Analysis and Interpretation of Bearing Vibration Data Using Principal Component Analysis and Self-Organizing Map

Raj Kumar Patel*
Department of Electrical Engineering,
Madan Mohan Malaviya University of Technology, India
E-mail: rajkp007@gmail.com
* Corresponding Author

V. K. Giri
Department of Electrical Engineering,
Madan Mohan Malaviya University of Technology, India
E-mail: girivkmmm@gmail.com

Abstract: Induction motor bearing is one of the key parts of the machine and its analysis and interpretation are important for fault detection. In the present work vibration signal has been taken for the classification i.e. bearing is Healthy (H) or Defective (D). For this purpose, clustering based classification of bearing vibration data has been carried out using Principal Component Analysis (PCA) and Self Organising Map (SOM). From the vibration signal, twelve statistical features have been extracted from both the healthy and the defected condition of the bearing. Further, these data are subjected to PCA to extract significant features relevant to cluster structure. It is observed that out of twelve features only four features are found significant which is feed to the SOM model. The SOM based classification is able to achieve an accuracy of 100%. This cluster-based method of feature reduction and classification could be useful in assessing the induction motor incipient bearing fault detection with large data set.

Keywords: Bi plot, Principal component analysis, Self organising map, Statistical features vibration signal

Biographical notes: Raj Kumar Patel obtained his B. Tech (Electrical and Electronics) degree from Uttar Pradesh Technical University (India) in 2006, and M. Tech (Condition Monitoring Control and Protection of Electrical Apparatus) degree from NIT, Hamirpur (India) in 2012. He is presently pursuing PhD degree from Uttar Pradesh Technical University, Lucknow (India). His research interests include Signal processing, Condition monitoring and Fault diagnosis of rotating machines. **V. K. Giri** obtained his B.E. (Electrical) Degree from REC, Surat (Gujrat) in 1988, M.E. Hons. Degree from University of Roorkee, Roorkee in 1997 and PhD degree from Indian Institute of Technology Roorkee, Roorkee in 2003. Presently, he holds the position of Professor in the same department since, 2008. His research interests include Digital Signal Processing, Measurement and Instrumentation, Biomedical Instrumentation, ECG, Data Compression and Telemedicine.

1   INTRODUCTION

Induction motors are the workhorse of industries, so that it is required to keep running these machines for long hours. To meet the demand it is essential to monitor the parts of machines, specially rotating parts because the failure of a single part may lead to the shutdown of the machine [1], [2]. The most critical part of any rotating machine is bearing. Failure of bearing ceases the production in plant causing production losses. Various features e.g. vibration, temperature, lubricant oil analysis, thermography, electric current, acoustic emission, etc, measures how well the moving part of the machine is working [3]. The vibration recording performs the complete evolution of the bearing system of the machine hence it is most commonly used for assessing the bearing of an induction motor. It monitors fault progression in different parts of the bearing and perform incipient fault assessment. Fault diagnosis using vibration signal is a reliable method of identifying faults and is used to grade the fault severity [4].

Different data analysis technique have been applied and developed to provide significant data analysis for condition monitoring which includes time domain method, frequency domain method and time-frequency method [5][6]. Automated diagnosis methods that evaluate bearing fault for appropriate induction motor are very much essential [7], [8]. Data clustering for common technique has been employed for multi-label classification. It has been used in many fields, including machine learning, data mining, pattern recognition, image analysis, and bio informatics [9]. Clustering performs classification of data by finding natural groups using unsupervised learning [10][11]. Cluster analysis divides data into clusters such that similar data objects belong to the same cluster and dissimilar data objects to different clusters. The Principal Component Analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observation of possible correlated variables into a set of values of uncorrelated variables called principal components [14], [15].

The mathematical approach used in PCA is called eigen analysis. The eigen vector associated with the largest eigenvalue has the same direction as the first principal component. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the variability as possible. The main objectives of the PCA are to reduce the dimensionality of the data set and to identify new meaningful underlying variables [16][17]. The SOM is an unsupervised neural network that characterizes a relatively massive amount of data and performs clustering based classification [18]. It projects a high dimensional input space on a two dimensional output space. The Unique feature of SOM is that it can use neighbourhood kernels to preserve and also control the topological structure of high dimensional input data. The projection is topological preserving, that is, where patterns that are similar in terms of the input space are mapped to geographically close locations in the output space. The SOM is composed of a set of nodes arranged in a geometric pattern, typically a two dimensional lattice. Each node is associated with a weight vector W with same dimensions as the input space. Learning in SOM is based on competitive learning where these output nodes of the network compete among themselves to be activated or fired. Only one output node or one node per group is ON at any time. The output nodes that win the competition are called winner take all nodes. SOM algorithm can serve as an abstraction of data set, scalable to large datasets and it is robust to the missing data and rare outliers [19].

In this work, vibration signal of induction motor bearing has been taken to investigate the defected condition of the bearing using SOM model as a clustering based classification. The twelve statistical features have been extracted and feed to the classifier. As a large number of features could be extracted from vibration, there may be a chance of classification of bearing fault. Asaithambi et al. have been observed that all extracted features or features are not equally significant and may degrade the performance of the classifier [19]. Hence, there is a need for data reduction of significant features for appropriate interpretation. Hence, in the present work feature has been selected using PCA and finally found four significant features out of twelve features. Reduced features are feed to the SOM model and it has been found that performance of reduced feature SOM model is better than the performance of SOM model with all the features.

2   METHODOLOGY

The vibration signal data are carried out from Case Western Reserve University Bearing Data Center for the present study [25]. The data chosen from the data are for the healthy and defective bearing. The sampling frequency of the recording was 12000 samples per second. Statistical features have been utilized as a tool for fault diagnosis in rolling element [20]. The statistical calculated features in this paper are given in Table 1. In this work, acquired data is pre-processed using PCA before feeding it to the clustering algorithm. The most significant features of recorded data are derived using PCA. Principal component analysis is a statistical method to estimate the most influenced common stochastic series by analysing cross-sectional correlation. It is utilized for extracting high dimensionality by constructing a linear projection of the bearing data set in the two dimensional space.

Principal component analysis is performed on the symmetric Covariance matrix or on the symmetric Correlation matrix. These matrices can be calculated from the data matrix. The induction motor bearing data for 100 segments is arranged in the form of a matrix. Each segment

consists of 3000 points. PCA transforms the data by extracting statistically independent components and arranging them in the order of relative significance. The original feature space consists of 12 statistical features of the healthy and defective bearing.

Table 1 Statistical Features calculated for Healthy and Defective Bearing cases

Features	Definition	Measures
F1	Mean	Measure of
F2	Median	Central
F3	Mode	tendency
F4	Standard Deviation	Measure of
F5	Variance	Variability
F6	Range	
F7	Kurtosis	Measures of
F8	Skewness	Dispersion
F9	Crest Factor	Dimension less
F10	Impulse Factor	Features
F11	Clearance Factor	
F12	Shape Factor	

PCA is employed in transforming this feature space into a new space and the components that account for most of the variability are retained whereas the remaining components are ignored. In PCA method, bearing vibration data arranged in the form of matrix of size 100×12 is reduced into a covariance matrix 'A' of size 12×12, where the columns correspond to a number of principal components and rows are associated with the vibration signal variables. The covariance matrix 'A' is given by:

$$\phi_{ij} = \frac{1}{n} \sum_{k=1}^{n} \left(x_{ik} - \overline{x_i} \right) \left(x_{kj} - \overline{x_j} \right), \ i \neq j \qquad (1)$$

Where, x is mean of n variables in the matrix. The Principal Components (PCs) are determined as the eigenvectors of the covariance matrix A. Eigen values $\left(\lambda_1, \lambda_2, \lambda_3 .... \lambda_n \right)$ can be found by solving the determinant Eq. $|\phi - \lambda I| = 0$. Then PCs of the covariance matrix ϕ can be generated using a process of singular value decomposition, which is given by :

$$\phi = E \wedge E^T \qquad (2)$$

The set of PCs is then represented as a linear combination of the original variables. PCA uncovers combinations of the original variables which describe the dominant patterns and the main trends in the data. This is done through an Eigen vector decomposition of the covariance matrix of the original variables. The extracted latent variables are orthogonal and they are sorted according to their Eigen values [21]. The percentage of variance (PV) for the principal components can be calculated by:

$$PV = \frac{\lambda_k}{\lambda_1 + \lambda_2 + ... + \lambda_n} \times 100\% \qquad (3)$$

The PCA from the perspective of statistical pattern recognition is an effective technique for dimension reduction. The principal components that explain the maximum percentage variance are chosen and the corresponding component magnitudes are analyzed. The statistical features variables corresponding to these components are considered as the most significant features. The three featuress having highest magnitudes in the loadings of the principal components are chosen for further cluster based classification using SOM.

SOM consists of two layers; the input layer and the output layer. The output layer is composed of a set of nodes arranged in a two dimensional lattice. Each neuron from the output layer has a double representation, one is its position in the grid and another is its weight vector. The dimension of the weight vector equals the dimension of the input data vectors. In the training process, the weights are gradually changed in order to span the weight vectors across the input data set. The training is based on competitive and cooperative learning [22] .

In the SOM network, the input pattern x represented by a vector of length 3 corresponding to significant features is presented to a two dimensional map of 25×25 nodes. During each training cycle t_k, every input pattern corresponding to 100 subjects is considered in turn and the best matching weight vector W, also called winner node is determined such that :

$$\|x - w_i\| = \min \|x_i - w_i\|, (0, 1, ... N) \qquad (4)$$

The weights of the neuron, which resembles the most to that input data, are updated. The weight vectors are updated using the following adaptation function.

$$w_i(t_k + 1) = \begin{cases} w_t(t_k) + [a(t_k) w_t(t_k) - x] & for \ i \ \varepsilon N_i(t_k) \\ w_i(t_k), & otherwise \end{cases} \qquad (5)$$

Where, $a(t_k)$ is the learning coefficient that decreases over time and $N_i(t_k)$ is the set of nodes considered to be in the topological neighbourhood of the node i, the winner node. Node i represents the neuron that maximally responds to the input signal, i.e., its weight vector matches most closely, among all the Kohonen layer nodes, to that of the input vector. N_I contains all nodes that are within a certain radius from node i. The weight vectors of all the nodes within the set N_I are updated at the same rate [23].

When the SOM has been trained, the centroids of the input patterns that create winning neuron are identified as cluster centers. The SOM network is used for classification of normal and obstructive subjects. The performance of classification is evaluated by accuracy, sensitivity, and

specificity. The performance of a clustering and classification system can be measured by the clustering accuracy and the classification accuracy [24]. The clustering accuracy is defined as Eq. 6.

Clustering Accuracy = Number of correct clustering combinations/Number of total combinations   (6)

Correct clustering indicates the number of clusters determined by the algorithm that is matched to the number of classes present in the vibration data.

Table 2: Mean and Standard Deviation for Healthy and Defective Bearing

Principal Component	Mean ± Standard Deviation			
	Healthy		Defective	
PC1	0.1860	±0.2442	-0.1694	±0.2442
PC2	0.0691	±0.2880	-0.0854	±0.2880
PC3	0.0165	±0.2832	-0.0992	±0.2832
PC4	0.1665	±0.2712	0.1262	±0.2712
PC5	-0.0443	±0.2713	0.1260	±0.2713
PC6	0.0869	±0.2956	0.0571	±0.2956
PC7	0.0774	±0.2995	0.0332	±0.2995
PC8	0.0164	±0.2998	-0.0309	±0.2998
PC9	0.0141	±0.3009	-0.0191	±0.3009
PC10	-0.0017	±0.3015	0.0038	±0.3015
PC11	0.0012	±0.3015	-0.0023	±0.3015
PC12	-0.0021	±0.3015	0.0024	±0.3015

3   RESULTS AND DISCUSSIONS

Twelve Statistical features are calculated and they are subjected to PCA. The statistical analysis consisting mean and standard deviation of the principal components of the healthy and defective cases of the bearing are presented in Table 2. The mean values of these principal components in healthy case are distinctly higher than those of the defective case. The standard deviations for all the principal components for both healthy and defective cases are found to be high.

The percentage variance obtained by various principal component for normal and obstructive data are shown in Fig.1. It is illustrated that the first two principal Component (PCs) account for 60% of the variance. The higher percentage variance indicated that these PCs are sufficient to explain the similarity pattern of each data and capture most of the discrimination capability. These principal components alone are enough to guide clustering and removing features with low variance provides a more robust clustering.

PC1 and PC2 obtained for Healthy case represented as a weighted linear combination of original featuress are given in Eq. 7 and Eq. 8.

$$PC1 = -0.04(F1) - 0.17(F2) - 0.04(F3) + 0.07(F4) + 0.07(F5) \\ + 0.40(F6) + 0.38(F7) + 0.20(F8) + 0.40(F9) \\ + 0.42(F11) + 0.34(F12) \quad (7)$$

$$PC2 = -0.03(F1) - 0.03(F2) - 0.001(F3) + 0.68(F4) + 0.68(F5) \\ + 0.17(F6) - 0.04(F7) - 0.03(F8) - 0.12(F9) - 0.12(F10) - \\ 0.11(F11) + 0.00(F12) \quad (8)$$

PC1 and PC2 obtained for defective case represented as a weighted linear combination of original featuress are given in Eq. 9 and Eq.10.

$$PC1 = -0.04(F1) - 0.02(F2) - 0.04(F3) - 0.01(F4) + 0.01(F5) \\ + 0.40(F6) + 0.40(F7) - 0.02(F8) + 0.43(F9) + 0.45(F10) \\ + 0.45(F11) + 0.31(F12) \quad (9)$$

$$PC2 = -0.26(F1) - 0.24(F2) + 0.05(F3) + 0.61(F4) + 0.61(F5) \\ + 0.27(F6) - 0.06(F7) + 14(F8) - 0.07(F9) - 0.07(F10) - \\ 0.07(F11) - 0.04(F12) \quad (10)$$
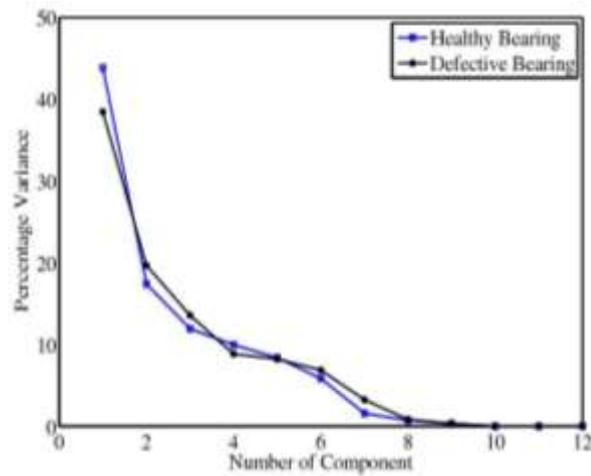

Fig. 1   Percentage Variance Obtained by Various Principal Components

In this article, a correlation between the bearing data and its twelve statistical features has been anaylysed by principle component analysis using bi-plot. Bi-plot of PCA are shown in Fig. 2 and 3. Biplot analysis was used to examine the relationships between the statistical features and its samples. In Fig. 2, we can see that F4,F5,F6 and F12, are highly correlated and their variability across healthy case is accounted mainly by PC1. Features F4,F5,F6,F12 are grouped on the positive Principle component 1 axis of the biplot, suggesting strong relationships among them and form cluster 1. Similarly, all other featuress F7,F8,F9,F10 and F10 have negative magnitude form cluster 2 for the healthy bearing. In Fig. 3, Features F4, F6 are found significant and form cluster 1 and all other features F7,F9,F10,F11,F12 have negative magnitude and form cluster 2 for defective case of bearing. The features with small and negative magnitudes are insignificant.
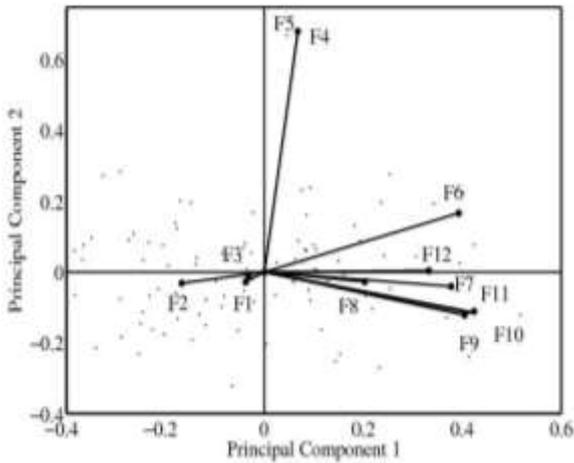
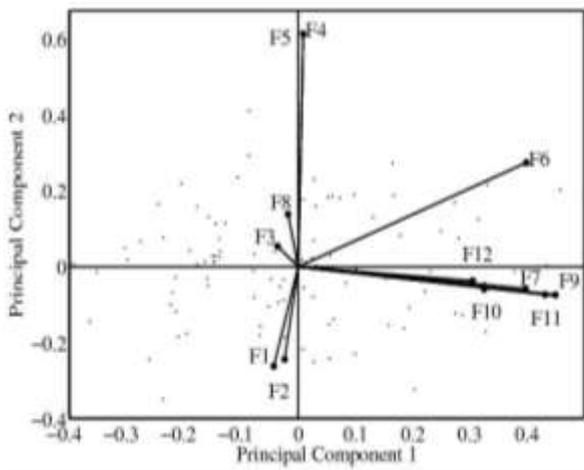**Fig. 2** Component loading plot for healthy bearing

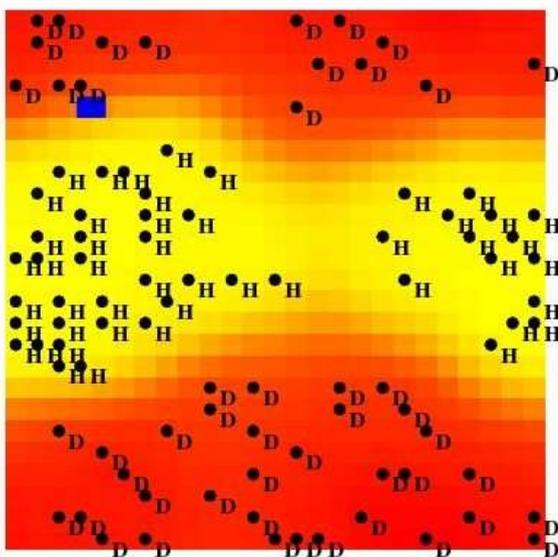**Fig. 3** Component loading plot for defected bearing

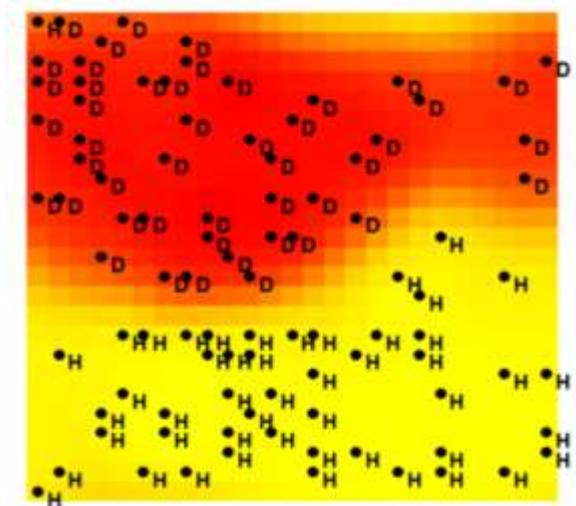**Fig. 4** Classification of Healthy and Defective Bearing using SOM
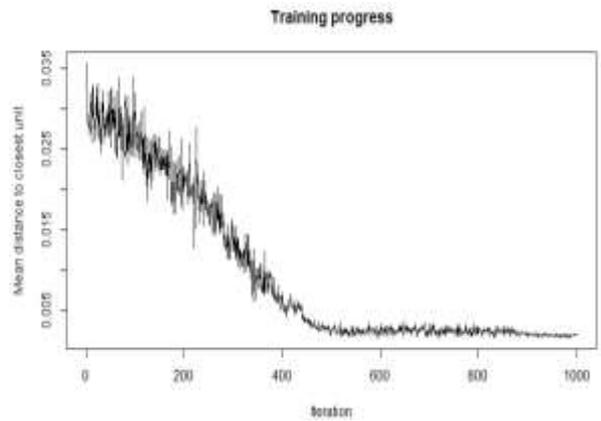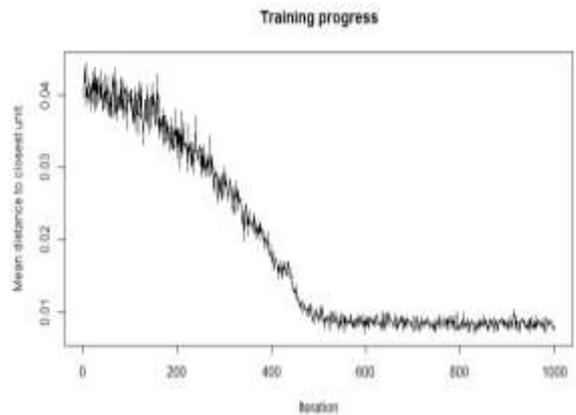

Fig. 5 Classification of Healthy and Defective Bearing using SOM

(a) For Dimension 25X25

(b) For Dimension 30X30

Fig. 6 Training process of SOM

The significant features obtained from the PCA i.e. F4, F5, F6, and F12 are given as the inputs to the self organizing map. Each pattern is presented and the best matching weight vector W for the present input is determined. The visualization map of SOM for classification of healthy (H)

and defective (D) bearings is shown in Fig. 4. The selection of node are based on trial and error and are presented for 25×25. It is found that nodes corresponding to each class are clustered and are mapped to geographically close locations in the output space. It is observed that the best results obtained for dimension of 25×25 for normal and defected bearings. Fig. 5 represent the visualization of SOM when all twelve features are given as input and it is observed that one healthy data has been misclassified.

The SOM model has been also observed based on its training. As the SOM training iterations progress, the distance from each node's weights to the samples represented by that node is reduced. Ideally, this distance should reach a minimum . Fig. 6 shows the training process progress over time for node 25×25 and 30×30 . If the curve is continually decreasing, more iterations are required. It has been observed from training curve that node 30×30 required more iteration.

5   CONCLUSIONS

Bearing vibration signal is the most frequently used mothed for the condition monitoring for the rotating machine and is an essential tool for the diagnosis of a defect in bearing. The condition monitoring of rotating machine depends on the accuracy, performance of the subject and on the measured and predicted values. Hence, there is a need to provide automated diagnostic support to engineer using cluster based classification schemes. In this work, clustering based classification of healthy and defective cases are analysed using PCA and SOM. Results show that PCA could extract significant features that are required for classification. These components provide a good stable clustering solution without any significant compromise on the quality of clusters. The clustering based SOM classification could achieve an accuracy of 100 % with a dimension of 25X25 neurons. It is observed that the proposed model can be used to enhance the diagnostic relevance of rotating machines used in industry.

ACKNOWLEDGMENTS

REFERENCES

[1] Randall, R. B., "Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications", JohnWiley & Sons, Ltd., UK, 2011, Chaps. 2.

[2] Gupta, K., "Vibration—A tool for machine diagnostics and condition monitoring," Sadhana, vol. 22, no. June, pp. 393–410, 1997.

[3] Mehala, N. and Dahiya, R., "Condition monitoring methods , failure identification and analysis for Induction machines," Int. J. circuits, Syst. Signal Process., vol. 3, no. 1, 2009, pp. 10–17.

[4] Sharma,A., Amarnath, M. and Kankar, P., "Feature extraction and fault severity classification in ball bearings," Journal of Vibration and Control, no. April, 2014, pp. 1–17.

[5] Patel, R. K. and Giri,V. K., "Induction Motor Bearing Fault Diagnosis Using Cascaded EMD and DWT techniques," International Journal of Applied Engineering Research., vol. 10, no. 11, 2015, pp. 28317–28330.

[6] Norouzi, R., Rafeeyan, M., Dalayeli, H., "Vibration Analysis of a Multi-Disk, Bearing and Mass Unbalance Rotor Using Assumed Modes Method," International Journal of Advanced Design and Manufacturing Technology, vol. 8, no. 1, 2015, pp. 25-33.

[7] C. Castejón, O. Lara, and J. C. García-Prada, "Automated diagnosis of rolling bearings using MRA and neural networks," Mechanical Systems and Signal Processing., vol. 24, no. 1, 2010, pp. 289–299.

[8] Jahedi, R., and Adibnazari, S., "Multi Layered Finite Element Analysis of Graded Coatings in Frictional Rolling Contact," International Journal of Advanced Design and Manufacturing Technology, vol. 8, no. 1, 2015, pp. 1-12.

[9] P. Konar, J. Sil, and P. Chattopadhyay, "Knowledge extraction using data mining for multi-class fault diagnosis of induction motor," Neurocomputing, vol. 166, 2015, pp. 14–25.

[10] L. Jiang, Y. Cao, H. Yin, and K. Deng, "An Improved Kernel K-Mean Cluster Method and Its Application in Fault Diagnosis of Roller Bearing," Engineering, vol. 2013, 2013, pp. 44–49.

[11] J. M. Bossio, C. H. Angelo, and G. R. Bossio, "Self-organizing map approach for classification of mechanical and rotor faults on induction motors," Neural Computing and Applications., vol. 23, no. 1, Nov. 2012, pp. 41–51.

[12] Q. He, "Time–frequency manifold for nonlinear feature extraction in machinery fault diagnosis," Mechanical Systems and Signal Processing., vol. 35, no. 1–2, Feb. 2013, pp. 200–218.

[13] P. Chattopadhyay and P. Konar, "Feature Extraction using Wavelet Transform for Multi-class Fault Detection of Induction Motor," Journal of the Institution of Engineers. Series. B, vol. 95, no. 1, May 2014, pp. 73–81.

[14] W. Zhou and B. Lu, "Incipient bearing fault detection via motor stator current noise cancellation using wiener filter," IEEE Trans on Industry., vol. 45, no. 4, 2009, pp. 1309–1317.

[15] Z. Hu, Z. Chen, W. Gui, and B. Jiang, "Adaptive PCA based fault diagnosis scheme in imperial smelting process.," ISA transactions., Jan. 2014, pp. 1–10.

[16] S. Dong, D. Sun, B. Tang, Z. Gao, W. Yu, and M. Xia, "A Fault Diagnosis Method for Rotating Machinery Based on PCA and Morlet Kernel SVM," Mathematical Problems in Engineering, vol. 2014, 2014, pp. 1-8.

[17] Akbari, M., Homaei, H., Heidari, M., "Faults Diagnosis of a Girth Gear Using Discrete Wavelet Transform And Artificial Neural Networks," International Journal of Advanced Design and Manufacturing Technology, vol. 7, no. 3, 2007, pp. 45-55.

[18] P. Sarlin, "Self-organizing time map: An abstraction of temporal multivariate patterns," Neurocomputing, vol. 99, 2013, pp. 496–508.

[19] Asaithambi, M., Manoharan, S. C., and Subramanian, S., "Clustering Based Analysis of Spirometric Data Using Principal Component Analysis and Self Organizing Map," in

4th International Conference, SEMCCO 2013, Chennai, India, December 19-21, 2013, Proceedings, Part II, 2013, pp. 523–533.

[20] Ahmad, K., Gupta, S. and Kumar, V., "Experimental Investigations on Outer Race Fault in Ball Bearings of Electrical Machines Using Statistical Analysis.," IETE J. Res. (Medknow Publ.), vol. 58, no. 2, 2012, pp. 96–106.

[21] Abouhnik, A., Ibrahim, G. R., Shnibha, R., & Albarbar, A.,"Novel Approach to Rotating Machinery Diagnostics Based on Principal Component and Residual Matrix Analysis," ISRN Mechanical Engineering,2012, pp.1-8.

[22] Kohonen, T., "The self-organizing map," Proceedings of the IEEE, vol. 78, no. 9, 1990, pp. 1464–1480.

[23] Aguado, D., Montoya,T., Borras, L., Seco, A., and Ferrer, J., "Using SOM and PCA for analysing and interpreting data from a P-removal SBR," Engineering Applications of Artificial Intelligence., vol. 21, no. 6, Sep. 2008, pp. 919–930.

[24] Chattopadhyay,M., Dan, P. K., and Mazumdar, S., "Principal Component Analysis And Self-Organizing Map For Visual Clustering Of Machine-Part Cell Formation In Cellular Manufacturing System," Systems Research Forum, vol. 05, no. 01, Jan. 2011, pp. 25–51.

[25] Case Western Reserve University, Bearing Data center [online database],
Available:URL:http://www.eecs.cwru.edu/laboratory/bearing/download. html, 2011.