Indian Journal of Science and Technology, *Vol 9(33), DOI: 10.17485/ijst/2016/v9i33/101337, September 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Fault Diagnosis of Roller Bearings with Sound Signals using Wavelets and Decision Tree Algorithm

Abhijit D. Gaikwad[1*] , V. Sugumaran[1] and M. Amarnath[2]

[1]School of Mechanical and Building Sciences, VIT University, Chennai Campus, Chennai - 600127, Tamil Nadu, India; gaikwadabhijit.devidas2015@vit.ac.in; v_sugu@yahoo.com
[2]IIITDM Jabalpur, Jabalpur - 482005, Madhya Pradesh, India; amarnath.cmy@gmail.com.om

Abstract

Objectives: Use of an appropriate fault diagnosis methods alerts in advance about malfunctioning and failure of bearings. Vibration and Sound signals of rotating machines contain the dynamic information about their operating conditions. There are many articles reporting suitability of vibration signals for fault diagnosis applications; however, the transducers (accelerometers) and data acquisition equipment used for vibration signals analysis are costly. This prevents small scale industries and low cost equipment from using diagnostic tools on affordability ground. On the other hand, transducers used for acquiring sound signals (microphones) are relatively low cost or/and affordable. Hence, there is a need for studying the use of sound signal for fault diagnosis applications. This paper uses sound signals acquired from roller bearings in good and simulated faulty conditions for the fault diagnosis purpose. **Methods/Analysis**: Sound signals from bearings having defects on inner race and outer race have been considered for analysis. Since the characteristic sound signals of faulty bearings are complex and are struck in the noise and high frequency structural resonance, simple signal processing techniques cannot be used to detect bearing fault. Hence, wavelet features are used for extracting features from sound signals. The energy levels at various levels of wavelet decomposition are used to define features from sound signals. The most contributing features were selected and their classification is done using decision tree algorithm. This paper also discusses the effect of features, effect of various classifier parameters on classification accuracy. **Findings:** In feature classification of the fault signals the RBIO 2.4 wavelet has given the highest classification accuracy of 96.66%. Out of the 120 total instances, 116 (96.66%) were correctly identified while 4 instances were incorrectly classified with an error margin of (3.33%). **Application/Improvements**: An extensive investigation has been made by a J48 algorithm which produced better predictive performance than the other algorithms. The training and the optimization of J48 model with their essential parametric measures are reported. Based on the overall study, J48 with variation in number of objects (from 1 to 6) feature was found as the most successful classification algorithm that achieved the best classification accuracy of 96.66%. The classification accuracy of the proposed algorithm has been found better with only 4 misclassified features. The classification capability and the performance evaluation of J48 algorithm with confusion matrix and detailed classification accuracy is reported and discussed for further study.

Keywords: Bearings, Classification Accuracy, Decision Tree, Fault Diagnosis, Feature Selection, Sound Signals, Wavelet Features

1. Introduction

Roller bearings are essential components for rotary machines. Heavy workload on machines make-high loading and unloading of bearings. This frequent fluctuation in applied load leads to development of faults in bearings. The most common defects are produced by fatigue in material after long running time. This starts with development of minute cracks under the surface of bearings. During operation, these cracks progress to the

**Author for correspondence*

surface of bearings due to cyclic loads leading to surface spalling and pitting[1]. When the bearing is operating under various speeds and loads, it is difficult to measure the severity of localized faults directly; therefore, certain physical parameters such as vibration, sound, Acoustic Emission and wear debris have been considered in detection and diagnosis of incipient faults. Vibration and Acoustic Emission (AE) signals are widely used in condition monitoring of rotating machines[1,2]. The fault detection is possible by comparing the signals of a machine running in normal and faulty conditions. Statistically, 90% of the total amount of the bearing faults is related either to an inner race or the outer race, while the remaining are mostly due to a rolling element fault[3]. Hence, the main faults considered in the present study are inner race fault and outer race fault. The fault diagnosis is capable of successfully finding the failure of a component in a machine or system as well as predicts failure from their symptoms[4]. Basically fault diagnosis approach consists of three important stages *viz.* feature extraction, feature selection and feature classification. A lot of works have been carried out previously in feasibility study on diagnostic methods for detection of bearing faults. In a study, fault diagnosis of roller bearing using fuzzy classifier and histogram features with the focus on automatic rule learning has been presented[5]. In another study, statistical features extracted from the vibration signals for the brake fault diagnosis was reported[6]. The application of Discrete Wavelet Transform (DWT) has been considered in this study for the bearing fault diagnosis. Wavelet based analysis is an exciting new problem solving tool. Among all available time–frequency analysis methods, the wavelet transforms may be the best one and have been widely used for fault diagnosis. Recently, the application of Discrete Wavelet Transform has emerged in the context of bearing damage detection[7], spalling on the ball bearing[8] using vibration signals. In a study, a de-noising method based on Morlet wavelet transform have been proposed for finding faults in roller bearing[9]. The Shannon sampling theorem states that a high sampling rate is needed and subsequently, large size samples are required for the fault detection. The theorem applies to mathematical functions having Fourier Transform. Therefore, it is expected that the desired method should have good computing efficiency for discrete-time signals. The computation of Continuous Wavelet Transform (CWT) is somewhat time consuming and is not suitable for analysis of big size data and on-line fault diagnosis. The Discrete Wavelet Transform (DWT) is found to yield a fast computation of wavelet transforms. The DWT is very easy to implement for analysis and reduces the computation time and resources required. In the present study, the diagnosis of single and multiple ball bearing race faults have been investigated using Discrete Wavelet Transform. After feature extraction and feature selection, the classification of faults is an important process. There are various classifier algorithms are available to classify the faults in the roller bearings, to name a few, decision tree algorithm[10], Support Vector Machine (SVM)[11], Bayes classifier[12], fuzzy logic[13], best first tree[14] etc. The decision tree algorithm is one of the main classifiers used to determine the faults in the parts and components. J48 decision tree is one of the commonly used decision trees in fault diagnosis, which can do feature selection as well as feature classification. One of the main advantages of this classification algorithm is that they provide human-readable rules of classification. The classifier model has to be considered in a way, so that it should give higher classification accuracy with minimum training time. This paper deals about the fault diagnosis of bearings through sound signals using J48 decision tree and wavelet features.

2. Methodology

Sound signals were taken from the good bearings and from bearings with different faulty conditions with the help of a microphone. By analyzing these signals, feature extraction was done. J48 decision tree algorithm was used as feature selection and classification tool. The methodology followed in fault diagnosis process is explained in Figure 1.

3. Experimental Study

Experimental tests were carried out on a set of four bearings, all SKF R7 NB62, to record sound signals. A rolling bearing element consists of two rings, out of which the one in outer periphery is called the outer race way and the one in inner periphery of bearing called the inner race way. A set of rolling elements of cylindrical or spherical shape are rotating in between the tracks. The shape of rolling elements of bearing depends on design of bearing, size, shape and application. Initially bearings were fixed on the test rig. The motor is operated at a constant speed of 1200 rpm. The four conditions tested are 1. Healthy (Good) bearings, 2. Bearings with

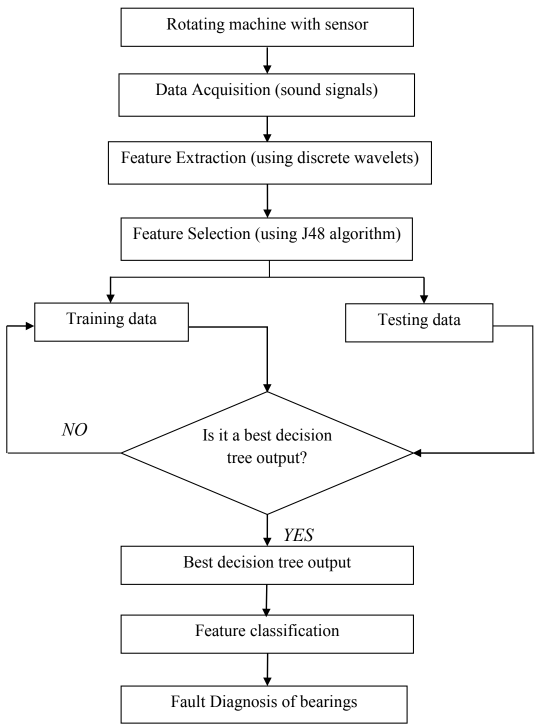

Figure 1. Floechart for roller bearing fault diagnosis.

One Inner Race fault, 3. Bearings with One Outer Race fault and 4. Bearings with Two Outer Race faults. These bearing faults were simulated using Electric Discharge Machining (EDM). The 'pits' were introduced in the inner and outer races of bearings. The size of the cylindrical pit is approximately 0.7 mm. During testing the sound signals were measured using a B&K 4117 microphone which was installed close to the test bearing. The sound signals were acquired using Agilent FFT analyzer and signals were sampled at a sampling frequency of 16.4 kHz. After the first test, healthy bearing was replaced by each defective bearing and then one by one signals were recorded for all the four cases separately, each one under the same operating conditions.

4. Feature Extraction

Feature extraction is simply the process in which different features are extracted from data which have been collected by transducer and stored. The sound signals were obtained from bearings, corresponding to respective good and faulty bearing conditions. The fault diagnosis was performed using the sound signal recorded from the bearing using the microphone. The sound signals which are in 'time-domain' were converted into 'time-frequency-domain' data by using Discrete Wavelet Transform (DWT) through wavelet decomposition. The wavelet decomposition results in the trend and details. Thus, obtained trends were again decomposed into next level trend and details. The same methodology was repeated for multiple levels of trends to give multiple levels of details. For the current study, a signal length of 8192 (2^{13}) was chosen and therefore, the signals can be decomposed into 13 levels. At each level, the detail co-efficient were used to compute energy content using the following formula.

$$V_i = \sum_{i=1}^{n} X_i^2$$

Where $X_{i\,=}$ details coefficients.

N = number of details coefficients.

Then the features were defined as the energy content at each level. The feature vector was defined as:

$$V = (V_1, V_2, V_3, \ldots\ldots\ldots, Vm)$$

Where m – (number such that length of signal) = 2^m

$V_1, V_2, V_3 \ldots$ are energy content at given level.

The following Discrete Wavelet Transformations were used in this study:

- Biorthogonal wavelet: BIOR – 1.1, 1.3, 1.5, 2.2, 2.4, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.5, 6.8
- Reverse biorthogonal: RBIO – 1.1, 1.3, 1.5, 2.2, 2.4, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.5, 6.8
- Coiflet: COIF – 1, 2, 3, 4, 5
- Daubechies wavelet: DB – 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- Sym wavelet: SYM – 2, 3, 4, 5, 6, 7, 8
- HAAR
- Discrete Meyer: DMEY

5. Feature Selection

The signals were processed using 54 different Discrete Wavelet Transforms from the seven wavelet families mentioned. The extracted features from each of the wavelet transform were then given as input to J48 algorithm to find the maximum classification accuracy. The details are presented in results and discussion section. Out of all the DWTs mentioned above, features extracted using RBIO 2.4 and RBIO 3.3 gave the best classification accuracy of 96.66% when used with J48 decision tree and used for the subsequent operations. Reverse biorthogonal wavelet, represented as 'RBIO n' is a family of compactly supported biorthogonal spline wavelets for which symmetry and exact reconstruction are possible with FIR filters.

6. Feature Classification: J48 Decision Tree

A decision tree is a tree based knowledge methodology used to represent classification rules. A standard tree induced with J48 algorithm consists of a number of branches, one root, a number of nodes and a number of leaves. One branch is a chain of nodes from root to a leaf and each node involves one attribute. The occurrence of an attribute in a tree provides the information about the importance of the associated attribute. J48 decision tree algorithm has two phases, the building phase and the pruning phase. In the building phase, J48 builds decision tree by using the concept of information theory. The tree has a single root node for the entire training set. For every division, a new node is added to the decision tree. J48 uses entropy based information gain as the selection criteria. As per information theory, entropy is a measure of the uncertainty in a random variable. The expected reduction in entropy due to the partitioning of the examples according to the given feature gives the information gain. It is a measure of the capability of a given attribute to separate its training examples according to the target function.

Information gain (S,A) of a feature A relative to a collection of examples S, is defined as:

$$Gain\ (S,A) = Entropy\ (S) - \sum_{V \in\ values\ (A)} \frac{|Sv|}{|S|}\ Entropy\ (Sv)$$

Where Sv = ({s ∈ S | A (s) = m}).

Entropy is a measure of homogeneity of the set of examples and it is given by:

$$Entropy\ (S) = \sum_{i=1}^{C} -P_i log_2 P_i$$

Where '*Pi*' is the proportion of '*S*' belonging to the class '*i*' and '*c*' is the number of classes. The second term in the equation above is the expected entropy after *S* is partitioned using feature *A*. When the data becomes large, the decision tree becomes large leading to more inaccuracy due to under-fitting or overtraining. Thus for better classification accuracy, the trees must be pruned to remove less reliable branches.

7. Results and Discussion

7.1 Feature Selection

Out of the total descriptive wavelet features selected for feature selection from sound signals, for classification purpose all of them may not be significant. As it cannot be predicted this of the parameters will be proved best for classification. Hence, extraction of all wavelet features was done and then good features which were contributing the most were selected. Here, decision tree was used for feature selection. The classification accuracy obtained for all the different wavelets is tabulated in Table 1. The fault diagnosis problem is analyzed using continues wavelet transform and J48 decision tree. Figure 2 shows decision tree of J48 algorithm for feature classification of RBIO 2.4 wavelet. The number of leaves of tree is 5 while the total size of the decision tree is 9. From Table 1, it can be

Table 1. Feature classification accuracy of different wavelets

Wavelet function	Accuracy %	Wavelet function	Accuracy %	Wavelet function	Accuracy %
BIOR 1.1	95.00	COIF 4	90.83	DB 1	95.00
BIOR 1.3	95.83	COIF 5	95.00	DB 2	92.50
BIOR 1.5	89.16	RBIO 1.1	95.00	DB 3	95.83
BIOR 2.2	90.00	RBIO 1.3	89.16	DB 4	89.16
BIOR 2.4	86.66	RBIO 1.5	90.00	DB 5	88.33
BIOR 2.6	88.33	RBIO 2.2	92.50	DB 6	91.66
BIOR 2.8	90.83	RBIO 2.4	96.66	DB 7	87.50
BIOR 3.1	89.16	RBIO 2.6	93.33	DB 8	94.16
BIOR 3.3	87.50	RBIO 2.8	90.00	DB 9	94.16
BIOR 3.5	90.00	RBIO 3.1	90.00	DB 10	92.50
BIOR 3.7	89.16	RBIO 3.3	96.66	SYM 2	92.50
BIOR 3.9	86.66	RBIO 3.5	88.33	SYM 3	94.16
BIOR 4.4	92.50	RBIO 3.7	87.50	SYM 4	94.16
BIOR 5.5	92.50	RBIO 3.9	89.16	SYM 5	87.50
BIOR 6.8	90.00	RBIO 4.4	95.83	SYM 6	91.66
COIF 1	92.50	RBIO 5.5	90.83	SYM 7	95.00
COIF 2	90.83	RBIO 6.8	86.66	SYM 8	92.50
COIF 3	88.33	DMEY	90.83	HAAR	95.00

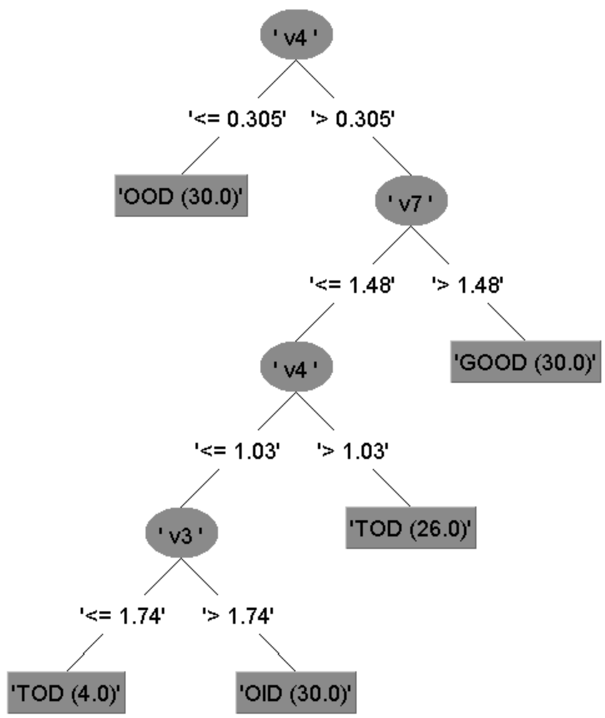

Figure 2. J48 decision tree for RBIO 2.4 wavelet.

seen that RBIO 2.4 has the highest feature classification accuracy of 96.66% out of all. Total 120 signals were collected from bearing for analysis with different conditions as good signals (GOOD), signals with One Inner Race Defect (OID), One Outer Race Defect (OOD) and Two Outer Race Defects (TOD).

7.2 Effect of Classifier Parameters Training on Classification Accuracy

Referring to the decision tree in Figure 2, the first feature V4 is contributing the most for classification (90.83%) and is called as root node. From Figure 2, it can be seen that only three features are enough to classify the bearing conditions viz. 'V4', 'V7' and'V3'. Then, effect of training of classifier parameters on classification accuracy was studied. First, the best feature 'V4' alone was used with decision tree and classification accuracy of 90.83% was obtained. Then top two features 'V4', 'V7' were selected and classification accuracy of 90.83% was obtained. Then next three top features 'V4', 'V7', 'V3' were selected and classification accuracy of 96.66% was obtained. This was repeated for all the 13 features and classification accuracy of 96.66% was obtained which is the highest. The classification accuracy increased from 90.83% to 96.66% for first three features 'V4', 'V7' and 'V3' and remained constant after that as 96.66% for all the remaining features. Hence, these features are successfully selected and classified. Moreover, the results are shown in Figure 3. For the feature classification with J48 algorithm only two classifier parameters can be trained, which can affect classification accuracy, viz. Minimum number of objects and Confidence Factor (CF). Hence, here both Minimum number of objects and Confidence Factor (CF) are varied and feature classification is tested with resultant accuracy. For variation in the minimum number of objects the objects were varied from 1 to 30 and the classification accuracy changed accordingly from 49.16% as lowest value to 96.66% as the highest. The Figure 4 shows variation of accuracy for change in minimum number of objects. For many of the objects (1-6, 14-21), accuracy value is same, indicating that there is no

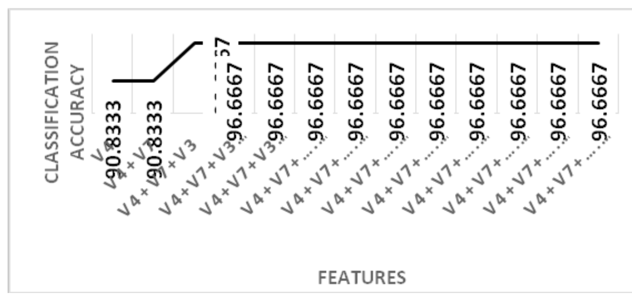

Figure 3. Contribtion of features in classification.

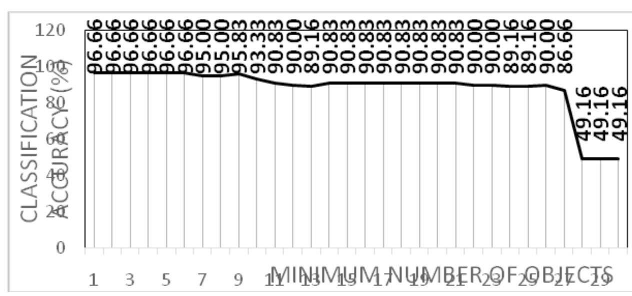

Figure 4. Variation of accuracy over minimum number of objects.

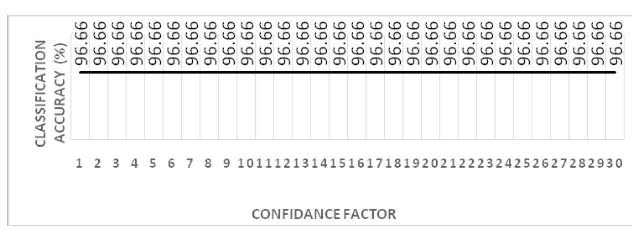

Figure 5. Variation of accuracy over confidence factor.

change in classification though there is change in minimum number of objects. The confidence factor was varied from 0 to 1 with the step size of 0.05. It can be seen from Figure 5 that the classification accuracy has not changed throughout the range of Confidence Factor (CF) from 0 to 1 and remained the same as 96.66%.

7.3 Detailed Accuracy and Confusion Matrix of Feature Classification

The confusion matrix of training 120 signals has shown in Table 2. For each tested bearing condition 30 signals have been taken. In the matrix, the diagonal elements represent the number of correctly classified instances. The diagonal elements in the confusion matrix show the number of correctly classified instances. The first element of the first row 29, shows the number of data points belonging to 'GOOD' class and correctly classified by J48 decision tree as 'GOOD'. The forth element of the same row 1, shows the number of data points belonging to 'GOOD' class but misclassified as TOD. This misclassification of features distinguishes between the number of elements which are actually contributing towards the feature classification and the elements which are not contributing in classification. The elements which do not contribute in classification lead to wrong feature classification. Similarly the second element in TOD column shows the number of data points misclassified as TOD but actually belong to OID class.

Total Number of Instances:	120	
Correctly Classified Instances:	116	96.67%
Incorrectly Classified Instances:	4	3.33%
Kappa statistic:	0.9556	
Mean absolute error:	0.0205	
Root mean squared error:	0.1283	
Relative absolute error:	5.4738%	
Root relative squared error:	29.6357%	
Number of Leaves:	5	
Size of the tree:	9	

Time taken to build model: 0.12 seconds

The class-wise detailed accuracy of J48 algorithm is presented in Table 3. 'TP rate' and 'FP' rate are very important. TP stands for True Positive and its value should be close to 1 for better classification accuracy. FP stands for False Positive and its value should be close to 0 for better classification accuracy. The True Positive (TP) rate explains the percentage of instances that are correctly classified and False Positive (FP) rate explains the misclassification percentage. Here, the TP rate value for GOOD class is 0.967 (Table 3), which is very close to 1, indicating better classification accuracy. The FP rate for GOOD class is 0.011, which is very low and close to 0. This implies that very low error in classification. Similarly, for other classes the TP rate is higher and close to value 1 and the FP rate is low and close to 0. The weighted average of TP rate for all classes is 0.967 Table 3, while for FP rate the weighted average value is as low as 0.011. Both the model confirms that build model is better one.

Table 2. Confusion matrix for J48 classification tree

GOOD	OID	OOD	TOD	
29	0	0	1	GOOD
0	**28**	0	2	OID
1	0	**29**	0	OOD

Table 3. Detailed accuracy of fault diagnosis by class

		Detailed	**Accuracy**	**by**	**class**	
Class	TP Rate	FP Rate	Precision	Recall	F - measure	ROC Area
GOOD	0.967	0.011	0.967	0.967	0.967	0.978
OID	0.933	0	1	0.966	0.966	0.978
OOD	0.967	0	1	0.983	0.983	0.983
TOD	1	0.033	0.909	0.952	0.952	0.979
Weighted Avg.	**0.967**	**0.011**	**0.969**	**0.967**	**0.967**	**0.98**

8. Conclusion

The fault diagnosis of the roller bearing was carried out using wavelet features and J48 decision tree and the results are presented along with the confusion matrix along with the detailed class wise accuracy. Since the J48 algorithm has given good accuracy (96.66%) in results hence it can be used for practical implementation. The remaining 3.33% error is low and can be neglected in case of large set of elements. The results obtained show a promising future in fault diagnosis of roller bearing applications. The results of the decision tree algorithm can be practically used for diagnosis of bearing the conditions successfully. Compared to vibration signals, sound signals (microphone) analysis is very cost effective and hence, the result will be more useful and acceptable.

9. References

1. Al-Badour F, Sunar M, Cheded L. Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques. Mechanical Systems and Signal Processing. 2011 Aug; 25(6):2083–101.
2. Al-Ghamd AM, Mba D. A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. Mechanical Systems and Signal Processing. 2006 Oct; 20(7):1537–71.
3. Sugumaran V, Ramachandran KI. Fault diagnosis of roller bearing using fuzzy classifier and histogram features with focus on automatic rule learning. Expert Systems with Applications. 2011 May; 38(5):4901–7.
4. Bentley D. Predictive maintenance through the monitoring and diagnostics of rolling element bearings. Bently Nevada Co. Application note. 1989; 44:2–8.
5. Sugumaran V, Ramachandran KI. Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. Mechanical Systems and Signal Processing. 2007 Jul; 21(5):2237–47.
6. Jegadeeshwaran R, Sugumaran V. Vibration based fault diagnosis study of an automobile brake system using K-STAR (K*) Algorithm – A Statistical approach. Recent Patents on Signal Processing. 2014 Apr; 4(1):44–56.
7. Nikolaou NG, Antoniadis IA. Roller bearing fault analysis using wavelet packets. NDT and E International. 2002; 35(3):197–205.
8. Mori K, Kasashima N, Yoshioka T, Ueno Y. Prediction of spalling on a ball bearing by applying Discrete Wavelet Transform to vibration signals. Wear. 1996 Jul; 195(1-2):162–8.
9. Lin J, Qu L. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. Journal of Sound and Vibration. 2000 Jun; 234(1):135–48.
10. Amarnath M, Sugumaran V, Hemantha Kumar R. Exploiting sound signals for fault diagnosis of bearings using decision tree. Measurement. 2013 Apr; 46(3):1250–6.
11. Sugumaran V, Ramachandran KI. Effect of number of features on classification of roller bearing faults using SVM and PSVM. Expert Systems with Applications. 2011 Apr; 38(4):4088–96.
12. Li Z, Zhu J, Shen X, Zhang C, Guo J. Fault diagnosis of motor bearing based on the Bayesian network. International Workshop on Automobile, Power and Energy Engineering. Procedia Engineering. 2011 Nov; 16:18–26.
13. Sakthivel NR, Sugumaran V, Nair BB. Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. Mechanical Systems and Signal Processing. 2010 Aug; 24(6):1887–906.
14. Jegadeeshwaran R, Sugumaran V. Comparative study of using decision tree and best first tree as a classifier for fault diagnosis in automobile hydraulic brake system using statistical features. Measurements. 2013 Nov; 46(9):3247–60.