Perspectives in Science (2016) 8, 334—337

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc

# Feature selection and classification of mechanical fault of an induction motor using random forest classifier☆

**Raj Kumar Patel***, **V.K. Giri**

*EED, MMMUT, Gorakhpur 273010, U.P., India*

Received 4 February 2016; accepted 10 April 2016
Available online 27 April 2016

**KEYWORDS**
CART;
Neural Network;
Random Forest;
Sensitivity;
Specificity

**Summary** Fault detection and diagnosis is the most important technology in condition-based maintenance (CBM) system for rotating machinery. This paper experimentally explores the development of a random forest (RF) classifier, a recently emerged machine learning technique, for multi-class mechanical fault diagnosis in bearing of an induction motor. Firstly, the vibration signals are collected from the bearing using accelerometer sensor. Parameters from the vibration signal are extracted in the form of statistical features and used as input feature for the classification problem. These features are classified through RF classifiers for four class problems. The prime objective of this paper is to evaluate effectiveness of random forest classifier on bearing fault diagnosis. The obtained results compared with the existing artificial intelligence techniques, neural network. The analysis of results shows the better performance and higher accuracy than the well existing techniques.


## Introduction

Now a day, industries are more dependent on the rotating machinery. The most popular machine used in the industries is induction motor staring from fractional horse power to large rating. In order to make continuous production, the monitoring of health of these induction motor are most challenging task for the health monitoring engineers (Mortazavizadeh and Mousavi, 2014). Generally, the parameter which is associated with the induction motors are recorded and used for analysis to know the internal condition of the motor. The vibration produced by the motor is one of the vital parameter which is widely used to detect the fault in induction motor. Mostly, the first action on recorded signal is to calculate parameters in time domain but this analysis do not claimed as effective method. Hence, the success of diagnosis depends on the other signal processing

☆ This article belongs to the special issue on Engineering and Material Sciences.

* Corresponding author.
*E-mail addresses:* rajkp007@gmail.com (R.K. Patel), girivkmmm@gmail.com (V.K. Giri).

http://dx.doi.org/10.1016/j.pisc.2016.04.068

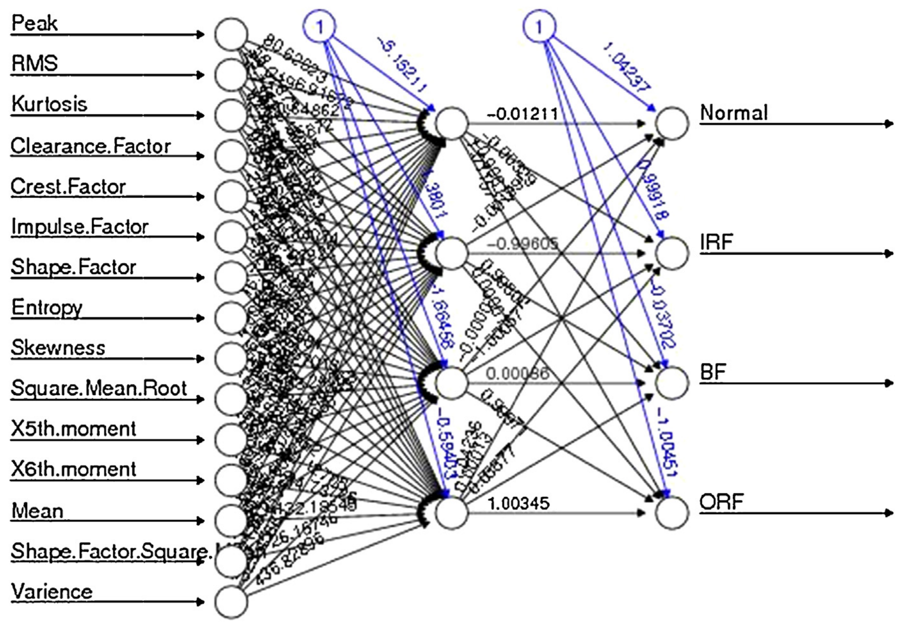


**Figure 1** Trained neural network with input, output and training process.

techniques. During last decade, frequency spectrum, envelope detection using Hilbert transform, Empirical Mode Decomposition (EMD), Short Time Fourier Transform (STFT), Wavelet Transform (WT) methods have been used for extracting the features and to identify the fault frequencies (Aiyu et al., 2013; Li et al., 2012; Wang et al., 2013; Xu et al., 2013). Principal Component Analysis (PCA), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) are very popular for selecting the significant feature for the classification of fault type and its severity (Hu et al., 2014; Unal et al., 2014). Automatic fault detection and online condition monitoring Artificial Neural Network (ANN) and Support Vector Machine (SVM) are enormously used in the literature. The existing classifier used for fault detection is affected when the number of class is more. It has also been reported that the response become sluggish. In order to sort out the mentioned problem in the present paper Random Forest (RF) classifier method has been used. Its performance has been compared with the one of the conventional classifier, i.e. ANN.

## Data description

In the present study, bearing vibration records have been selected from the Case Western Reserve Lab data centre for analysis and classification purpose. The data have been recorded from 2 HP three phase induction motor bearing using accelerometer. The accelerometer was mounted on the motor housing at the drive end of the motor. The samples have been taken at 12,000 per second for each of a 16 channels digital audio tape (DAT) recorder. The speed of the shaft is measured as 1797—1752 rpm from no load to full load, respectively. The fault seeded into the races and rolling ball. The size of fault is 0.007 inch, 0.014 inch and 0.021 inch for each cases of fault. The recorded data have been prepared for the classification pupose. The four cases have been taken as Normal, Inner Raceway Fault (IRF),

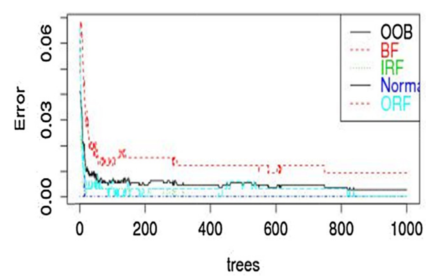


**Figure 2** OOB-error of random forest classifier.

Bearing Ball Fault (BF) and Outer Raceway Fault (ORF). The data for classification of Normal case is $160 \times 3000$ samples and for IRF, BF, ORF the size is chosen of $480 \times 3000$ for each case. From the prepared data, the 15 statistical feaures have been calculated, and feeded to the neural network and random forest classifier, these features are shown in Fig. 1. The output of the classifier chosen as four related to the condition of the bearing of an induction motor.

## Artificial neural network model

Artificial neural networks model is inspired from biological learning process of the human brain and have been developed in form of parallel distributed network. During the training process of the neural networks are fitted to the data by learning algorithms. In this article, supervised learning has been used with input of 15 feature and four output. In the present work, the traditional backpropagation algorithm has been used which is regarded by the usage of a given output that compared to the predicted output and by adjusted of all parameters according to comparision. The parameter of ANNs such as weigts are usully initialized with random values drawn from a standard normal distribution. The first step of the ANN training process is that to compute an output for given inputs and its current weights and output is compared

**Table 1** Confusion matrix of classifiers output.

| Prediction | Neural network | | | | Random forest | | | |
|---|---|---|---|---|---|---|---|---|
| | Normal | IRF | BF | ORF | Normal | IRF | BF | ORF |
| Normal | 160 | 0 | 01 | 0 | 160 | 0 | 0 | 0 |
| IRF | 0 | 479 | 08 | 01 | 0 | 480 | 0 | 0 |
| BF | 0 | 01 | 468 | 0 | 0 | 02 | 478 | 0 |
| ORF | 0 | 0 | 03 | 479 | 0 | 0 | 0 | 480 |

**Table 2** Accuracy measures of classifiers.

| Prediction | Neural network | | | | Random forest | | | |
|---|---|---|---|---|---|---|---|---|
| | Normal | IRF | BF | ORF | Normal | IRF | BF | ORF |
| Senstivity | 1.0 | 0.9917 | 0.9915 | 0.9917 | 1.0 | 0.9931 | 1.0 | 1.0 |
| Specificity | 0.9993 | 0.991 | 0.9911 | 0.991 | 1.0 | 1.0 | 0.9970 | 1.0 |
| Positive predictive value | 0.9938 | 0.9796 | 0.9791 | 0.9796 | 1.0 | 1.0 | 0.9931 | 1.0 |
| Negative predictive value | 1.0 | 0.9964 | 0.9964 | 0.9964 | 1.0 | 1.0 | 1.0 | 1.0 |
| Prevalance | 0.1 | 0.3019 | 0.295 | 0.3019 | 0.1 | 0.1 | 0.3 | 0.3 |
| Balance accuracy | 0.9994 | 0.9913 | 0.9913 | 0.9913 | 1.0 | 0.9965 | 0.9985 | 1.0 |

with predicted output. The second step is to calculate the error and as per algorithm weigts are adjusted and checked error below the threshold value. However, the training time is relatively long and it is also susceptible to local minimum traps (Guenther and Fritsch, 2012; Prieto et al., 2013). Fig. 1. shows the trained neural network with their weights, input and output structure.

## Random forest model

The random forest (RF) model is based on the grouping of trees for classification and regression (CART) and introduced by Breiman (Guenther and Fritsch, 2012). The method is based on the tree-type classifier. Each tree classifier is named a component predictor. A large number tree makes RF from sub-dataset. The distinctive training set is completed by using bagging. The procedure bagging means random sampling with replacement which is a meta-algorithm to improve classification, and regression models according to stability and classification accuracy. The process bagging reduces variance and aids to elude over-fitting synchronously. The final decision makes by totalling the votes of component predictors on each class and then choosing the winner class in terms of the number of votes to it. The

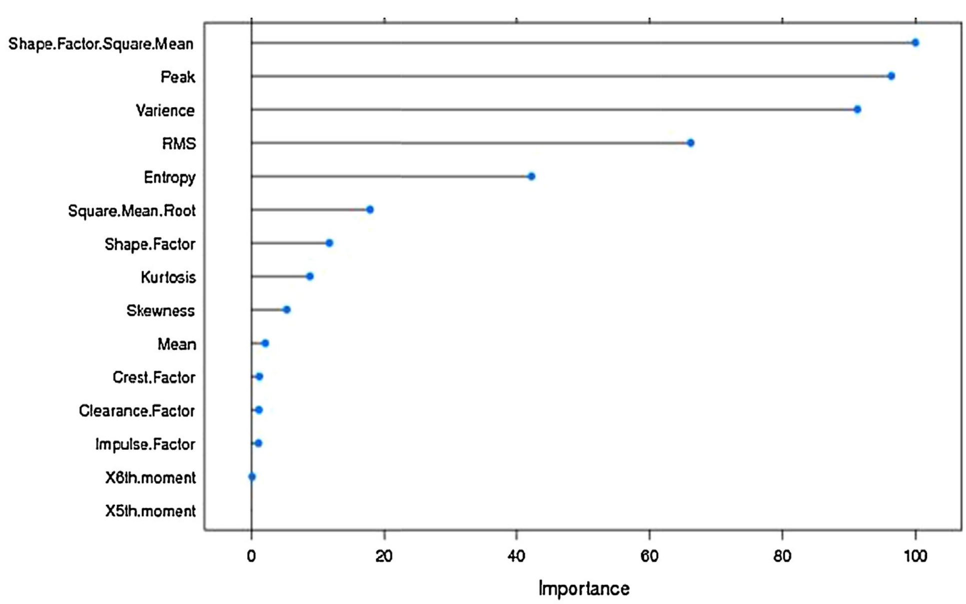


**Figure 3** Features importance.

RF performance measure by a metric called out of bag error (oob-error) calculated as the average of the rate of error in each weak learner (Cabrera et al., 2015; Yang et al., 2008).

## Results and discussions

In the present work, feature selection and classification of bearing fault of an induction motor using random forest and neural network have been presented. Table 1. shows confusion matrix of both the classifier which indicates us how well the classifer classified Normal, IRF,BF, and ORF labels with the actual labels. From the table, we observed that the false prediction of RF is only 2 out of 1600 data set and for ANNs it is 14. Other Accuracy measures derived from a confusion matrix have been shown in Table 2.

The RF classifier has been found more suitalbe for the fault diagnosis of any rotating machine. Fig. 2 shows the aggregate OOB-error visualize the error rates as the number of trees increase and it is find decreasing in nature and appraches to 0.01. Further, we may observe which variables were most important in classifying the correct labels. From Fig. 3, we can see the most important features are shape factor square mean, peak, varience, rms and entrpy. The least important features are impulse factor, x6th moment and x5th moment. Another step has been performed by using the top four most important variables and find 100% prediction.

## Conclusions

This paper has presented random forest classifier an approach for classification of bearing fault. The statistical features from bearing vibration signal have been calculated and feed to the RF and ANNs classifier. From obtained results, it has been observed that RF performance is better than the ANNs. From the results, it is also concluded that RF runs efficiently on large data bases. The RF found to be unexcelled in accuracy among current algorithms. It may estimate the importance of each variable in the classification. The results obtained using RF method is promising therefore; proposed method may be used for bearing fault detection and diagnosis.

## Acknowledgements

This work is carried out at M.M.M. University of Technology, Gorakhpur and supported by TEQIP- II, at a program of the government of India. The authors would like to thank the Case Western Reserve University for providing free access to the bearing vibration experimental data from their website.

## References

Aiyu, W., Hongxia, P., Huiling, L., 2013. Fault diagnosis of automaton based on EMD and close degree. Sensors Transducers 160, 612—618.

Cabrera, D., Sancho, F., Sánchez, R.-V., Zurita, G., Cerrada, M., Li, C., Vásquez, R.E., 2015. Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition. Front. Mech. Eng. 10, 277—286.

Guenther, F., Fritsch, S., 2012. neuralnet: training of neural networks. R J. 2, 30—38.

Hu, Z., Chen, Z., Gui, W., Jiang, B., 2014. Adaptive PCA based fault diagnosis scheme in imperial smelting process. ISA Trans. 53, 1446—1455, http://dx.doi.org/10.1016/j.isatra.2013.12.018.

Li, N., Zhou, R., Hu, Q., Liu, X., 2012. Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. Mech. Syst. Signal Process. 28, 608—621.

Mortazavizadeh, S.A., Mousavi, S.M.G., 2014. A review on condition monitoring and diagnostic techniques of rotating electrical machines. Phys. Sci. Int. J. 4 (3), 310—338.

Prieto, M.D., Cirrincione, G., Espinosa, A.G., Ortega, J.A., Henao, H., 2013. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. IEEE Trans. Ind. Electr. 60, 3398—3407.

Unal, M., Onat, M., Demetgul, M., Kucuk, H., 2014. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. Measurement.

Wang, D., Tse, P.W., Tsui, K.L., 2013. An enhanced Kurtogram method for fault diagnosis of rolling element bearings. Mech. Syst. Signal Process. 35, 176—199.

Xu, B., Sun, L., Xu, L., Xu, G., 2013. Improvement of the Hilbert method via esprit for detecting rotor fault in induction motors at low slip. IEEE Trans. Energy Convers. 28, 225—233.

Yang, B.-S., Di, X., Han, T., 2008. Random forests classifier for machine fault diagnosis. J. Mech. Sci. Technol. 22, 1716—1725.