO	NW	IR	OR	RE	CA	
NO	80	0	0	0	0	0	100
NW	0	80	0	0	0	0	100
IR	0	0	80	0	0	0	100
OR	0	0	0	76	4	0	95
RE	0	0	0	0	80	0	100
CA	0	0	0	0	0	80	100

(b) Sampling rate $\alpha = 0.2$

Bearing classes	True classes						Class Prediction (%)
	NO	NW	IR	OR	RE	CA	
NO	80	0	0	0	0	0	100
NW	0	80	0	0	0	0	100
IR	0	0	80	0	0	0	100
OR	0	0	0	80	0	0	100
RE	0	0	0	0	80	0	100
CA	0	0	0	0	0	80	100

(c) Sampling rate $\alpha = 0.3$

For further evaluation of the efficiency of our proposed method, Table IV presents the comparisons with some recently published results using the same bearing dataset as in [16]. Three methods were used to report the classification results in [16]: (1) first method uses the whole original signal, (2) second uses compressed measurements for $\alpha = 0.5$ then recover the original signals, and (3) third method uses compressed measurements for $\alpha = 0.25$ then recover the original signals. In [17] a method uses a combination of CS with $\alpha = 0.4$ and PCA, and in [19] uses a combination of CS with $\alpha = 0.3$ and Laplacian score (LS) feature selection. Finally, results reported in [31] by means of a hybrid method involves Fuzzy Min-Max (FMM) neural network and Random Forest (RF) using all the original data.

It is clear that results from our proposed method with $\alpha = 0.3$ and 0.4 are better than those results obtained using the methods

listed in Table IV. Also, the classification results from the proposed method with $\alpha = 0.2$ are better than those reported in [16, 17, and 19] and remain competitive with results reported in [31] although we are using only 20% of the original data which is not matched by the method in [31] using 100% of the data .

TABLE IV. A COMPARISON WITH THE CLASSIFICATION RESULTS FROM LITERATURE ON BEARING DATASET

	Testing Classification Accuracy (%)
Raw Vibration [16]	98.9 ± 1.2
Compressed Sensed ($\alpha = 0.5$) followed by reconstruction [16]	92.4 ± 0.5
Compressed Sensed ($\alpha = 0.25$) followed by reconstruction [16]	84.6 ± 3.4
CS-PCA ($\alpha = 0.4$) [17]	98.8 ± 0.7
CS-LS ($\alpha = 0.3$) with 300 selected features [19]	99.8 ± 0.3
FMM-RF (SampEn +PS) [31]	99.81 ± 0.41
The proposed method	
($\alpha = 0.2$) with 15 selected features.	99.4 ± 0.5
($\alpha = 0.3$) with 26 selected features.	99.9 ± 0.2
($\alpha = 0.4$) with 28 selected features.	99.9 ± 0.1

V. CONCLUSION

The main goal of the current study was to propose a vibration-based three-stage method for rotating machine health condition monitoring combining CS, a process of GMST, SPE, and NCA, and SVM, and to examine its efficacy using experimental vibration data of roller bearings. As of the experimental results, the proposed method has achieved high classification accuracies with significantly reduced feature sets and its classification accuracy outperformed existing methods. The findings of this study suggest that the proposed method can be efficiently used for various high-dimension pattern recognition applications by means of MMV-CS based sample compression, the process of GMST, SPE, and NCA based feature selection and SVM classifier based accurate condition classification.

REFERENCES

- [1] J. Nilsson and L. Bertling, "Maintenance Management of Wind Power Systems Using Condition Monitoring Systems—Life Cycle Cost Analysis for Two Case Studies," in IEEE Transactions on Energy Conversion, vol. 22, no. 1, pp. 223-229, 2007.
- [2] R. Giussani, L. Renforth, M. Seltzer-Grant and C. Zachariades, "A holistic combined electrical and mechanical condition monitoring technique for oil and gas high voltage rotating machines," IEEE Electrical Insulation Conference (EIC), Seattle, WA, pp. 283-287, 2015.
- [3] D. Coats, K. Cho, Y. J. Shin, N. Goodman, V. Blechertas and A. M. E. Bayoumi, "Advanced Time-Frequency Mutual Information Measures for Condition-Based Maintenance of Helicopter Drivetrains," in IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 8, pp. 2984-2994, 2011.
- [4] W. Zhou, T. G. Habetler and R. G. Harley, "Bearing Condition Monitoring Methods for Electric Machines: A General Review,"

- 2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Cracow, pp. 3-6, 2007.
- [5] S. J. Lacey, "An overview of bearing vibration analysis", *Maintenance Asset Manage.*, vol. 23, no. 6, pp. 32-42, 2008.
- [6] R. B. Randall, J. Antoni, "Rolling element bearing diagnostics—A tutorial", *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485-520, 2011.
- [7] M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers*, U.K., Cambridge: Cambridge Univ. Press, pp. 94, 1989.
- [8] Z. Feng, M. Liang, F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples", *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165-205, Jul. 2013.
- [9] J. Harmouche, C. Delpha and D. Diallo, "A global approach for the classification of bearing faults conditions using spectral features," *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Vienna, pp. 7352-7357, 2013.
- [10] N. R. Sakthivel, B. B. Nair, M. Elangovan, V. Sugumaran, Y. S. Saravanmurugan, "Comparison of Dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibrations signals", *Elsevier Engineering Science and Technology an International Journal*, pp. 30 – 38, 2014.
- [11] Q. He, "Time–frequency manifold for nonlinear feature extraction in machinery fault diagnosis", *Mech. Syst. Signal Process.*, vol. 35, no. 1–2, pp. 200-218, Feb. 2013.
- [12] Z. X. Li, X. P. Yan, C. Q. Yuan, Z. X. Peng, L. Li, Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet- autoregressive model and principal component analysis method, *Mechanical Systems and Signal Processing*, Vol. 25, No. 7, 2589-2607, 2011.
- [13] L. Ciabattini, 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, pp. 771-776. 2015.
- [14] L. B. Jack and A. K. Nandi, "Genetic algorithms for feature selection in machine condition monitoring with vibration signals," in *IEE Proceedings - Vision, Image and Signal Processing*, vol. 147, no. 3, pp. 205-212, 2000.
- [15] M. Van and H. J. Kang, "Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection," in *IET Science, Measurement & Technology*, vol. 9, no. 6, pp. 671-680, 2015.
- [16] M.L.D. Wong, M. Zhang, A.K. Nandi, "Effects of compressed sensing on classification of bearing faults with entropic features," *EUSIPCO proceeding*, IEEE, pp. 2296–2300, 2015.
- [17] H.O.A. Ahmed, M.L.D. Wong, A.K. Nandi, "Compressive sensing strategy for classification of bearing faults". *ICASSP proceedings*, IEEE. 2017.
- [18] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi. "Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features." *Mechanical Systems and Signal Processing* 99, pp. 459-477, 2018.
- [19] H. O. A. Ahmed, M. L. D. Wong and A. K. Nandi, "Classification of bearing faults combining compressive sampling, laplacian score, and support vector machine," *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, Beijing, pp. 8053-8058, 2017.
- [20] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," in *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30, 2008.
- [21] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simplified proof of the restricted isometry property for random matrices (aka the Johnson-Lindenstrauss lemma meets compressed sensing)," *Constructive Approximation*, 2007.
- [22] J.A. Costa, A.O. Hero, "Geodesic Entropic Graphs for Dimension and Entropy Estimation in Manifold Learning", *IEEE Trans. Signal Processing*, vol. 52, no. 8, pp. 2210-2221, 2004.
- [23] J. A. Costa and A. O. Hero, "Manifold learning using Euclidean k-nearest neighbor graphs [image processing examples]," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 988-991 vol.3, 2004.
- [24] D. K. Agrafiotis, "Stochastic proximity embedding", *J. Comput. Chem.*, vol. 24, pp. 1215-1221, 2003.
- [25] Yang, W., K. Wang, W. Zuo. "Neighborhood Component Feature Selection for High-Dimensional Data." *Journal of Computers*. Vol. 7, Number 1, 2012.
- [26] D. K. Agrafiotis, H. Xu, F. Zhu, D. Bandyopadhyay, P. Liu, "Stochastic proximity embedding: Methods and applications", *Molecular Informatics*, vol. 29, pp. 758-770, 2010.
- [27] "Feature selection using neighborhood component analysis for classification–MATLAB fscna", *Mathworks.com*, 2018 [Online]. Available: <https://www.mathworks.com/help/stats/fscnca.html>.
- [28] C. Cortes, V. Vapnik, "Support vector networks", *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, 1995
- [29] "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>.
- [30] D. Needell, J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", *Appl. Computat. Harmon. Anal.*, vol. 26, no. 3, pp. 301-321, 2009.
- [31] Seera. M, M. L. D. Wong, and A. K. Nandi. "Classification of ball bearing faults using a hybrid intelligent model." *Applied Soft Computing* 57, pp. 427-435, 2017.