Use of Feature Ranking Techniques for Defect Severity Estimation of Rolling Element Bearings

Aditya Sharma, M. Amarnath and Pavan Kumar Kankar

Machine Dynamics and Vibration Lab, Mechanical Engineering Discipline, PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Jabalpur-482005, India.

(Received 29 June 2015; accepted 28 June 2016)

Bearings are the most common components used in rotating machines. Their malfunction may result in costly shutdowns and human causalities which can be avoided by effective condition monitoring practices. In present study, attempt has been made to estimate the severity of defect in bearing components by a two-step process. Initially, defects of various severities in all bearing components are classified. In the next step, if defect exist in any of the bearing components, i.e. inner race, outer race, and rolling elements, level of severity of defect is estimated. Various statistical features are extracted from the raw vibration signals. Two machine learning techniques; support vector machine and artificial neural network, along with four feature ranking techniques; Chi-square, gain ratio, ReliefF and principal component analysis are used and employed for the analysis. Results show the potential of the proposed methodology in defect severity estimation and classification of rolling element bearings.

NOMENCLATURE

C	Penalty parameter
c	Number of classes
O_{ij}	Observed value in the i^{th} interval
E_{ij}	Expected frequency of O_{ij}
N	Number of instances
n	Number of samples
p	Distance measurement
$Z_{t,i}$	Value of instance x_i on feature f_i
ξ_i	Slack variable
J	Number of elements in a neuron
w_i	Interconnection weights of vector v_i
b	Bias for the neuron

1. INTRODUCTION

Rolling element bearings are the backbone of almost all the rotating machinery. Studies show that around 40% of the failures in rotating machines are due to bearing faults.[1] If the defect severity is diagnosed well in advance, bearing failure and thus machinery shutdowns can be reduced significantly by avoiding catastrophic failure. Various online health monitoring (OHM) techniques[2] are available which respond as the fault initiates, but it is impossible to estimate the defect severity at fluctuating speed and load during operation. Having various techniques, vibration based condition monitoring techniques are preferred due to ease of use and higher responsive towards the faults.

Various remarkable vibration based fault diagnosis methodologies have been developed for bearings.[3–6] However, in these studies, authors have not considered the severity of faults in their analysis. Classification and estimation of the specific defect is an important part of machinery maintenance systems. Faults with deferent severity levels in the same component give same characteristic frequency which makes the estimation of defect severity a challenging task. Inaccurate defect severity classification misleads the maintenance program. Various authors have classified the single level severity in rolling element bearing with 100% classification accuracy. Saxena and Saad,[7] Wu et al.,[8] and Liu et al.[9] have classified single level severity in rolling element bearing with 100% classification accuracy. Few attempts have been reported in the literature which attempts to classify various defect sizes in the same component with higher classification accuracy. Skewness, kurtosis, standard deviation, crest factor, and other statistical measures have been utilized by Sharma et al.[10] and Amarnath et al.[11] In order to increase the computational efficiency, feature ranking techniques are used to select the appropriate features which contain most significant information about the system.[12–15] In their extensive study, Zhao et al.[12] proposed various feature selection techniques; such as Chi-square, ReliefF, etc. Samanta et al.[13] have employed the genetic algorithm for condition monitoring of machines. The concept of mutual information has been applied by Kappaganthu and Nataraj[14] and the authors concluded that the fault detection accuracy improved significantly with the use of feature ranking methods. Sharma et al.[15] have examined various feature ranking techniques for the analysis of bearing faults and summarized that performance of analysis can be improved in the presence of ranked features.

Catastrophic failure of the bearing and the associated system can be reduced significantly with known defect severity. Hong and Liang[16] and Wang et al.[17] used the Lempel-Ziv complexity and continuous wavelet transform (CWT) based model to quantify the defect severity. The authors conclude that the Lempel-Ziv measure, as non-dimensional index, can be used for fault severity estimation. Jiang et al.[18] have extracted residual signals and statistical features from the conducted experiments to quantify the defect severity. The multi-frequency band energies (MFBEs) are also extracted from the acquired signals and summarize that varying trend of residual signals can be a useful tool for defect severity estimation. Yaqub et al.[19] presented a defect severity estimation model based on wavelet packet decomposition and support vector machine (SVM). The authors also extracted various statistical features for the severity estimation. Moshou et al.[20] have extracted statistical features for the defect severity estimation in rolling element bearing. The authors have quantified the defect severity by graphical representation of self-organizing maps (SOMs).

This paper presents a new methodology for defect severity classification and defect severity estimation of rolling element bearings. Four feature ranking techniques are used to select the most appropriate features. The selected features are further used as an input to two machine learning techniques, support vector machine (SVM), and artificial neural network (ANN) for classification and estimation purposes.

2. FEATURE RANKING TECHNIQUES

A number of features are extracted from raw signals to interpret them in meaningful results. However, not all the features are equally important for a specific purpose. Thus, optimal feature selection is the important task in fault diagnosis and severity estimation. The objective of feature ranking techniques is to rank the features based on information and physical spacing. In this study four feature ranking techniques; Chi-square, gain ratio (GR), ReliefF and principal component analysis (PCA), are employed to select the most appropriate features from the extracted features. The selected features are fed as input to the machine learning techniques for defect classification and defect severity estimation analysis. Feature ranking techniques used in this study are described as follows:

2.1. Chi-square

Chi-square is a very commonly used feature selection method. It evaluates the importance of a feature with respect to the class by calculating the value of Chi-squared statistic. Mathematically;

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}. \tag{1}$$

The necessary steps for the analysis using Chi-square technique are summarized as follows:[21]

1. Calculate the Chi-square value for every pair of adjacent intervals in a signal using Eq. (1).
2. It considers a high significance level for all numeric attributes for discretization.
3. A significance level for each of the attribute ($a = 1, \ldots, n$) is calculated and merged.
4. The consistency checking is performed after each attribute's merging.
5. Consistency checking is conducted to ensure that the discretized data set accurately represent the original one.
6. If the inconsistency rate is not exceeded, significance level (a) is determined for attribute a's next round of merging; otherwise attribute a will not be involved in further merging.
7. The process is continued until no attribute's value is merged to only one value. When the discretization ends, feature ranking is accomplished.

Chi-square discretize the relevant attributes and remove irreverent attributes. It automatically select the Chi-square value, determine the interval of numeric attribute as well as select features according to the characteristics of the data. It ensures that the fidelity of the training data can remain after Chi-square is applied. Chi-square feature ranking technique is a useful and reliable tool for discretization and feature selection of numeric attribute.

2.2. Gain Ratio

Gain ratio is based on the principle of information gain. In GR, features are selected in an incremental manner based on the iteration and the iteration ends when a predefined number of features remain.[22] Higher GR value indicates the higher applicability of features in a feature set, as well as improves the information gain by taking the inherent information of a split into account and is expressed as;

$$Gain\ ratio = \frac{Information\ gain}{Splitted\ information} \tag{2}$$

where $Information\ gain = Unsplitted\ information - Splitted\ information$. Gain ratio is an entropy based feature selection technique and calculates the usefulness of a feature by evaluating the performance of feature randomly in its presence. In GR, features are ranked based on maximizing the feature's information gain with minimizing the number of its value. The GR values lies between the range $(0, 1)$, where higher GR value of a feature indicates its higher ranking in a feature set.[23]

2.3. ReliefF

ReliefF evaluates the worth of an attribute by frequently considering an instance and by taking the value of given attribute for the nearest instance of the same and different class. Basically, it is defined for the two-class problem, but can also be used for multiple class problems.[24] For two class problem ReliefF is;

$$RF(Z_i) = \frac{1}{2} \sum_{i=1}^{N} p(Z_{t,i} - Z_{dc(xi)}) - p(Z_{t,i} - Z_{sc(xi)}) \tag{3}$$

where, $Z_{dc(xi)}$ and $Z_{sc(xi)}$ indicates the value of i^{th} feature of nearest points to x_i with different and same class label, respectively.

ReliefF is a supervised feature ranking technique. It is employed in data preprocessing as a feature subset selection method. During the features evaluation process, a weight is assigned to each feature based on the ability of the feature to distinguish among the classes and selects those features whose weight exceed a predefined threshold as a relevant feature. The weight computation is executed based on the probability of nearest neighbors from two different classes having different values for a feature and the probability of two nearest neighbors of the same class having the same value of the feature.

The higher the difference between two probabilities represents the more importance of the feature.[23]

ReliefF feature ranking technique is more robust and can deal with the noisy and incomplete data. However, its larger computational complexity can reduce the efficiency.

2.4. Principal Component Analysis

Principal component analysis is one of the major linear unsupervised dimensionality reduction techniques. It tries to set the data point from a higher dimensional space to a lower dimensional space with keeping all of the important information intact.[25] It considers the eigenvector to evaluate the influence, to the feature extraction result of each feature element. In PCA, the eigenvector corresponding to a large eigenvalue is able to capture more information of samples.[26]

3. MACHINE LEARNING TECHNIQUES

A variety of machine learning techniques such as support vector machine,[3,4] artificial neural network,[27,28] fuzzy logic,[29,30] genetic algorithm,[31,32] and others have been successfully employed in many engineering applications. Among them, support vector machine (SVM) and artificial neutral network (ANN) are most widely used artificial intelligence (AI) techniques due to their proven outstanding performance on rolling element bearings applications.[4,33] These two supervised soft computing techniques are considered in this study.

3.1. Support Vector Machine

Support vector machine is a supervised machine learning method based on structural risk minimization principle derived in statistical learning theory. SVM is extensively used for classification and regression problems due to its high generalization performance, robustness, ability to model non-linear relationships, and potential to handle very large feature space.[34,35] For a two-class problem SVM can be formulated as following optimization problem;

$$\text{Minimize} \quad \frac{1}{2}||W||^2 + C\sum_{i=1}^{n}\xi_i \tag{4}$$

$$\text{Subject to} \quad \begin{cases} y_i\left(W^T x_i + q\right) \geq 1 - \xi_i \\ \xi_i \geq 0, \quad i = 1, 2, \ldots, n \end{cases}; \tag{5}$$

where x_i, y_i is the data set and q is a real constant.

The sequential minimal optimization (SMO) is an improved faster training algorithm, used for solving the dual problem arising from the derivation of the SVM.

3.2. Artificial Neural Network

Artificial neural network is a group of especially interconnected artificial nodes, called neurons. ANN is an adaptive system that changes its structure according to the information flows through the network. Having various architectures of ANN, multilayer feed forward back propagation algorithm is widely used for rotary machine elements.

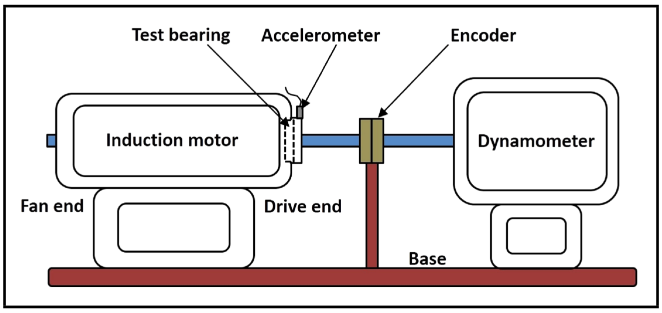

Figure 1. Schematic representation of experimental setup.

Table 1. Drive-end side test bearing specifications.

Parameter	Physical value
Bearing specification	6205-2RS JEM
Inner race diameter	25 mm
Outer race diameter	52 mm
Width	15 mm
Ball diameter	7.94 mm
Pitch circle diameter	39.04 mm
Contact angle	0°

A single neuron consists of synapses, summing function, and an activation function. Mathematically a neuron can be represented as:

$$K = Z\left(\sum_{i=1}^{J} w_i v_i + b\right). \tag{6}$$

4. EXPERIMENTAL SETUP

The bearing vibration data used for analysis in this study are collected from Case Western Reserve University Bearing Data Centre website.[36] Figure 1 shows the brief outlines of the experimental test rig. The test rig has a 2HP three phase induction motor, an encoder, and a dynamometer. The drive-end side of the motor consists the test bearing and is loaded by the dynamometer. Accelerometer, having magnetic base, is mounted on the housing of the test bearing and used for acquiring the vibration signals. Healthy bearing data are considered as the baseline data in the analysis. The drive-end side test bearing parameters used in this study are listed in Table 1. The schematic representation of various bearing components defects, i.e. inner race defect, outer race defect, and ball defect, are shown in Fig. 2.

Various single point bearing defects considered for the analysis are:

(i) Inner race defects having 0.1778 mm, 0.3556 mm and 0.5334 mm in diameter,

(ii) Outer race defects having 0.1778 mm, 0.3556 mm and 0.5334 mm at 6 o'clock position in diameter, and

(iii) Ball defects having 0.1778 mm, 0.3556 mm and 0.5334 mm in diameter.

5. FEATURE EXTRACTION AND SELECTION

A wide set of statistical features is extracted from the vibration signals. The extracted features are described as follows:

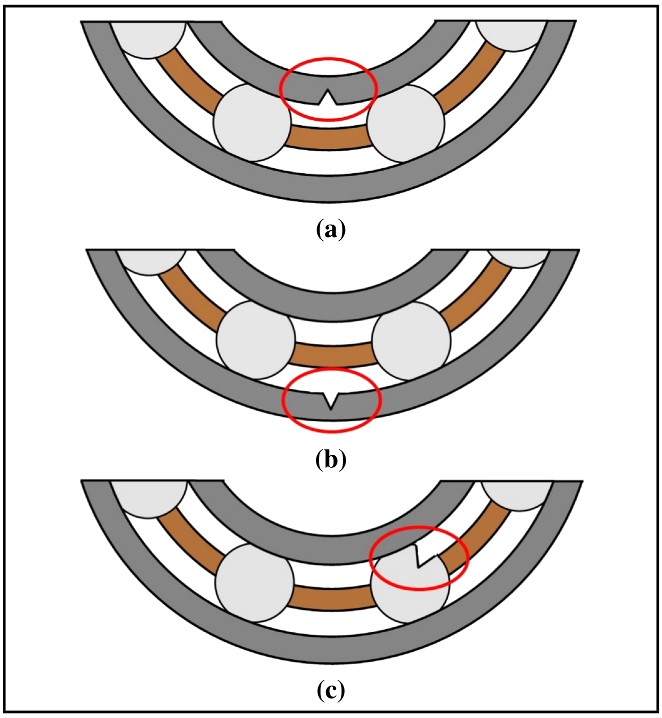

Figure 2. Schematic representations of defects in bearing components: (a) inner race defect, (b) outer race defect and (c) ball defect.

5.1. Mean

Mean is referred as the average value of the signal.

$$Mean = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{7}$$

5.2. Root Mean Square (RMS)

Root mean square is the square root of the average of the squared values of the signal.

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}. \tag{8}$$

5.3. Standard Deviation (SDEV)

Standard deviation is a measure of energy contain in the signal.

$$SDEV = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}. \tag{9}$$

5.4. Kurtosis

Kurtosis is used to describe the distribution of observed data around the mean and is defined as the degree to which a statistical frequency curve is peaked.

$$Kurtosis = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4}{(n-1)(SDEV)^4}. \tag{10}$$

5.5. Skewness

Skewness measures the symmetry of a distribution around its mean. Skewness can be negative or positive.

$$Skewness = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3}{(n-1)(SDEV)^3}. \tag{11}$$

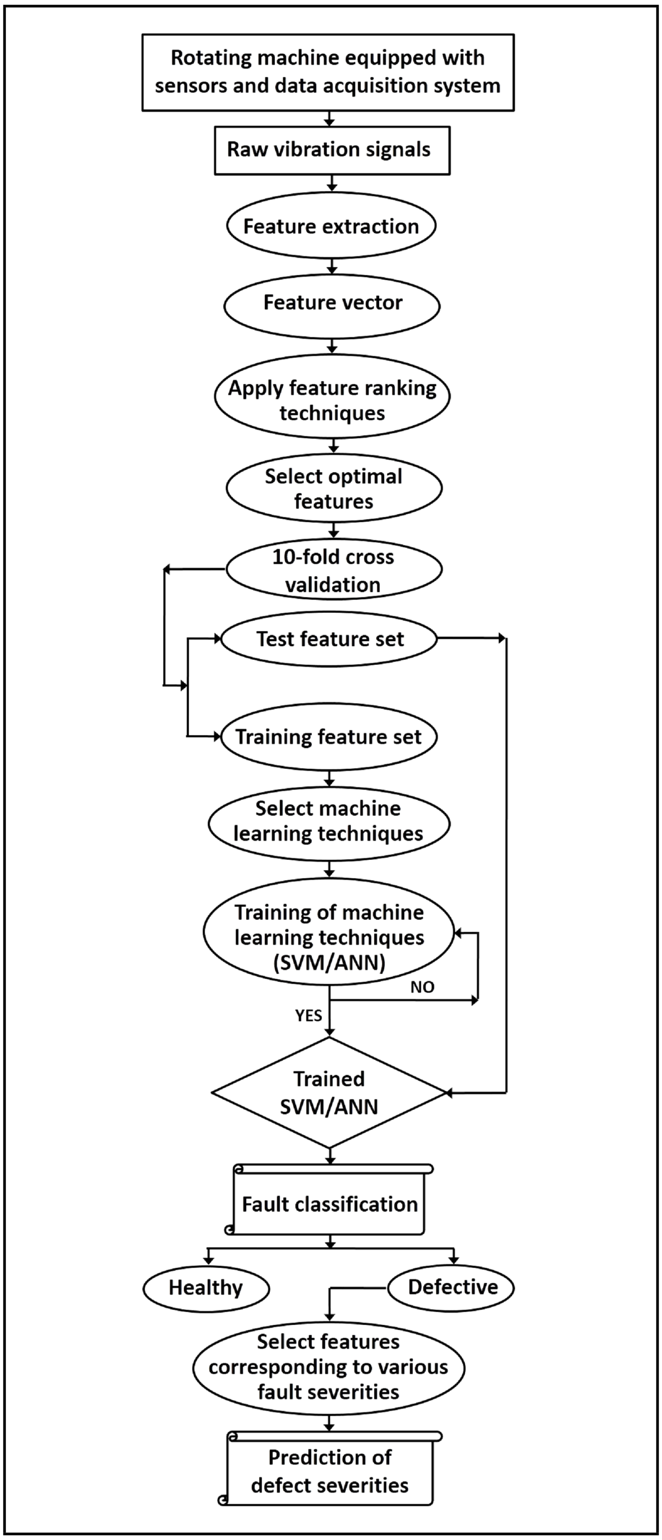

Figure 3. Overview of the methodology for multiclass defect severity classification and multiclass defect severity estimation.

5.6. Crest Factor

Crest factor is the ratio of peak value to RMS value of the signal and indicates the shape of the waveform.

$$Crest\ factor = \frac{\max |x_i|}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i)^2}}. \tag{12}$$

5.7. Minimum Value (MIN)

Minimum value represents the minimum value of the signal.

5.8. Maximum Value (MAX)

Maximum value represents the maximum value of the signal.

5.9. Covariance (COV)

Covariance is a measure that represents the strength of the correlation between two random variables in a signal.

5.10. Shape Indicator

Shape indicator is defined as the ratio of the RMS value to the mean value of the signal.

$$Shape\ indicator = \frac{RMS}{Mean}. \tag{13}$$

These statistical features are initially used to form a feature vector. To improve the defect classification and defect severity estimation efficiency, these extracted features are fed as input to various feature ranking techniques as discussed; thereafter, the features are shortlisted and selected as per their ranking. The selected features are then fed as input to machine learning techniques, i.e. SVM and ANN. The overview of the methodology for defect severity classification and defect severity estimation is shown in Fig. 3.

6. RESULTS AND DISCUSSION

In this study, defects in all bearing components, i.e. inner race, outer race and rolling element, with various defect severity levels, i.e. 0.1778 mm, 0.3556 mm and 0.5334 mm, and with healthy bearing, are considered for the fault diagnosis and defect severity estimation. Various statistical features are extracted from the considered bearing conditions. Further, features are selected as per their ranking using four feature ranking techniques. All the extracted features are supplied to four feature ranking techniques for their ranking. As suggested by Wang et al.,[37] $(\log_2 f_n)$ number of features may be used for classification with various learning algorithms, where f_n is the number of features. Table 2 summarizes the ranking of features corresponding to various feature ranking methods. A comparative study between SVM and ANN with all feature ranking techniques is carried out for defect classification and defect severity estimation.

6.1. Defect Classification

As a part of analysis, first the classification among all the considered cases is carried out, which includes the following forty bearing conditions: four corresponding to healthy bearing, twelve corresponding to inner race defects, twelve corresponding to outer race defect, and twelve corresponding to ball defect, having localized defects of 0.1778 mm, 0.3556 mm, and 0.5334 mm in inner race, outer race, and ball and each corresponding to four speeds, i.e. 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm. A sample training/testing vector used in the investigation is shown in Table 3 (where, HY = healthy bearing, ID = bearing having inner race defect, OD = bearing having outer race defect and BD = bearing having ball race defect).

The results for the two machine learning techniques, i.e. SVM and ANN, using 10-fold cross validation are shown. In 10-fold cross validation, data is randomly divided into ten equal sized training and testing folds. During iterations, nine of the 10-folds are used for training and remaining one fold is used for testing the dataset and finally it provides a single value after averaging all the iterations. 10-fold cross validation is preferred due to its capability of eliminating any biasness while dividing data into training and testing set. The detailed accuracy for SVM and ANN using the following four feature ranking techniques: Chi-square, GR, ReliefF, and PCA and these are shown in Table 4. It represents 100% classification accuracy and 0% incorrectly classified instances for each of SVM and ANN. Results also indicate the value of Kappa statistics for each of SVM and ANN with all the feature ranking techniques as 1 (or 100%), which indicates the perfect categorization of the data with the highest accuracy. Kappa statistics is an important measure which is used to predict the agreement between actual and predicted classes.[38]

6.2. Defect Severity Estimation

In the previous section, classification between defective inner race, outer race and rolling elements have been carried out. The classification accuracies of both SVM and ANN with all the feature ranking techniques are obtained as 100%. In this section, defect severities in bearing components are estimated. The estimation is carried out on three different defect severities, i.e. 0.1778 mm, 0.3556 mm, and 0.5334 mm of inner race, outer race, and rolling elements at four different speeds, i.e. 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm. A sample training/testing vector used for defect severity estimation purpose is shown in Table 5 (where IR = inner race, OR = outer race, and Ball = rolling element).

The detailed accuracies of defect severity estimation of SVM and ANN for inner race defects using various feature ranking techniques, i.e. Chi-square, GR, ReliefF, and PCA, are listed in Table 6. The correlation coefficients show a good agreement between the actual class and the predicted class, as its value is observed as 1 for all the ranking techniques for both of SVM and ANN. The maximum percentage error is reported as 0.2812% for ANN with PCA ranking technique. It indicates highly correlated results having very few errors.

The results of SVM and ANN for defect severity estimation of outer race defects with various feature ranking techniques are listed in Table 7. The results show the superior relationship between the actual class and the predicted class than that for inner race. For both of the artificial intelligence techniques, the correlation coefficient is observed as 1. Also, the maximum percentage error is reported as 0% for both of SVM and ANN. It shows perfectly correlated results for both SVM and ANN.

Table 8 indicates the results of defect severity estimation of SVM and ANN for rolling element with various feature ranking techniques. Results show that prediction capability of SVM is better than that of ANN. The correlation coefficient shows a perfect synchronization between the actual and predicted class for SVM while ANN has fewer prediction capabilities in this case. The maximum percentage error for SVM is observed as 0% and for ANN it is found as 3.3746%. The

Table 2. Ranking of features using different feature ranking techniques.

Feature ranking	Feature ranking technique			
	Chi-square	**Gain Ratio**	**ReliefF**	**PCA**
1	SDEV	SDEV	SDEV	Mean
2	RMS	RMS	RMS	RMS
3	MAX	MAX	Kurtosis	SDEV
4	Crest factor	Crest factor	MIN	Kurtosis
5	Shape indicator	Shape indicator	MAX	Skewness
6	Kurtosis	Kurtosis	COV	Crest factor
7	MIN	MIN	Mean	MIN
8	Mean	Mean	Shape indicator	MAX
9	COV	COV	Skewness	COV
10	Skewness	Skewness	Crest factor	Shape indicator

Table 3. Sample training/testing vector for SVM and ANN for fault severity classification.

	Features										Class
	Mean	RMS	SDEV	Kurtosis	Skewness	Crest factor	MIN	MAX	COV	Shape indicator	
Amplitude of features	0.0126	0.0738	0.0727	2.7643	-0.0362	4.2196	-0.2867	0.3113	0.0053	1.2399	HY
	0.0126	0.0664	0.0652	2.9306	-0.1731	5.213	-0.3459	0.3175	0.0043	1.2409	HY
	0.0135	0.2915	0.29131	5.3959	0.1641	5.9653	-1.3799	1.7391	0.0848	1.3957	ID
	0.0058	0.2929	0.2929	5.5423	0.1305	5.3973	-1.4029	1.5808	0.0858	1.41055	ID
	0.0232	0.6695	0.6691	7.6495	0.0569	5.4226	-3.4088	3.6305	0.4477	1.651	OD
	0.0041	0.5919	0.5919	7.5949	0.03341	5.2577	-3.0119	3.1123	0.3504	1.6153	OD
	0.0126	0.1393	0.1387	2.9847	-0.0089	4.3598	-0.6071	0.6039	0.0193	1.2526	BD
	0.0039	0.1391	0.1391	2.9638	0.0075	4.7434	-0.6597	0.6397	0.0194	1.2504	BD

maximum and minimum values of correlation coefficients for ANN are also noticed as 0.9997 and 0.999, respectively, which are very close to 1 and show good agreement between the actual and predicted class.

7. CONCLUSIONS

The present study deals with defect severity classification and estimation in various rolling element bearing components. Defects having three fault severities in inner race, outer race, and rolling elements are considered for the analysis. A wide set of statistical features is extracted from vibration signals. Four feature ranking techniques are used to rank the extracted features and the performance of two machine learning techniques, support vector machine and artificial neural network, are evaluated. The following conclusions are drawn from the present study:

- Both SVM and ANN show good performance for defect severity classification and estimation, but ANN performs a bit underneath for estimating the defect severity with principal component analysis feature ranking technique. The results obtained from SVM are superior due to its inherent capability of generalization.
- Results indicate that the two features of standard deviation and root mean square are proven to be the best two indicators irrespective of feature ranking method.
- The classification and quantitative assessment of fault severity of rolling element bearings can be improved significantly with feature ranking techniques. Results show that Chi-square method outperforms other techniques in terms of correlation coefficient as well as the minimum Maximum error (%).
- Proposed methodology can be effectively used for dimensionality reduction of features without compromising the performance and it would be beneficial in real practices to analyze the defect severities accurately.

ACKNOWLEDGEMENT

The authors would like to thank Prof. K.A. Loparo and Case Western Reserve University for their efforts to make bearing data set available and permission to use data set.

REFERENCES

[1] Singh, G. K. and Kazzaz, S. A. S. A. Induction machine drive condition monitoring and diagnostic research-a survey, *Electric Power Systems Research*, **64** (2), 145–158, (2003). https://dx.doi.org/10.1016/s0378-7796(02)00172-4

[2] Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., and Wu, B. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, John Wiley & Sons, Inc., Hoboken, New Jersey, (2006).

[3] Sugumaran, V., Muralidharan, V., and Ramachandran, K. I. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing, *Mechanical Systems and Signal Processing*, **21** (2), 930–942, (2007). https://dx.doi.org/10.1016/j.ymssp.2006.05.004

[4] Harsha, S. P., Nataraj, C., and Kankar, P. K. The effect of ball waviness on nonlinear vibration associated with rolling element bearings, *International Journal of Acoustics and Vibration*, **11** (2), 55–66, (2006). https://dx.doi.org/10.20855/ijav.2006.11.2191

[5] Singh, S. and Kumar, N. Combined rotor fault diagnosis in rotating machinery using empirical mode decomposition, *Journal of Mechanical Science and Technology*, **28** (12), 4869–4876, (2014). https://dx.doi.org/10.1007/s12206-014-1107-1

Table 4. Detailed accuracy of SVM and ANN for defect classification using different feature ranking techniques.

Feature ranking technique	Correctly classified instances		Incorrectly classified instances		Kappa statistics		Error (%)		Classification accuracy (%)	
	SVM	ANN	SVM	ANN	SVM	ANN	SVM	ANN	SVM	ANN
Chi-square	40	40	0	0	1	1	0	0	100	100
Gain Ratio	40	40	0	0	1	1	0	0	100	100
ReliefF	40	40	0	0	1	1	0	0	100	100
PCA	40	40	0	0	1	1	0	0	100	100

Table 5. Sample training/testing vector for SVM and ANN for fault severity estimation.

	Features												Defect severity (mm)
	Mean	RMS	SDEV	Kurtosis	Skewness	Crest factor	MIN	MAX	COV	Shape indicator	Speed (rpm)	Condition	
Amplitude of features	0.004	0.314	0.314	5.291	-0.0133	5.329	-1.536	1.672	0.098	1.399	1721	IR	0.1778
	0.004	0.163	0.163	21.686	0.0235	11.532	-1.88	1.854	0.026	1.652	1752	IR	0.3556
	0.003	0.449	0.449	8.345	0.303	8.056	-3.087	3.614	0.201	1.482	1728	IR	0.5334
	0.005	0.58	0.58	7.964	-0.003	5.576	-3.003	3.236	0.337	1.634	1725	OR	0.1778
	0.004	0.097	0.097	3.024	0.0002	4.932	-0.409	0.478	0.009	1.255	1749	OR	0.3556
	0.004	0.559	0.559	23.542	0.131	11.902	-6.654	6.653	0.313	2.054	1721	OR	0.5334
	0.004	0.154	0.154	2.84	0.02	4.69	-0.72	0.672	0.024	1.245	1722	Ball	0.1778
	0.005	0.144	0.144	9.753	0.144	12.819	-1.386	1.839	0.02	1.422	1749	Ball	0.3556
	0.005	0.118	0.118	3.101	0.025	4.886	-0.493	0.577	0.014	1.259	1729	Ball	0.5334

Table 6. Detailed accuracy of SVM and ANN for defect severity estimation of inner race using different feature ranking techniques.

Feature ranking technique	Correlation coefficient		Maximum error (%)	
	SVM	ANN	SVM	ANN
Chi-square	1	1	0	0
Gain Ratio	1	1	0	0
ReliefF	1	1	0	0
PCA	1	1	0	0.2812

Table 7. Detailed accuracy of SVM and ANN for defect severity estimation of outer race using different feature ranking techniques.

Feature ranking technique	Correlation coefficient		Maximum error (%)	
	SVM	ANN	SVM	ANN
Chi-square	1	1	0	0
Gain Ratio	1	1	0	0
ReliefF	1	1	0	0
PCA	1	1	0	0

Table 8. Detailed accuracy of SVM and ANN for defect severity estimation of rolling element using different feature ranking techniques.

Feature ranking technique	Correlation coefficient		Maximum error (%)	
	SVM	ANN	SVM	ANN
Chi-square	1	0.9992	0	0.2812
Gain Ratio	1	0.9992	0	3.0934
ReliefF	1	0.9997	0	1.9685
PCA	1	0.999	0	3.3746

[6] Singh, S. and Kumar, N. Rotor faults diagnosis using artificial neural networks and support vector machines, *International Journal of Acoustics and Vibration*, **20** (3), 153–159, (2015). https://dx.doi.org/10.20855/ijav.2015.20.3379

[7] Saxena, A. and Saad, A. Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems, *Applied Soft Computing*, **7** (1), 441–454 (2007). https://dx.doi.org/10.1016/j.asoc.2005.10.001

[8] 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** (8), 1343–1356, (2012). https://dx.doi.org/10.3390/e14081343

[9] Liu, Z., Cao, H., Chen, X., He, Z., and Shen, Z. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings, *Neurocomputing*, **99** (1), 399–410, (2013). https://dx.doi.org/10.1016/j.neucom.2012.07.019

[10] Sharma, A., Amarnath, M., and Kankar, P.K. Feature extraction and fault severity classification in ball bearings, *Journal of Vibration and Control*, **22** (1), 176–192, (2016). https://dx.doi.org/10.1177/1077546314528021

[11] Amarnath, M., Sugumaran, V., and Kumar, H. Exploiting sound signals for fault diagnosis of bearings using decision tree, *Measurement*, **46** (3), 1250–1256, (2013). https://dx.doi.org/10.1016/j.measurement.2012.11.011

[12] Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., and Liu, H. Advancing feature selection research-ASU feature selection repository, Technical Report, (2011).

[13] Samanta, B., Al-Balushi, K. R., and Al-Araimi, S. A. Artificial neural networks and genetic algorithm for bearing fault detection, *Soft Computing*, **10** (3), 264–271, (2006). https://dx.doi.org/10.1007/s00500-005-0481-0

[14] Kappaganthu, K. and Nataraj, C. Feature selection for fault detection in rolling element bearings using mutual information, *ASME Journal of Vibration and Acoustics*, **133** (6), 061001, (2012). https://dx.doi.org/10.1115/1.4003400

[15] Sharma, A., Amarnath, M., and Kankar P. K. Novel ensemble techniques for classification of rolling element bearing faults, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **39** (3), 709–724, (2016). https://dx.doi.org/10.1007/s40430-016-0540-8

[16] Hong, H. and Liang, M. Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform, *Journal of Sound and Vibration*, **320** (1–2), 452–468, (2009). https://dx.doi.org/10.1016/j.jsv.2008.07.011

[17] Wang, J., Cui, L., Wang, H., and Chen, P. Improved complexity based on time-frequency analysis in bearing quantitative diagnosis, *Advances in Mechanical Engineering*, **5**, 1–11, (2013). https://dx.doi.org/10.1155/2013/258506

[18] Jiang, F., Li, W., Wang, Z., and Zhu, Z. Fault severity estimation of rotating machinery based on residual signals, *Advances in Mechanical Engineering*, **4**, 1–8, (2012). https://dx.doi.org/10.1155/2012/518468

[19] Yaqub, M. F., Gondal, I., and Kamruzzaman, J. Machine fault severity estimation based on adaptive wavelet nodes selection and SVM, *Proc. of the IEEE International Conference on Mechatronics and Automation*, Beijing, China, (2011). https://dx.doi.org/10.1109/icma.2011.5986279

[20] Moshou, D., Kateris, D., Sawalhi, N., Loutridis, S., and Gravalos, I. Fault severity estimation in rotating mechanical systems using feature based fusion and self-organizing maps, *Proc. of Artificial Neural Networks-ICANN, 20th International Conference*, Thessaloniki, Greece, (2010). https://dx.doi.org/10.1007/978-3-642-15822-3_49

[21] Liu, H. and Setiono, R. Chi2: feature selection and discretization of numeric attributes, *Proc. of 7th IEEE International Conference on Tools with Artificial Intelligence*, Herndon, VA, USA, (1995). https://dx.doi.org/10.1109/tai.1995.479783

[22] Sharma, A. and Dey, S. Performance investigation of feature selection methods and sentiment lexicons for sentiment analysis, *International Journal of Computer Applications*, Special Issue on Advanced Computing and Communication Technologies for HPC Applications, 15–20, (2012).

[23] Novaković, J., Strbac, P., and Bulatović, D. Toward optimal feature selection using ranking methods and classification algorithms, *Yugoslav Journal of Operations Research*, **21** (1), 119–135, (2011). https://dx.doi.org/10.2298/yjor1101119n

[24] Robnik-Šikonja, M. and Kononenko, I. Theoretical and empirical analysis of ReliefF and RReliefF, *Machine Learning*, **53** (1–2), 23–69, (2003). https://dx.doi.org/10.1023/A:1025667309714

[25] Boutsidis, C., Mahoney, M. W., and Drineas, P. Unsupervised feature selection for principal components analysis, *Proc. of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, Nevada, USA, (2008). https://dx.doi.org/10.1145/1401890.1401903

[26] Song, F., Guo, Z., and Mei, D. Feature selection using principal component analysis, *Proc. of the International Conference on System Science, Engineering Design and Manufacturing Informatization*, Yichang, China, (2010). https://dx.doi.org/10.1109/icsem.2010.14

[27] Wu, J.-Da, and Liu, C.-H. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network, *Expert Systems with Applications*, **36** (3), 4278–4286, (2009). https://dx.doi.org/10.1016/j.eswa.2008.03.008

[28] Samanta, B. and Al-Balushi, K. R. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, *Mechanical Systems and Signal Processing*, **17** (2), 317–328, (2003). https://dx.doi.org/10.1006/mssp.2001.1462

[29] Liu, T. I., Singonahalli, J. H., and Iyer, N. R. Detection of roller bearing defects using expert system and fuzzy logic, *Mechanical Systems and Signal Processing*, **10** (5), 595–614, (1996). https://dx.doi.org/10.1006/mssp.1996.0041

[30] Mechefske, C. K. Objective machinery fault diagnosis using fuzzy logic, *Mechanical Systems and Signal Processing*, **12** (6), 855–862, (1998). https://dx.doi.org/10.1006/mssp.1998.0173

[31] Jiang, J. and Zhang, B. Rolling element bearing vibration modeling with applications to health monitoring, *Journal of Vibration and Control*, **18** (12), 1768–1776, (2011). https://dx.doi.org/10.1177/1077546311422241

[32] Zhang, Y. and Randall, R. B. Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram, *Mechanical Systems and Signal Processing*, **23** (5), 1509–1517, (2009). https://dx.doi.org/10.1016/j.ymssp.2009.02.003

[33] Tian, Z., Wong, L., and Safaei, N. A neural network approach for remaining useful life prediction utilizing both failure and suspension histories, *Mechanical Systems and Signal Processing*, **24** (5), 1542–1555, (2010). https://dx.doi.org/10.1016/j.ymssp.2009.11.005

[34] Widodo, A. and Yang, B.-S., Support vector machine in machine condition monitoring and fault diagnosis, *Mechanical Systems and Signal Processing*, **21** (6), 2560–2574, (2007). https://dx.doi.org/10.1016/j.ymssp.2006.12.007

[35] Üstün, B., Melssen, W. J., and Buydens, L. M. C. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel, *Chemometrics and Intelligent Laboratory Systems*, **81** (1), 29–40, (2006). https://dx.doi.org/10.1016/j.chemolab.2005.09.003

[36] Bearing vibration data set, Case Western Reserve University bearing data centre website, available at: http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website (Accessed May 2, 2015).

[37] Wang, H., Khoshgoftaar, T. M., and Gao, K. A comparative study of filter-based feature ranking techniques, *Proc. of the 11th IEEE International Conference on Information Reuse and Integration*, Las Vegas, Nevada, USA, (2010). https://dx.doi.org/10.1109/iri.2010.5558966

[38] Witten, I. H. and Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, (2005).