Indian Journal of Science and Technology, Vol 9(33), DOI: 10.17485/ijst/2016/v9i33/101334, September 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Fault Diagnosis of Helical Gearbox through Vibration Signals using Wavelet Features, J48 Decision Tree and Random Forest Classifiers

Ayush Kimothi[1*], Ameet Singh[1], V. Sugumaran[1] and M. Amarnath[2]

[1]School of Mechanical and Building Sciences, VIT University, Chennai Campus, Chennai - 600127, Tamil Nadu, India; kimothiayush75@yahoo.in,ameetsinghsassan123@gmail.com, v_sugu@yahoo.com

[2]Indian Institute of Information Technology Design and Manufacturing Jabalpur, Jabalpur – 482005, Madhya Pradesh, India; amaranth.any@gmail.com

Abstract

Objective: Gearbox being the backbone of transmission system is designed and manufactured very carefully so that there is minimum compliance in the system. However, there are still faults and failures which usually occur in the system. The failure in helical gearbox is more prominent in bearings rather than gears which are the main components of the system. Gearbox is susceptible to failures because of reasons like misalignment, vibration and shocks. In this paper wavelet feature extraction is used along with random forest algorithm to diagnose faults in gearbox. The vibration signals were used for extracting wavelet features. Features were selected using J48 Decision Tree and were classified using random forest algorithm. A detailed study has been done to ensure that the optimum number of features was used and the factor was iterated so that maximum classification accuracy is obtained. The results are presented along with the conclusion. **Method Analysis:** The classification accuracy is obtained by 3 steps namely, feature extraction, feature selection and feature classification. By obtaining the Decision Tree the most important factors are selected to obtain maximum classification accuracy at minimum number of features to reduce calculations in real time application. The number of features and depth of data is iterated to obtain the maximum classification accuracy. **Findings:** Through this research random forest algorithm was tested for fault diagnosis of gearbox and a better classification accuracy was obtained. These results can be further used for fault diagnosis in industries for any gearbox related problems. **Application/Improvements:** An extensive investigation is done by a random forest algorithm which produced better forecasting than the other algorithms. Based on the overall study, random forest was found as the most preferred classification algorithm that achieved the best classification accuracy of 93.08% which is better than the other algorithms.

Keywords: Fault Diagnostics, Gearbox Fault Diagnostics, J48 Decision Tree, Machine Learning, Random Forest, Vibration Signals, Wavelet Feature Extraction

1. Introduction

In modern day production, the use of computers for improvement and continuity in production process is increasing. Fault in a system or machinery can delay the production process hence causing a harm of million dollars within hours to the production process. In this research, the aim is at using vibration signals for fault diagnostics in gearbox.

Gearbox is the soul of transmission system. If there is a fault in the gearbox, it will immediately effect the working of machinery and thereby the manufacturing process in an industry. The fault in gearbox reduces transmission efficiency and the performance. Gearbox is used to reduce the speed of output shaft thereby increasing the torque at low speed or vice-versa or just simply change the direction (reverse gear) i.e. adjusting the rotation of output shaft in power band or torque band. More torque

*Author for correspondence

is required to run vehicle at lower speed while power is the necessity at higher speeds. The study is aimed at forecasting the problem by analyzing vibration signals and giving the accuracy to which it is working by using wavelet features[1,2]. While operating naturally, the rotation of gears will create a vibration which will be different if the working deviates from natural i.e. the difference due to a fault in the system. The fault in gearbox generally occurs in the gear teeth and bearings which are susceptible to problems due to misalignment and jerks etc. Reading and manipulating the vibration pattern of gearbox will help to detect in prior any fault in the system. Fault diagnostics using vibration patterns consists of three steps namely feature extraction, feature selection and feature classification. There are three types of features which can be primarily used for such kind of study namely wavelet features[3], histogram features[4] and statistical features[5] out of which we have used the wavelet features. The technique for feature selection includes Principal Component Analysis (PCA)[6], Decision Tree (DT)[7], fuzzy and artificial neural network[8] and Genetic Algorithm (GA)[9,10] has explained the statistical condition indicators like RMS, nRMS etc. do not reveal the linear increase over the wind turbine gear fault progress[10, 11]have obtained 100% accuracy by using Hilbert transform for fault diagnostics in spur gearbox[11,12] has analyzed broken system in gearbox diagnosis methods based on SVM and wavelet lifting. Their approach is hybrid approach for small sample sizes[12,13] have used vibrational signals for condition monitoring of a gearbox using statistical features and obtained good classification accuracy for different gears[13]. In have presened diagnostic method based on Bayesian networks. The proposed BN based diagnostic mechanism is capable of diagnosing faults with high accuracy[14]. In have done a comparative experimental study of the effectiveness of ANN (Artificial Neural Network) and SVM (Support Vector Machine)[15]. In proposed use of Binary Particle Swarm Optimization (BPSO) and Binary Genetic Algorithm (BGA) in feature selection process using different fitness functions in the field of bearing fault diagnosis[16]. In has presented the use of the J48 algorithm for fault diagnosis through discrete wavelet features extracted from vibration signals of good and faulty conditions of the components of a centrifugal pump and compared the classification accuracy of discrete wavelet families[17]. In have done Condition Monitoring using Wavelet Transform and Fuzzy Logic by Vibration Signals and obtained good accuracy[18]. The features were extracted and selected using visualization tree under J48 algorithm. Then, they were classified using random forest algorithm and the classification accuracy was thus obtained.

The procedure for current work is as follows:

- Experiment is conducted so as to obtain the required vibration signal with different faults.
- The features are extracted using MATLAB.
- The extracted features were classified using J48 Decision Tree which is also known as feature selection.
- In the order of importance of various factors random forest algorithm was used for feature classification and analyzes effect of that feature on overall accuracy.

2. Experimental Study

Test rig was set up to get the vibration readings and conduct the study of the gearbox. The details of the test rig and set up are given in subsection.

2.1 Experimental Setup

The experimental setup consists of a 5 HP, 2 stage helical gearbox which is driven by a 5.5 HP, 3 phase induction motor that is rated at 1440 rpm. Inverter drive is used to control it and using that it is operated at a reduced speed of 80 rpm. As the step up ratio is 1:15, the speed of pinion shaft in second stage of generator is obtained to be 1200 rpm[19]. Table 1 gives the specifications of the test rig. The pinion gear is connected to a DC motor which acts as a power generator to generate 2 kW power, which is dissipated in a resistor bank. Hence, the acting load on the gearbox is only 2.6 HP which is 52% of the rated power of motor i.e. of 5 HP.

In industries the utilization of load varies from 50%-100%. In the case of traditionally used dynamometer, the additional torsional vibrations can occur due to torque fluctuations. This is avoided in this study by using D.C motor and resistor bank. Figure 1 shows the experimental setup used for the study. To restrict the backlash in the system to gears, tire couplings are fitted between the electrical machines. The motor, the gearbox and the generator are mounted on I-beams, which are strongly anchored to a massive foundation so as to provide rigidity to the system and thereby reduce the unnecessary vibrations. Vibration signals are measured using a Bruel and Kjaer accelerometer which is installed close to the test bearing

to reduce gathering of unnecessary information. Signals are sampled at a sampling frequency of 8.2 kHz. The overhaul time of a new gearbox is more than one year which is not suitable for the study. It is very difficult to study the fault detection procedures without seeded fault trials i.e. without introduction of faults. Local faults in a gearbox can be classified into three categories. 1. Cracked tooth, 2. Surface wear and 3. Loss of a part of tooth due to breakage of tooth at root or at a point on working tip (broken tooth or chipped tooth). There are different methods to simulate faults in gears i.e. by grinding and adding iron particles in gearbox lubricant, Electric Discharge Machining (EDM) and over loading the gearbox i.e., accelerated test condition. The simplest approach is partial tooth removal. This simulates the partial tooth break, which is most common in many industrial applications[20–23].

Table 1. Experimental setup specifications

	First stage	Second stage
Number of teeth	44/13	73/16
Pitch Circle diameter (mm)	198/65	202/48
Pressure angle (degree)	20	20
Helix angle (degree)	20	15
Modulus	4.5/5	2.75/3
Speed of shaft	80 rpm input	1200 rpm output
Mesh frequency	59 Hz	320 Hz
Step up ratio	1:15 5 HP	
Rated power		
Power transmitted	2.6 HP	

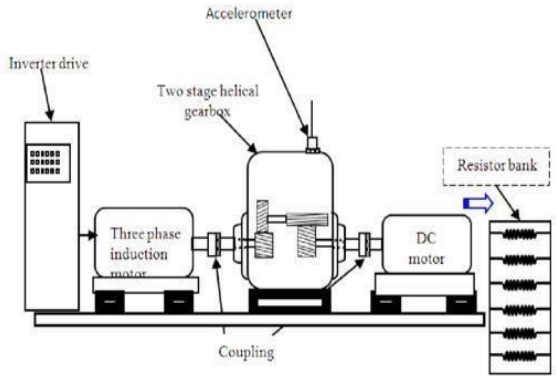

Figure 1. Experimental setup.

3. Methodology

3.1 Feature Extraction

The vibration signals taken are time domain signals which are to be converted into time frequency domain data. Using DWT (Discrete Wavelet Transform), time domain signals were converted into time-frequency domain data. Wavelet decomposition was performed on signals using DWT. The trends and details are the outcome of decomposition. For further details of the next level, the previous level domains are decomposed again. Continuing many levels of details in time-frequency domain data are obtained. The features were extracted using Symlet (Sym) features. At each level of time domain signal, the detail coefficients were used to compute the energy content using the following formulae:

$$Vi = \sum n\ i=1\ X\ i2 \quad (1)$$

The features were defined as energy content on each level. The features are in form of vectors i.e. V = V1, V2, V3 etc.

The results obtained were observed and classified out of which "SYM8" was the signal with highest classification accuracy. Families of wavelets taken into account for the fault diagnosis are:

- Haar wavelet.
- Discrete Meyer wavelet.
- Daubechies wavelet – Db1, db2, db3, db4, db5, db6, db7, db8, db9, db10.
- Biorthogonal wavelet – bior1.1, bior 1.3, bior 1.5, bior 2.2, bior 2.4, bior 2.6, bior 2.8, bior 3.1, bior 3.3, bior 3.5,bior 3.7, bior 3.9, bior 4.4, bior 5.5, bior 6.8.
- Reversed Biorthogonal wavelet - rbio1.1, rbio 1.3, rbio 1.5, rbio 2.2, rbio 2.4, rbio 2.6, rbio 2.8, rbio.
- 3.1, rbio 3.3, rbio 3.5, rbio 3.7, rbio 3.9, rbio 4.4, rbio 5.5, rbio 6.8.
- Coiflet – coif 1, coif 2, coif 3, coif 4, coif 5.
- Symlets – sym 2, sym 3, sym 4, sym 5, sym 6, sym 7, sym 8.

The methodology can be seen from Figure 2.

The wavelet selection is explained by subsection.

For wavelet selection, time domain signals were processed from seven different wavelet families using 54 discrete wavelets. The extracted features were classified using J48 Decision Tree algorithm using Weka 3.6 and

maximum classification accuracy was obtained which can be observed from Figure 3 to Figure 9.

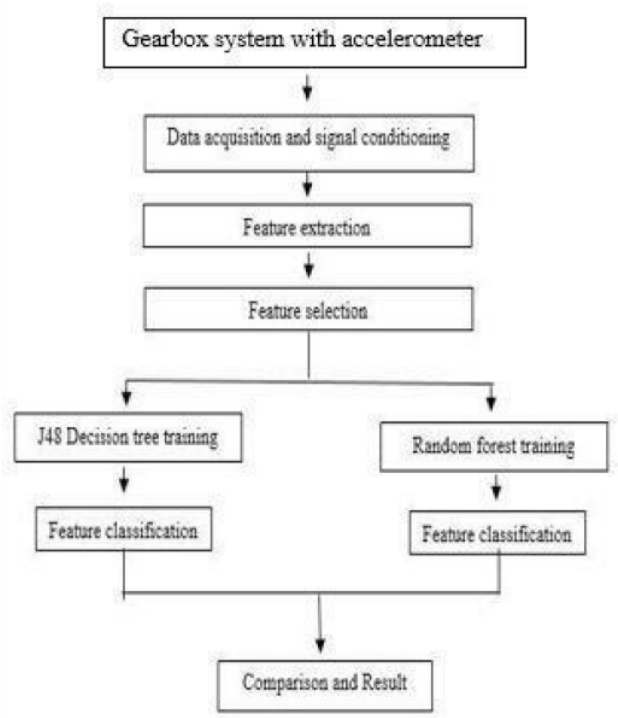

Figure 2. Methodology.

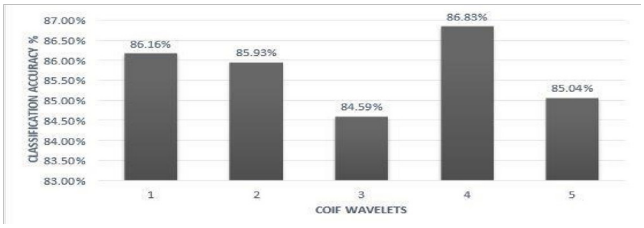

Figure 3. Classification accuracy of COIF wavelets.

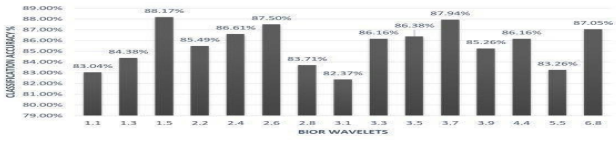

Figure 4. Classification accuracy of BIOR wavelets.

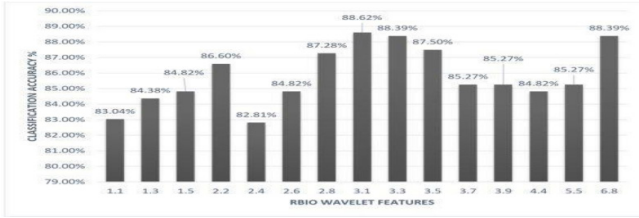

Figure 5. Classification accuracy of RBIO wavelets.

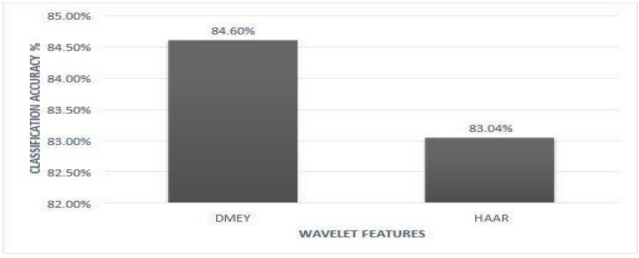

Figure 6. Classification accuracy of HAAR and DMEY wavelets.

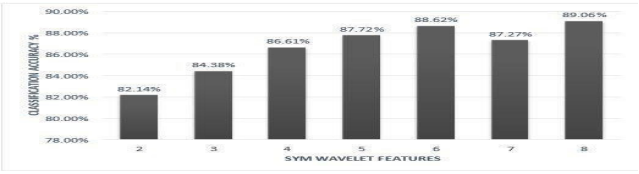

Figure 7. Classification accuracy of SYM wavelet features.

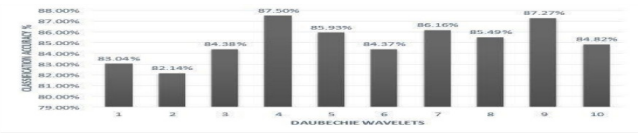

Figure 8. Classification accuracy of DAUBECHIE wavelet features.

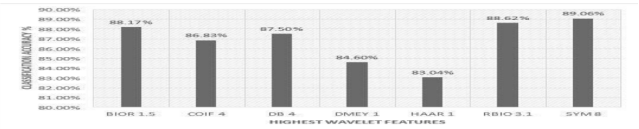

Figure 9. Comparison of classification accuracy of highest wavelet features.

3.2 Feature Selection using Decision Tree

- From the vibration signals, 13 wavelet features were obtained after feature extraction.
- There were 7 Sym features namely Sym2, Sym3, Sym4 etc. which were classified using J48 algorithm in Weka 3.6 and maximum classification accuracy was obtained which can be observed in Figure 10.
- The obtained signals were then classified using the j48 Decision Tree and the order of contribution of various factors in overall classification accuracy was obtained.
- Figure11 represents the Decision Tree obtained for the given data classified using j48 decision tree. As can be observed from the figure the top node is the root node and has the highest individual classification accuracy. In this study the

root node is (V3) with an individual classification accuracy of 60.268%.

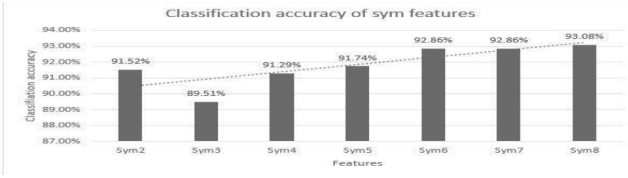

Figure 10. Classification accuracy of Symlet features.

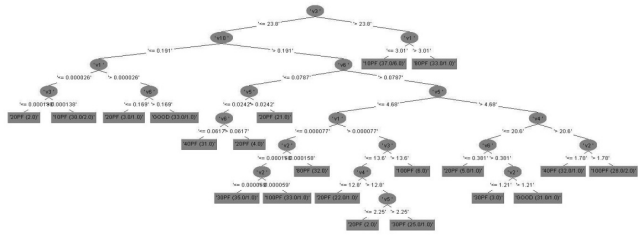

Figure 11. Decision Tree obtained from J48 classifier.

3.3 Feature Selection using Random Forest

Random forests algorithm consists of a combination of tree predictors such that each tree depends on the values of a vector which is sampled independently and with the same data distribution for all trees in the forest. The generalization of error for forests converges as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the classification of individual trees in the forest and the correlation between them[24,25]. Random tree creates assembly of trees during individual classification of data[26]. Because many trees are created, the problem of over fitting the variance is removed in random forest algorithm. For different range of the sets, different trees are grown. When the test variable is introduced, a mode value of the trees is calculated from different parts of the training set thus reducing the variance. Random forest classifier was used in Weka 3.6 to classify Symlet 8 features by varying the depth and the number of features used for study for obtaining the optimum data required for maximum classification accuracy.

4. Results and Discussions

The vibration signals were recorded for both abnormal and normal working of a gearbox. A total 448 samples were collected which were categorized as 100% accurate i.e. with no load condition, with 10% fault, 20% fault etc. These samples were then classified using Random forest algorithm.

4.1 Variation with Depth

To ensure that sufficient depth of data is under consideration the depth of the data was varied from 1-30. Selection of depth value less than this can provide us misleading results therefore depth value was selected to be 20 because it had highest classification accuracy and after this the classification accuracy became constant. Figure 12 represents the variation of classification accuracy with depth.

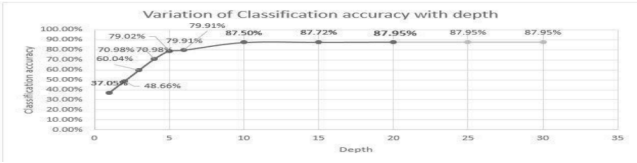

Figure 12. Variation of classification accuracy with depth of data.

4.2 Variation with Number of Features

Considering all the features lead to the consideration of unnecessary data which reduces the performance of classification algorithm. Also the number of computational resources required are also increased. To avoid this problem we checked the classification accuracy by varying the number of features for consideration. Keeping the depth as constant the number of features were varied from 1-14 and the maximum classification accuracy was observed at 4, Figure 13.

The Decision Tree provides us the order of importance of various features to the overall classification accuracy in the decreasing order. As can be observed from Figure 2 "V3" feature has highest individual accuracy followed by "V10" and "V1". This gives us the major classification factors for achieving maximum classification accuracy and removing unnecessary features i.e. the features having very less impact on overall classification accuracy. After fixing the depth and number of features, the features selected using feature selection process was separately analyzed and their contribution to overall classification accuracy was observed. The factors affecting the classification percentage were individually classified and their effect on the classification percentage was calculated. Figure 14 represents the classification accuracy of different features taken individually in order of their contribution to overall classification accuracy.

The classification accuracy obtained for individual features can be observed in Table 2.

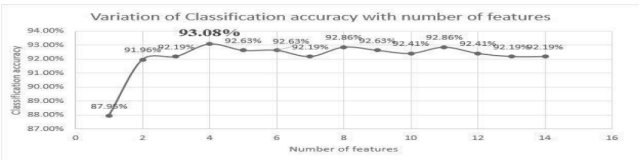

Figure 13. Variation of classification accuracy with number of features.

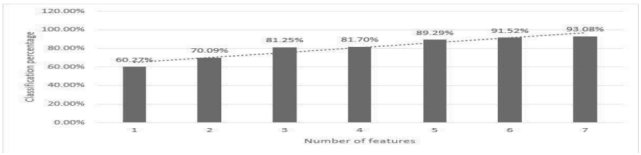

Figure 14. Classification accuracy of important features.

Table 2. Individual classification accuracy of features

V3	1	60.27%
V10+V3	2	70.09%
V1+V10+V3	3	81.25%
V6+V1+V10+V3	4	81.70%
V5+V6+V1+V10+V3	5	89.29%
V4+V5+V6+V1+V10+V3	6	91.52%
V2+V4+V5+V6+V1+V10+V3	7	93.08%

4.3 Feature Classification using Random Forest Algorithm

The selected features were classified using the random forest algorithm. In this case, eight features were used for classification. The obtained results were presented in the form of confusion matrix as shown in Table 3. The interpretation of the confusion matrix is as follows:

- The first element of the first row tells the number of vibration signals with no fault i.e. 0%.
- Fault.
- The second element of the first row tells the number of signals with 10% fault introduced.
- The sixth element of the first row tells the number of signals with 80% fault.

Zero value in first element of second, third, fifth, sixth and seventh row corresponds to misclassification of faulty conditions. Misclassifications in other conditions are given by elements other than diagonal elements and they constitute of only 6.69% which is acceptable for any fault diagnostic study for practical application.

P.F (Percentage Fault)

The details of the study are:

Correctly Classified Instances	418	93.3036%
Incorrectly Classified Instances	30	6.6964%
Kappa statistic	0.9219	
Mean absolute error	0.049	
Root mean squared error	0.1292	
Relative absolute error	20.0106%	
Root relative squared error	36.9094%	
Total Number of Instances	448	

Table 3. Confusion matrix using random forest algorithm

	Good	10 P.F	20 P.F	30 P.F	40 P.F	80 P.F	100 P.F
Good	61	0	1	2	0	0	0
10 P.F	0	59	4	0	0	1	0
20 P.F	0	8	56	0	0	0	0
30 P.F	2	0	1	57	3	0	1
40 P.F	0	0	0	3	59	0	2
80 P.F	0	1	0	0	0	63	0
100 P.F	0	0	0	0	1	0	63

5. Conclusion

Gearbox is an essential part of automobile which is subjected to faults. This paper presented the use of vibration signals for fault diagnostics by using wavelet features and random forest algorithm. From the acquired vibration signals, data features were extracted using wavelet features. Decision Tree was made to select the most important factors majorly affecting the classification accuracy. Random forest algorithm was then used to calculate the maximum classification accuracy of 93.0804% using the 7 features (V2+V4+V5+V6+V1+V10+V3). Hence, the results of random forest algorithm can be successfully used for fault diagnostics of gearbox with high accuracy.

6. References

1. Sugumaran V, Muralidharan V, Ramachandran KI. Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of

roller bearing. Mechanical Systems and Signal Processing. 2007 Feb; 21(2):930–42.
2. Mba D, Rao Raj BKN. Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines; bearings, pumps, gearboxes, engines and rotating structures. The Shock and Vibration Digest. 2006 Mar; 38(1):3–16.
3. Konga F, Chen R. A combined method for triplex pump fault diagnosis based on wavelet transform, fuzzy logic and neuronetworks. Mechanical Systems and Signal Processing. 2004 Jan; 18(1):161–8.
4. Sakthivel NR, Indira V, Nair BB, Sugumaran V. Use of histogram features for Decision Tree based fault diagnosis of mono-block centrifugal pump. International Journal of Granular Computing. Rough Sets and Intelligent Systems. 2011; 2(1):23–36.
5. 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.
6. Suykens JAK, Van Gestel T, Vandewalle J, De Moor B. A Support Vector Machine formulation to PCA analysis and its kernel version. IEEE Transactions Neural Network. 2003; 14(2):447–50.
7. Sakthivel NR, Sugumaran V, Babudevasenapati S. Vibration based fault diagnosis of monoblock centrifugal pump using Decision Tree. Expert Systems with Applications. 2010 Jun; 37(6):4040–9.
8. Sakthivel NR, Sugumaran V, Nair BB. Automatic rule learning using rough-set for fuzzy classifier in fault categorization of centrifugal pump. International Journal of Applied soft computing. 2012 Jan; 12(1):196–203.
9. Samanta B, Al-balushi KR, Al-araim SA. Artificial Neural Networks and Support Vector Machines with Genetic Algorithm for bearing fault detection. Engineering Applications of Artificial Intelligence. 2003 Oct–Dec; 16(7-8):657–65.
10. Bajric R, Zuber N, Skrimpas GA, Mijatovic N. Feature extraction using Discrete Wavelet Transform for gear fault diagnosis of wind turbine gearbox. Shock and Vibration. 2015 Sep; 2016:1–10.
11. Natarajan S. Gearbox fault diagnosis using Hilbert transform and study on classification of features by Support Vector Machine. International Journal of Hybrid Information Technology.2014; 7(4):69–82.
12. Gao L, Ren Z, Tang W, Wang H, Chen P. Intelligent gearbox diagnosis methods based on SVM, Wavelet Lifting and RBR. Sensors (Basel). 2010 May; 10(5):4602–21.
13. Praveenkumar T, Saimurugan M, Krishnakumar P, Ramachandran KI. Fault diagnosis of automobile gearbox based on machine learning techniques. Procedia Engineering. 2014 Dec; 97:2092–8.
14. Najafi M, Auslander DM, Bartlett PL, Haves P. Application of machine learning in fault diagnostics of mechanical systems. Proceedings of the World Congress on Engineering and Computer Science; 2008 Oct. p. 957–62.
15. Kankar PK, Sharma SC, Harsha SP. Fault diagnosis of ball bearings using machine learning methods. Expert Systems with Applications. 2011 Mar; 38(3):1876–86.
16. Devendiran S, Manivannan K, Kamani SC, Refai R. An early bearing fault diagnosis using effective feature selection methods and data mining techniques. International Journal of Engineering and Technology. 2015 Apr-May; 7(2):583–98.
17. Muralidharan V, Sugumaran V. Selection of discrete wavelets for fault diagnosis of Mono-block centrifugal pumps using the J48 algorithm. Applied Artificial Intelligence: An International Journal. 2013 Jan; 27(1):1–19.
18. Nassser M, Mohammadi M. Condition monitoring using wavelet transform and fuzzy logic by vibration signals. Life Science Journal.2012 Dec; 9(4):5680–5.
19. Jain D, Sugumaran V, Amarnath M, Kumar H. Fault diagnosis of helical gearbox using Decision Tree through vibration signals. International Journal of Performability Engineering. 2013 Mar; 9(2):221–34.
20. Staszewski WJ, Worden K, Tomlinson GR. Time-frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition. Mechanical Systems and Signal Processing. 1997 Sep; 11(5):673–92.
21. Yesilyurt I, Gu F, Ball AD. Gear tooth stiffness measurement using modal analysis and its use in wear fault severity assessment of spur gears. NDT&E International. 2003 Jul; 36(5):357–72.
22. Yesilyurt I. Gearbox fault detection and severity assessment using vibration analysis. EThos, e-these online Service; 1997.
23. Loutridis SJ. Damage detection in gear system using empirical mode decomposition. Engineering Structures. 2004 Oct; 26(12):1833–41.
24. Breiman L. Random Forests. Journal of Machine Learning. 2001 Oct; 45(1):5–32.
25. Sasikala S, Bharathidason S, Jothi Venkateswaran C. Improving classification accuracy based on random forest model through weighted sampling for noisy data with linear decision boundary. Indian Journal of Science and Technology. 2015 Apr; 8(8):614–9.
26. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: A classification and regression tool for compound classification and QSAR Modeling. Journal Chemical Information and Computer Science.2003 Nov; 43(6):1947–58.