Ball Bearing Fault Diagnosis Using Supervised and Unsupervised Machine Learning Methods

V. Vakharia, V. K. Gupta and P. K. Kankar

Mechanical Engineering Discipline, PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Jabalpur-482005, India

(Received 29 September 2013; accepted 9 June 2014)

This paper deals with the approach of using multiscale permutation entropy as a tool for feature selection for fault diagnosis in ball bearings. The coefficients obtained from the wavelet transformation of the vibration signals of the bearings are used for the calculation of statistical parameters. Based on the minimum multiscale permutation entropy criteria, the best scale is selected and statistical parameters such as crest factor, form factor, and permutation entropy are calculated. Finally, the faults are classified by considering the statistical parameters and permutation entropy as features in supervised and unsupervised machine learning methods, such as a support vector machine and self-organizing maps, respectively. Results revealed that the multiscale permutation entropy-based feature extraction techniques provide higher classification accuracy in comparison to the other methodologies that have been proposed in previous published works. The methodology proposed in this paper also gives good results for unsupervised learning methods, i.e. self-organizing maps.

NOMENCLATURE

C	penalty constant
m	embedding dimension
M	total number of samples
N	length of data
s	scale
t_1	time
z	time series
Y_j^s	coarse-grain time series
w	weight
ξ_i	slack variable
π	permutation pattern
τ	time delay

1. INTRODUCTION

Techniques designed to monitor the conditions of rolling element bearings receive considerable attention from researchers across the globe. Faults in the bearings are the major source of the breakdown of machinery. When a defect in the surface of one bearing strikes the surface of another, impulsive force is generated. This effect has been exploited by several vibration analysis methods, as well as various signal processing techniques.[1] Incipient fault diagnosis in rolling element bearings is essential for production efficiency and plant safety. Fault diagnosis depends mainly on the feature extraction techniques, because the signals carry dynamic information about the state of the machinery. The patterns of vibration signals, due to defects in various rotating parts, exhibit specific features. That is to say, faults can be identified by looking at pattern abnormalities in plant machinery and rotating parts. Signal processing techniques such as time domain, frequency domain, and combined time frequency domain (such as wavelet transformation) have been investigated by various researchers.[2–5] Due to variations in friction, loading conditions, interaction of various rotating elements, and clearance and nonlinear stiffness of the bearings,[6,7] the vibration signals generated by machinery are often characterized by nonlinearity. Thus, nonlinear parameter estimation techniques have been widely used by many researchers.[8–12] Numerous methods such as the correlation dimension[13] and the Lyapunov exponent[14] have been developed recently to detect nonlinearity. Entropy estimation is an important parameter for measuring system complexity. Analysis of the vibration signals generated from rotating machinery, using complexity measure such as approximate entropy[15] and multiscale entropy,[16] was used for the bearing fault diagnosis. Permutation entropy was introduced by Bandt and Pompe,[18] and is a new nonlinear parameter estimation tool that was efficiently used for the fault diagnosis.[17] By comparing neighbouring values, the complexity of a time series can be extracted using permutation entropy. Shannon entropy is useful for the estimation of the complexity of a time series based on a single scale, while multiscale permutation entropy is useful for calculating the complexity of a time series after comparing neighbouring values and entropy over multiple scales. Bandt and Pompe[18] presented permutation entropy, as a parameter of average entropy, to describe the complexity of a time series. It should also be noted that feature vectors consisting of multiscale permutation entropy provide better information about physical phenomena such as the occurrence of faults in the rotor bearing system.[19] The use of permutation entropy for chatter detection in the turning process,[20] known as electroencephalography (EEG) signal analysis,[21] has been used for chaotic time series. Signals obtained from complex mechanical systems that have several components are usually complicated. Approximate entropy and permutation entropy are based on a single scale, and are therefore inefficient in diagnosing the signals correctly. Multiscale permutation entropy was proposed by Costa, et al.,[22] and this concept was utilized by Wu, et al.[23] and Vakharia, et al.,[24] for bearing fault diagnosis and classification.

In this paper, raw vibration signals are used, and the concept of permutation entropy is utilized for the selection of scale. Wavelet coefficients are calculated by considering coiflet as

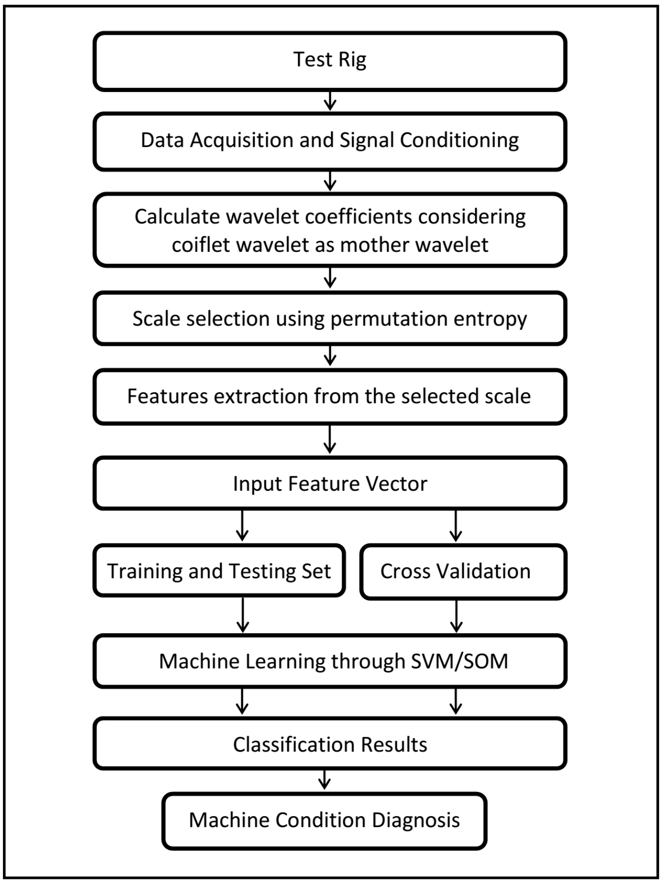

Figure 1. Proposed fault diagnosis strategy.

mother wavelet, and a scale giving the least permutation entropy is selected for the calculation of statistical parameters. A feature vector includes these statistical parameters: speed, loading condition, and permutation entropy. A feature vector is fed as the input for fault classification using a support vector machine (SVM) and self-organizing maps (SOM). Results revealed that the proposed feature extraction method gives improved results compared to the conventional feature extraction methods. The complete methodology for fault diagnosis is shown in Fig. 1.

2. PERMUTATION ENTROPY

Signals obtained from complex machinery parts using EEG show that stock markets are very complex in nature. Initially, entropy was used for quantifying the predictability of a time series on a single scale. It does not give insight between regularity and complexity. Costa, et al.[22] have developed multiscale entropy for the analysis of physiologic time series, in which initial sample entropy is calculated and, based on the concept of multiscale, various entropies can be calculated.

Permutation entropy was introduced as a computational efficient method for extracting the information from compound systems. For a given time series,

$$Z(t) = (Z_1, Z_2, Z_3, \ldots, Z_n); \tag{1}$$

at each time t_1, a vector composed of the m^{th} subsequent values is constructed as:[25]

$$t_1 \rightarrow (Z_{t_1}, Z_{t_1+1}, \ldots, Z_{t_1+(m-2)}, Z_{t_1+(m-1)}); \tag{2}$$

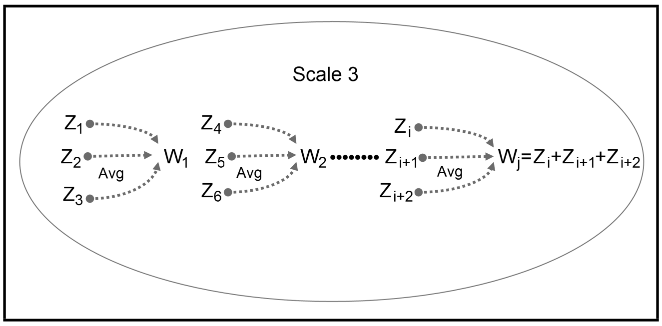

Figure 2. Coarse grain procedure.

where m is called embedding dimension and tells how much information is present in a vector. Further, by considering time delay of τ, Eq. (2) can be rewritten as

$$t_1 \rightarrow (Z_{t_1}, Z_{t_1+\tau}, \ldots, Z_{t_1+\tau(m-2)}, Z_{t_1+\tau(m-1)}). \tag{3}$$

For a given embedding dimension, there will be $m!$ possible permutation π of order m. Permutation entropy employs the concept of Shannon entropy by analysing the relative frequency of patterns generated from a time series. The permutation entropy is defined as:

$$\text{PE} = -\sum_{i=1}^{m!} \pi_i \ln \pi_i. \tag{4}$$

Permutation entropy depends mainly on the selection of embedding dimension m and time delay τ. Bandt and Pompe[18] suggested in their study that value of embedding dimension m should be $3 \leq m \leq 7$ and time delay $\tau = 1$.

Normalized permutation entropy is given by

$$\text{NPE} = \frac{\text{PE}}{\ln m!}; \tag{5}$$

where $\ln m!$ denotes maximum PE value.

2.1. Multiscale Permutation Entropy

The concept of multiscale entropy has been proposed by Costa, et al.[22] For a given time series and 3 scales, the data points are averaged by selecting non-overlapping windows of increasing length to form multiple coarse-grained time series as shown in Fig. 2.

For the scale factor s, the elements of coarse-grained time series are evaluated by

$$Y_j^s = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} z_i, \qquad 1 \leq j \leq \frac{N}{s}; \tag{6}$$

where N denotes the length of the data. For scale 1, the coarse-grained time series is simply the original time series.

3. MACHINE LEARNING TECHNIQUES

Machine learning is a type of artificial intelligence technique used essentially for classification and regression. An important task of machine learning is classification where algorithms are constructed between different data based on their specific patterns. Algorithms can be broadly categorized into supervised and unsupervised algorithms.

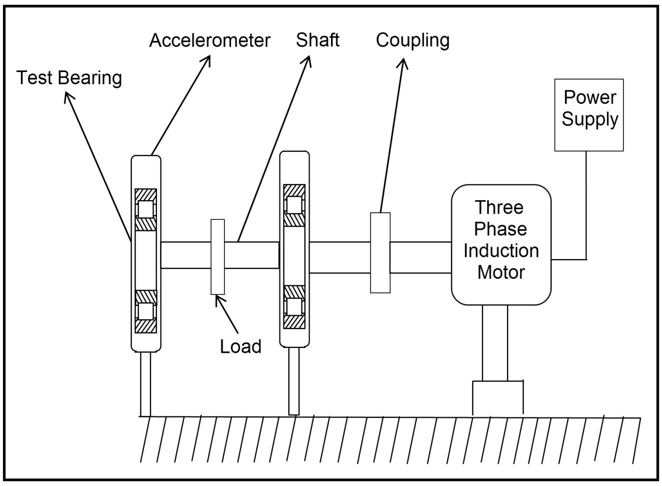

Figure 3. Schematic diagram of rotor bearing system.

3.1. Supervised Learning

In supervised learning, a label is associated with each feature of training data. Training data consists of input and desired result. The task of an algorithm is to search for patterns and develop mathematical models. Based on their prediction accuracy, models are evaluated. Naive Bayes, artificial neural networks, and support vector machines are some examples of supervised learning methods.

The support vector machine (SVM) is a statistical learning method based on the principle of structural risk minimization and was introduced by Vapnik.[26] SVM is a supervised learning algorithm in which a learning machine has allotted some set of features to a class of labels.

For linearly separable data, a hyperplane is constructed which separates hyperspace to achieve maximum separation between the classes known as the margin. The nearest data points that are used to define the margin are known as support vectors.

The optimal hyperplane separating the data can be obtained as a solution to the following optimization problem:
Minimize

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{M}\xi_i \tag{7}$$

subject to

$$y_i\left(w'x_i + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \ldots, M. \tag{8}$$

3.2. Unsupervised Learning

An unsupervised learning model is not provided with correct results during training. The task is to identify hidden patterns in unlabelled data. It can also be used for clustering the input data in to classes based on their statistical property. Self-organizing maps, K means, and blind signal separation are among the techniques which come under unsupervised learning.

A self-organizing map (SOM) is a type of neural network model commonly used for unsupervised classification.[27] In self-organizing maps, "self-organizing" means that no supervision is required. The models learn on their own through unsupervised competitive learning, while the "map" is used to map their weights according to the given input data.

Table 1. Parameters of bearing 6205 (SKF).

Parameter	Value
Outer race diameter	52 mm
Inner race diameter	25 mm
Ball diameter	7.94 mm
Ball number	10
Contact angle	0^o

In the initial stage, all node weights are initialized and a vector is randomly chosen from the set of training data and forwarded to network. Every node in the network is then used to calculate which weights are similar to the input vector. Nodes which are within the boundary of the best-matching unit are adjusted to make them like the input vector. Finally, the location of the most similar node is arranged in such a way that a topographic map is generated. The locations of the most similar nodes indicate statistical features contained in the input patterns.

4. EXPERIMENTAL SETUP AND DATA ACQUISITION

In the present study, an experimental test rig has been used, and vibration responses for healthy bearings and bearings with faults are obtained. A schematic diagram of a rotor bearing system is shown in Fig. 3. Table 1 shows the dimensions of the ball bearing used for this study. The vibration signals from the rig are taken after some hours of initial running.

The signals are measured at rotor speeds 1000, 1500, and 2000 rpm, and in all five classes with no loader, one loader, and two loader conditions, respectively. The following five bearing conditions are considered for the study:

1. Bearing with no defect (BND);
2. Bearing with spall on inner race (SI);
3. Bearing with spall on outer race (SO);
4. Bearing with spall on ball (SB);
5. Combined defects (CD).

The combined defects represent bearing conditions which have a spall on the inner race, the outer race, and ball together.

5. FEATURE EXTRACTION

Statistical analysis of vibration signals gives different primary and secondary parameters.

The continuous wavelet coefficients (CWC) of all signals were calculated at the 7th level of decomposition (2^7 scales). A scale giving the least permutation entropy was selected, and the statistical features of CWC corresponding to that scale were calculated for both horizontal and vertical directions. The selected embedded dimension m and the time delay τ of multiscale permutation entropy are 5 and 1, respectively.

The following features were selected for both horizontal and vertical conditions:

a. Permutation entropy, defined by Eq. (4).

b. Form factor, defined as the ratio of the RMS value and the average value of the signal.

Table 2. Sample input feature values for SOM/SVM.

Horizontal response			Vertical Response					
Permutation entropy	Crest factor	Form factor	Permutation entropy	Crest factor	Form factor	Loader	Speed	Class
0.6188	-0.9115	-1.0971	0.7100	0.8356	1.1968	0	1000	CD
0.5665	0.8123	1.2311	0.6356	-8.004	-1.2494	0	1500	CD
0.5823	-0.8453	-1.1831	0.7711	0.8259	1.2108	0	2000	CD
0.5169	-0.8519	-1.1738	0.5155	0.4019	2.4881	1	1000	SB
0.6154	0.4970	2.0121	0.8217	-0.6378	-1.5679	1	1500	SB
0.7166	-0.6240	-1.6025	0.7974	0.3539	2.8259	1	2000	SB
0.4255	0.8674	1.1528	0.4568	0.7593	1.3170	2	1000	SO
0.4120	0.8849	1.1301	0.7877	0.1105	9.0499	2	1500	SO
0.4939	-0.9028	-1.1076	0.8260	-0.5253	-1.9037	2	2000	SO
0.6045	0.8888	1.1251	0.5695	0.3862	2.5891	0	1000	BND
0.5901	0.2458	4.0691	0.5652	0.7917	1.2631	0	1500	BND
0.5666	-0.8682	-1.1518	0.6888	0.9204	1.0865	0	2000	BND
0.5093	0.7406	1.3503	0.7466	-0.7828	-1.2775	2	1000	SI
0.5712	-0.8311	-1.2032	0.5927	0.1317	7.5934	1	1500	SI
0.5241	0.4286	2.3334	0.5087	-0.6191	-1.6153	1	2000	SI

c. Crest factor, defined as the ratio of the peak value of the signal to its RMS value.

These features were extracted from the vibration signals and were fed as an input to machine learning techniques such as SOM and SVM for the classification of faults.

6. RESULTS AND DISCUSSION

Testing and cross validation of feature sets have been carried out using SOM and SVM as classifiers.[28] These features consist of permutation entropy, form factor, and crest factor, each for horizontal and vertical responses, and the number of loaders and rotor speed are also considered as the features for testing and cross-validation purposes. A sample input feature values are shown in Table 2.

The effects of bearing defects like ball defect, combined defect, outer race defect and inner race defect on multiscale permutation entropy are shown in Fig. 4. For the cases considered, it was found that for the vertical response, multiscale permutation entropy is higher as compared to the horizontal response.

With no loader and 1500 rpm, maximum multiscale permutation entropy is 0.7625 for ball defects under vertical response conditions, and minimum multiscale permutation entropy is 0.4015 for outer race defects under horizontal response conditions, as shown in Fig. 4(a). It can be interpreted that ball defects under no-loader conditions and 1500 rpm exhibit more disorder when compared to other classes, and outer race defects contain less disorder.

From Fig. 4(b), it is observed that the maximum multiscale permutation entropy, 0.8368, is for ball defects under vertical response conditions, and the minimum value of 0.521 for inner race defects is under horizontal response conditions. It can be interpreted that when speed is increased, disorder increased in ball defects and decreased in inner race defects. When the load is increased and the speed is 1500 rpm, the maximum multiscale permutation entropy is 0.836 for ball defects, the vertical response conditions and minimum multiscale permutation entropy is 0.3028 for combined defects, and the horizontal response conditions are shown in Fig. 4(c). With one loader and 2000 rpm, the maximum multiscale permutation entropy is 0.7974 for ball defects, the vertical response conditions and minimum multiscale permutation entropy is 0.5087 for inner race defects, and vertical response conditions are shown in Fig. 4(d). Thus, it is concluded that ball defects are severe as compared to other defects considered in the present study for all load conditions and speeds.

A total of 75 instances are considered which consist of 6, 16, 18, 17, and 18 cases of BND, SI, SO, SB, and CD, respectively. For coiflet wavelets, a scale is selected based on minimum multiscale permutation entropy. Testing and cross-validation results are shown in Tables 3 and 4 for SOM and SVM, respectively. Cross validation is a technique to evaluate the performance of classifiers. Therefore, 10-fold cross validation, which is the standard method of testing classifiers, was carried out.

From Table 3, for testing purpose 6/6, 16/16, 12/18, 17/17, and 18/18 cases were predicted correctly for BND, SI, SO, SB, and CD, respectively. Similarly, for cross-validation purposes, 4/6, 12/16, 11/18, 17/17, and 18/18 cases were predicted correctly. We infer that for SB and CD, SOM has correctly predicted 17/17 and 18/18 cases each for both testing and cross validation. It is also clear that for SO, the prediction accuracy rate is comparatively lower compared to other classes. It can be concluded that for SO, about 22% (4/18) of the data matches with CD, and about 16% (3/18) of the data matches with SI, which suggests that while performing cross validation, SOM is unable to distinguish between these fault classes efficiently. A possible reason for spalls on the outer race is that the vibration data collected contains more noisy data compared to other faults. Similarly, for SI about 12% (2/16) of the data matches with SO, and 12% (2/16) of the data matches with SB. Another possible reason for not correctly identifying these defects during cross validation may be due to over fitting; i.e. the optimization of the parameters for the SOM classifier is not done properly.

Table 4 shows the prediction accuracy when SVM is used as a classifier. For testing purposes, 6/6, 16/16, 18/18, 17/17, and 18/18 cases are predicted correctly for BND, SI, SO, SB, and CD, respectively. For cross validation, 6/6, 14/16, 14/18, 16/17, and 18/18 cases are predicted correctly for BND, SI, SO, SB, and CD, respectively. Thus, it is clear that the prediction accuracy of BND and CD is 100% for both testing and cross validation when SVM is used as a classifier. SI and SO are comparatively less accurately predicted. For SO, about 10% (2/18) of the data falls under CD. This means that 10% of the data of SO matches with the CD class. Similarly, 10% (2/18) of the data matches with the SI class during cross validation, which is an indication that SO contains more noisy data compared to other classes. Similarly, for SI about 13% (2/16)

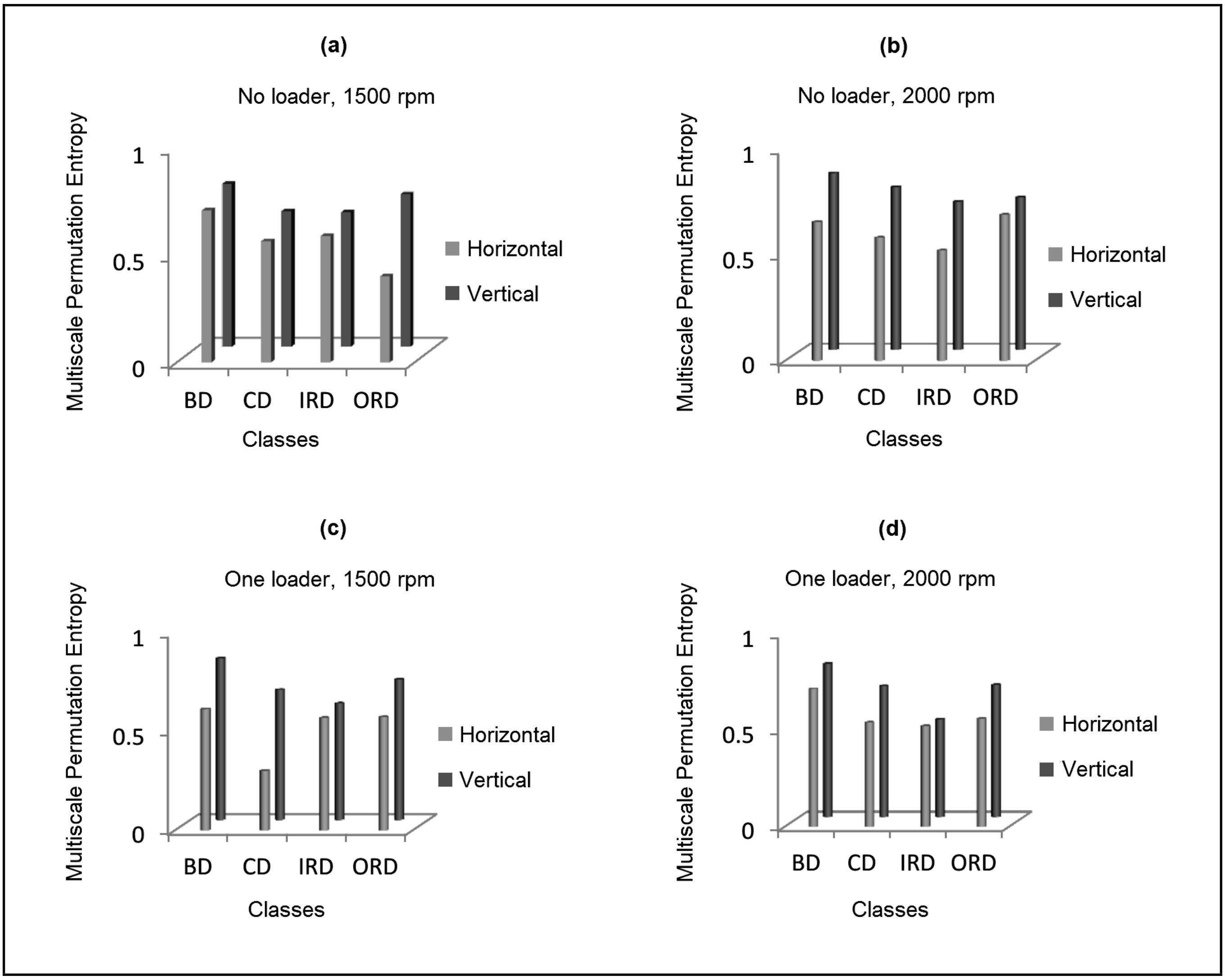

Figure 4. Multiscale permutation entropy with load condition and speed.

Table 3. Confusion matrix for SOM.

Using Test Set						Using 10 fold cross validation					
BND	SI	SO	SB	CD	Classified as	BND	SI	SO	SB	CD	Classified as
6	0	0	0	0	BND	4	0	1	0	1	BND
0	16	0	0	0	SI	0	12	2	2	0	SI
0	2	12	0	4	SO	0	3	11	0	4	SO
0	0	0	17	0	SB	0	0	0	17	0	SB
0	0	0	0	18	CD	0	0	0	0	18	CD

Table 4. Confusion matrix for SVM.

Using Test Set						Using 10 fold cross validation					
BND	SI	SO	SB	CD	Classified as	BND	SI	SO	SB	CD	Classified as
6	0	0	0	0	BND	6	0	6	0	6	BND
0	16	0	0	0	SI	0	14	0	2	0	SI
0	0	18	0	0	SO	0	2	14	0	2	SO
0	0	0	17	0	SB	0	0	0	16	1	SB
0	0	0	0	18	CD	0	0	0	0	18	CD

Table 5. Classification accuracy.

Parameters	SOM		SVM	
	Test set	10 fold cross validation	Test set	10 fold cross validation
Correctly classified instances	69 (92%)	63 (84%)	75 (100%)	68 (90.667%)
Incorrectly classified instances	6 (8%)	12 (16%)	0 (0%)	7 (9.33%)
Kappa statistic	0.8977	0.8161	1	0.8806
Total number of instances	75	75	75	75

Table 6. A comparative study between the presented work and published literature.

References	Machine Learning Method used	Faults considered	Efficiency of classification (%)	Techniques used for vibration analysis	Remarks
Kankar, et al.[29]	SOM, ANN, SVM	Spall in inner race, outer race, rolling element, and combined component fault, healthy bearing	70.66, 89.33, 90.66 by SOM, ANN, and SVM, respectively (cross validation)	Meyer, coiflet5, symlet2, gaussian, complex morlet, and shannon wavelet	Wavelets are compared
Kankar, et al.[30]	ANN, SVM	Spall in inner race, outer race, rolling element, and combined component fault, healthy bearing	71, 74 by ANN and SVM, respectively (test set)	NA	Time series data used
Seker, et al.[31]	NA	Fault at ball, inner race, outer race	71.33	Daubechies 15 and 20	Time series data used
Abbasion, et al.[32]	SVM	Bearing looseness, defects in rolling elements and bearing raceways	100 by SVM	Meyer wavelet	Wavelet denoising
Wu, et al.[23]	SVM	Fault at ball, inner race, outer race, and normal bearing	97–100 by SVM (training)	NA	Time series data used
Proposed work	SOM, SVM	Inner race, outer race, ball, combined fault, and bearing with no defect	92, 100 (test set) and 84, 91 (cross validation) by SOM and SVM, respectively	coiflet2	Best scale is selected using Permutation Entropy criterion

of the data matches exactly with SB, which also indicates that the classifier is not able to distinguish between them during cross validation.

Table 5 depicts the overall classification accuracy for both the test set and the 10-fold cross validation set using two classifiers. It can be observed that, for SOM, 69/75 test instances were classified correctly, which gives a 92% classification accuracy rate, and for 10-fold cross validation, 63/75 test instances were classified correctly, giving an 84% classification accuracy rate. Similarly, for SVM, 75/75 test instances were classified correctly, giving a 100% classification accuracy rate, and for 10-fold cross validation, 68/75 test instances were classified correctly, giving a 90.66% classification accuracy rate, as shown in Table 5.

The kappa statistic is used for assessing the degree to which two or more classes that are testing the same data match when it comes to assigning the data to classes. For complete matching, the corresponding value of the kappa statistic is 1, and for totally incomplete matching, its value is 0. For SVM, using a testing set, the ideal value of 1 is achieved for the kappa statistic. It is clear from the above mentioned results that the classification accuracy of SVM is much better compared to SOM, and is reported by Kankar, et al. in their study.[29] SOM is a type of unsupervised learning method in which the objective is to identify hidden structures in unlabelled data. Since the inputs given are unlabelled, it becomes difficult for the learning algorithm to train itself to correctly identify that particular feature belonging to a specific class. This makes it quite difficult to correctly predict the given feature set, and thus the classification accuracy is lower. On the other hand, the classification accuracy of SVM is high because of its good generalization capability. For demonstrating the effectiveness of the proposed methodology, a comparative study between the present work and some published literature is shown in Table 6. The proposed work is compared in terms of the machine learning method used, faults considered on the bearings, classification efficiency, and the vibration analysis technique.

7. CONCLUSIONS

In the present study, a methodology is proposed for comparing supervised and unsupervised learning methods for fault diagnosis of bearings. Raw vibration signals of various fault categories are used and the concept of permutation entropy is applied for the best scale selection of wavelet coefficients. Features based on the best scale are extracted for both horizontal and vertical response conditions. In total, 8 features have been considered, including permutation entropy, form factor, and crest factor for both horizontal responses and vertical responses, along with the number of loaders and shaft rotation. The classification results of SOM and SVM are compared, and the results show that SVM is able to give much better results due to its better generalization capability. It is observed that severe vibration is observed for ball defects. The prediction accuracy rate of both learning algorithms is lower for outer race defects and higher for combined defects. It can be concluded that the proposed methodology based on scale selection using multiscale permutation entropy along with supervised and unsupervised machine learning techniques has potential for application to the development of real-time fault diagnosis systems.

REFERENCES

1. Tandon, N. and Choudhury, A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, *Tribology International*, **32**, 469–480, (1999).

2. Bonnardot, F., Randall, R. B., and Antoni, J. Enhanced unsupervised noise cancellation using angular resampling for planetary bearing fault diagnosis, *International Journal of Acoustics and Vibration*, **9** (2), 51–60, (2004).

3. Harsha, S. P., Nataraj, C., Kankar, P. K. The effect of ball waviness on nonlinear vibration associated with rolling element bearings, *International Journal of Acoustics and Vibration*, **11** (2), 56–66, (2006).

4. Sharma Aditya, Amarnath M., and Kankar P. K. Feature extraction and fault severity classification in ball bearings, *Journal of Vibration and Control*, (2014). DOI: 10.1177/1077546314528021

5. Wang, J., Cui, L., Wang, H., and Chen, P. Improved complexity based on time-frequency analysis in bearing quanti-

tative diagnosis, *Advances in Mechanical Engineering*, **11**, (2013). DOI: 10.1155/2013/258506

[6] Kankar, P. K., Sharma S. C., and Harsha, S. P. Vibration based performance prediction of ball bearings caused by localized defects, *Nonlinear Dynamics*, **69** (3), 847–875, (2012).

[7] Harsha, S. P. and Kankar, P. K. Nonlinear dynamic analysis of a complex rotor bearing system, *International Journal of Acoustics and Vibration*, **10** (1), 33–39, (2005).

[8] Pincus, M. Approximate entropy as a complexity measure, *Chaos*, **5** (1), 110–117, (1995).

[9] Signorini, M. G., Magenes, G., Cerutti, S., and Arduini, D. Linear and nonlinear parameters for the analysis of fetal heart rate signal from cariotocographic recordings, *IEEE Transactions on Biomedical Engineering*, **50** (3), 365–374, (2003).

[10] Li, X., Cui, S., and Voss, L. J. Using permutation entropy to measure the electroencephalographic effects of sevoflurane, *Anesthesiology*, **109**, 448–456, (2008).

[11] Li, X., Ouyang, G., and Richards, D. A. Predictability analysis of absence seizures with permutation entropy, *Epilepsy Research*, **77**, 70–74, (2007).

[12] Quiroga, R. Q., Arnhold, J., Lehnertz, K., and Grassberger, P. Kulback-Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients, *Physical Review E*, **62**, 8380–8386, (2000).

[13] Rolo-Naranjo, A. and Otero, M. E. M. A method for the correlation dimension estimation for on-line condition monitoring of large rotating machinery, *Mechanical Systems and Signal Processing*, **19**, 939–954 (2005).

[14] Ding, R. Q. and Li, J. P. Nonlinear finite-time lyapunov exponent and predictability, *Physics Letter A*, **364**, 396–400, (2007).

[15] Yan, R. and Gao, R. X. Approximate entropy as a diagnostic tool for machine health monitoring, *Mechanical Systems and Signal Processing*, **21**, 824–839, (2007).

[16] Zhang, L., Xiong, G., Liu, H., Zou, H., and Guo, W. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference, *Expert Systems and Applications*, **37**, 6077–6085, (2010).

[17] Yan, R., Liu, Y., and Gao, R. X. Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines, *Mechanical Systems and Signal Processing*, **29**, 474–484, (2011).

[18] Bandt, C. and Pompe, B. Permutation entropy: A natural complexity measure for time series, *Physical Review Letters*, **88**, 174102, (2002).

[19] Tiwari, R, Gupta,V. K., and Kankar, P. K. Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier, *Journal of Vibration and Control*, (2013). DOI: 10.1177/1077546313490778

[20] Nair, U., Krishna, B. M., Namboothiri,V. N. N., and Nampoori, V. P. N. Permutation entropy based real-time chatter detection using audio signal in turning process, *International Journal of Advanced Manufacturing Technology*, **46**, 61–68, (2010).

[21] Bruzzo, A. A., Gesierich, B., Santi, M., Tassinari, C. A., Birbaumer, N., and Rubboli, G. Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients: A priminary study, *Neuro Science*, **29**, 3–9, (2008).

[22] Costa, M., Goldberger, A. L., and Peng, C. K. Multiscale entropy analysis of complex physiological time series, *Physical Review Letters*, **89**, 068102–1–4, (2002).

[23] Wu, S. D., Wu, P. H., Wu, C. W., Ding, J. J., and Wang, C. C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine, *Entropy*, **14**, 1343–1356, (2012).

[24] Vakharia,V., Gupta,V. K., and Kankar, P. K. A Multiscale entropy based approach to select wavelet for fault diagnosis of ball bearings, *Journal of Vibration and Control*, (2014). DOI: 10.1177/1077546314520830

[25] Zanin, M., Zunino, L., Rosso, O. A., and Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review, *Entropy*, **14**, 1553–1577, (2012).

[26] Vapnik, V. N., *Statistical learning theory*, John Wiley and Sons, (1998).

[27] Haykin, S. *Neural Networks: A comprehensive foundation*, Pearson Prentice Hall Publications, Ontario, Canada, (2005).

[28] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The WEKA data mining software: An update, *SIGKDD Explorations*, **11** (1), (2009).

[29] Kankar, P. K., Sharma, S. C., and Harsha, S. P. Fault diagnosis of ball bearings using continous wavelet transform, *Applied Soft Computing*, **11**, 2300–2312, (2011).

[30] Kankar, P. K., Sharma, S. C., and Harsha, S.P. Fault diagnosis of ball bearings using machine learning methods, *Expert Systems with Applications*, **38**, 1876–1886, (2011).

[31] Seker, S. and Ayaz, E. Feature extraction related to bearing damage in electric motors by wavelet analysis, *Journal of Franklin Institute*, **340** (2), 125–134, (2003).

[32] Abbasion, S., Rafsanjani, A., Farshidianfar, A., and Iranic, N. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*, **21**, 2933–2945, (2007).