ISSN: 0975-766X
CODEN: IJPTFI
Available Online through **Research Article**
www.ijptonline.com

FAULT DIAGNOSIS FOR WIND TURBINE BLADE THROUGH VIBRATION SIGNALS USING STATISTICAL FEATURES AND RANDOM FOREST ALGORITHM

A. Joshuva, V. Sugumaran*
School of Mechanical and Building Sciences (SMBS), VIT University, Chennai Campus,
Vandalur-Kelambakkam Road, Chennai 600127, India.
E-mail: v_sugu@yahoo.com

Received on: 14-02-2017 *Accepted on: 22-03-2017*

Abstract

Wind energy is one of the important renewable energy resources. The wind energy is converted into electrical energy using rotating blades which are connected to the generator. Due to environmental conditions and large structure, the blades are subjected to various faults and cause a lack of productivity. The downtime can be reduced when they are diagnosed periodically using structural health monitoring. These are considered as a pattern recognition problem which consists of three phases namely, feature extraction, feature selection and feature classification. In this study, statistical features were extracted from vibration signals, feature selection was carried out using a J48 decision tree algorithm and feature classification was performed using random forest algorithm.

Keywords: Fault diagnosis, Structural health monitoring, Wind turbine blade, Statistical feature, J48 algorithm, Random forest algorithm.

1. Introduction

Wind energy is one of the desirable renewable sources. The main focus of the wind turbine is to boost energy mining. To increase performance, accessibility, reliability, tolerability, the life of turbines must be enhanced [1]. Because of different catastrophic incidents and complex wind pattern, the system is exposed to vibration. Wind turbine blades gets affected when compared with other parts of wind turbine because of its extensive structure.The continuous and severe vibration will create damage to blade and leads to breakdown. Due to huge structure and operating condition, the vibration of the blades is difficult to evaluate on-line. Early acknowledgment of the faults can keep the system away from breakdown and the productivity can also be increased. Hence, identification of different fault conditions using a machine learning approach is essential [2]. The challenge in building a classifier model is to find the best 'feature-classifier' pair that will identify the condition of the blade.

Many studies were carried out using machine learning, to name a few, Andrew Kusiak and AnoopVerma [3] built a data-driven model for monitoring blade pitch faults in wind turbines. Two blade pitch faults were considered namely, blade angle asymmetry and blade angle implausibility and determine the associations between them. The study was carried out using bagging (72.5%), artificial neural network (76.2%), pruning rule-based classification tree (75.5%), K-nearest neighbour (73.5%) and genetic programming (74.7%) algorithms.Godwin and Matthews [4] have carried out a work on classification and detection of wind turbine pitch faults through SCADA data analysis and RIPPER algorithm which yield them 87.05% classification accuracy in pitch angle fault.Chen *et al.*, [5] built a model for wind turbine pitch fault prognosis using apriori knowledge based adaptive neuro-fuzzy inference system (ANFIS) and SCADA data. The classification accuracy which obtained using ANFIS is about 88.30% for the diagnosing blade pitch fault.

A comparative study on wind turbine power coefficient estimation by soft computing methodologies was carried out by Shamshirband *et al.*,[6]. In this study, support vector regression (radial basis function), support vector regression (polynomial), ANFIS (adaptive neuro-fuzzy inference system), NN (neural network) algorithms were used and the algorithms performance was compared. Correlation Coefficient of algorithms where found to be SVR (RBF)-0.997, SVR (Polynomial)-0.504, ANFIS-0.978 and NN-0.922.A study on structural health monitoring of wind turbine blades using acoustic source localization and wireless sensor networks was carried out by Bouzid*et al.*, [7]. In their study, they obtained an error rate of 7.98% for the condition monitoring of the wind turbine blade.

A study on wavelet transform based stress and time history editing of horizontal axis wind turbine blades was carried out by Pratumnopharat *et al.*, [8]. With wavelet transform, this method extracts fatigue damage parts from the stress-time history and generates the edited stress-time history with the shorter time length. In this study, Time correlated fatigue damage (89.82%), Mexican hat wavelet (79.23%), Meyer wavelet (79.76%), Daubechies 30th order (80.81%), Morlet wavelet (80.34%) and Discrete Meyer wavelet (80.30%) was used for the classification of crack on the blade. Hoell and Omenzetter [9] carried out a study on structural damage detection in wind turbine blades based on time series representations of dynamic responses using vibration data. This study was carried out using algorithms like cross-correlations, principal component analysis (PCA), genetic programming (GP) and the stuructural damage was analyzed.

Althoughsome researchers carried out fault diagnosis system using machine learning approach and they considered very few faults in their study. Only selected algorithms were used to perform the fault prediction on wind turbine blades and many algorithms are yet to be studied [10]. Figure 1 shows the methodology. The main contribution in this study is

1. This study considers five faults (blade crack, erosion, hub-blade loose connection, pitch angle twist and blade bend) for wind turbine blade fault diagnosis.
2. Statistical analysis was used for feature extraction.
3. J48 algorithm was used for feature selection.
4. Random forest algorithm was used for feature classification.

The rest of the paper is organized as follows. In section 2, experimental setup and experimental procedure are explained. Section 3 presents the feature extraction and feature selection procedure. The random forest classifier details are explained in section 4. The classification accuracy of the models is discussed and the suggestion of the better model is proposed in section 5. Conclusions are presented in the final section (section 6).

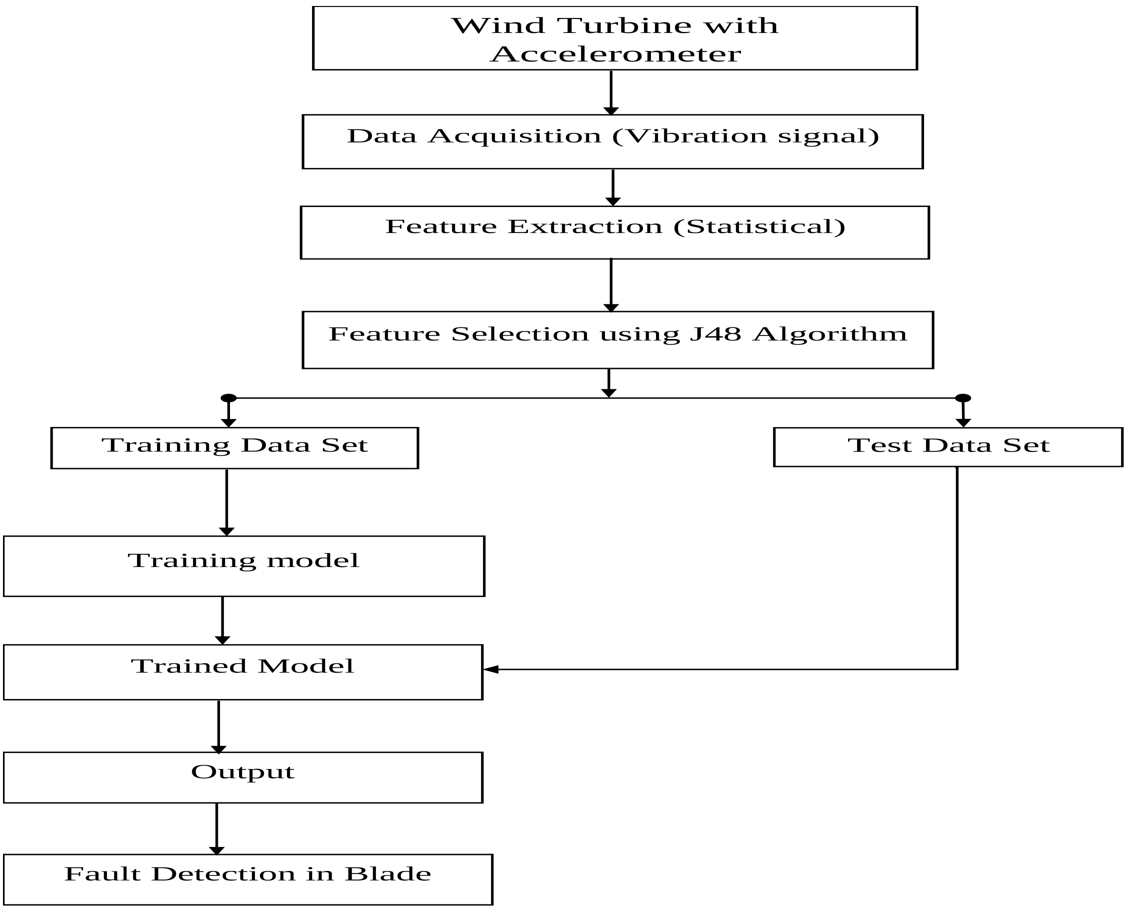


Figure 1: Methodology.

2. Experimental Studies

The main aim of this study is to classify whether the blades are in good condition or in a defective state. If it is defective, then the objective is to identify the condition of fault. The experimental setup and experimental procedure are described in the following subsections.

2.1. Experimental Setup

The experiment was carried out on a 50W, 12V variable speed wind turbine (MX-POWER, model: FP-50W-12V). The technical parameters of a wind turbine are given in Table 1. The wind turbine was mounted on a fixed steel stand. The wind speed ranges from 5m/s to 15 m/s where the turbine is placed for conducting the experiment with an open circuit wind tunnel being used as the wind source in this study. The piezoelectric accelerometer (DYTRAN 3055B1) was mounted on the nacelle near to the wind turbine hub to record the vibration signals using the adhesive mounting technique. The accelerometer was directly connected to the data acquisition unit (NI USB-4432), where the analogue signals were converted into digital signals and it was stored in computer memory. These signals were used to study the classification of the faults. Figure 2 shows the wind turbine setup.

Table 1. Technical parameters of wind turbine.

Model	FP-50W-12V
Rated Power	50W
Rated Voltage	12V
Rated Current	8A
Rated Rotating Rate	850 rpm
Max Power	150W
Start-up Wind Velocity	2.5 m/s
Cut-in Wind Velocity	3.5 m/s
Cut-out Wind Velocity	15 m/s
Security Wind Velocity	40 m/s
Rated Wind Velocity	12.5 m/s
Engine	Three-phase permanent magnet generator
Rotor Diameter	1050mm
Blade Material	Carbon fiber reinforced plastics

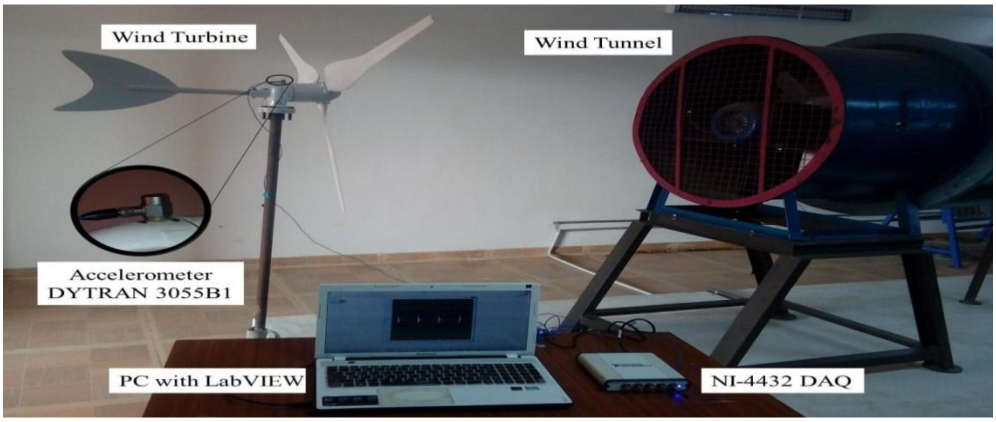


Figure 2. Wind turbine setup.

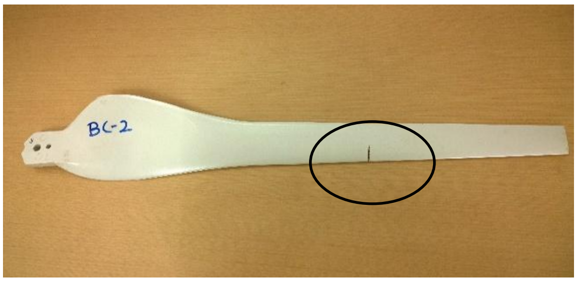
2.2. Experimental procedure

In the present study, three-blade variable horizontal axis wind turbine (HAWT) was used. Initially, the wind turbine was considered to be in good condition (free from defects, new setup) and the signals were recorded using an accelerometer. These signals were recorded with the following specifications:

1. Sample length: 10000 (Ten thousand data points).
2. Sampling Frequency: The sampling frequency should be at least twice the highest frequency contained in the signal as per Nyquist sampling theorem. By using this theorem sampling frequency was calculated as 12 kHz (12000Hz).
3. Number of samples: Minimum of 100 (hundred) samples were taken for each condition of the wind turbine blade and the vibration signals were stored in data files.

The following faults were simulated one at a time while all other components remain in good condition and the corresponding vibration signals were acquired. Figure 3 shows the different blade fault conditions which are simulated on the blade.

BG - Blade good condition

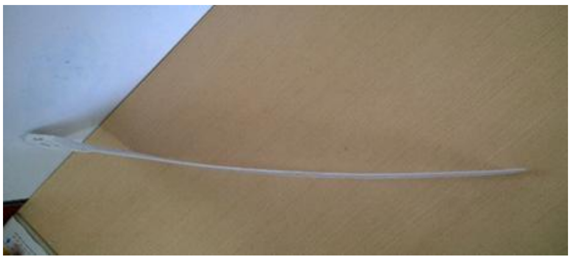


BC-2 – Blade crack

BPT – Blade pitch twist

BE – Blade erosion

BB – Blade bend (Front View)



Blade bend (Top View)

Figure 3. Various blade fault conditions.

a) Blade bend (BB): This fault occurs due to the high-speed wind and complex forces caused by the wind. The blade was made to flap wise bend with 10^0 angles.

b) Blade crack (BC-2): This occurs due to foreign object damage on blade while it is in operating condition. On blade, 15mm crack was made.

c) Blade erosion (BE): This fault is due to the erosion of the top layer of the blade by the high-speed wind. The smooth surface of the blade was eroded using emery sheet (320Cw) to provide an erosion effect on the blade.

d) Hub-blade loose contact: This fault generally occurs on a wind turbine blade due to an excessive runtime or usage time. The bolt connecting the hub and blade was made loose to obtain this fault.

e) Blade pitch angle twist (PAT): This fault occurs due to the stress on the blade caused by high-speed wind. This makes the pitch get twisted, creating a heavy vibration to the framework. To attain this fault, blade pitch was twisted about 12^0 with respect to the normal blade condition.

Figure 4 shows the time domain signals which were taken from different conditions of the wind turbine blade. They show the vibration signal plot (amplitude vs time of the vibration) for good condition blade, blade bend, blade erosion, hub-blade loose connection, blade crack and pitch angle twist respectively.

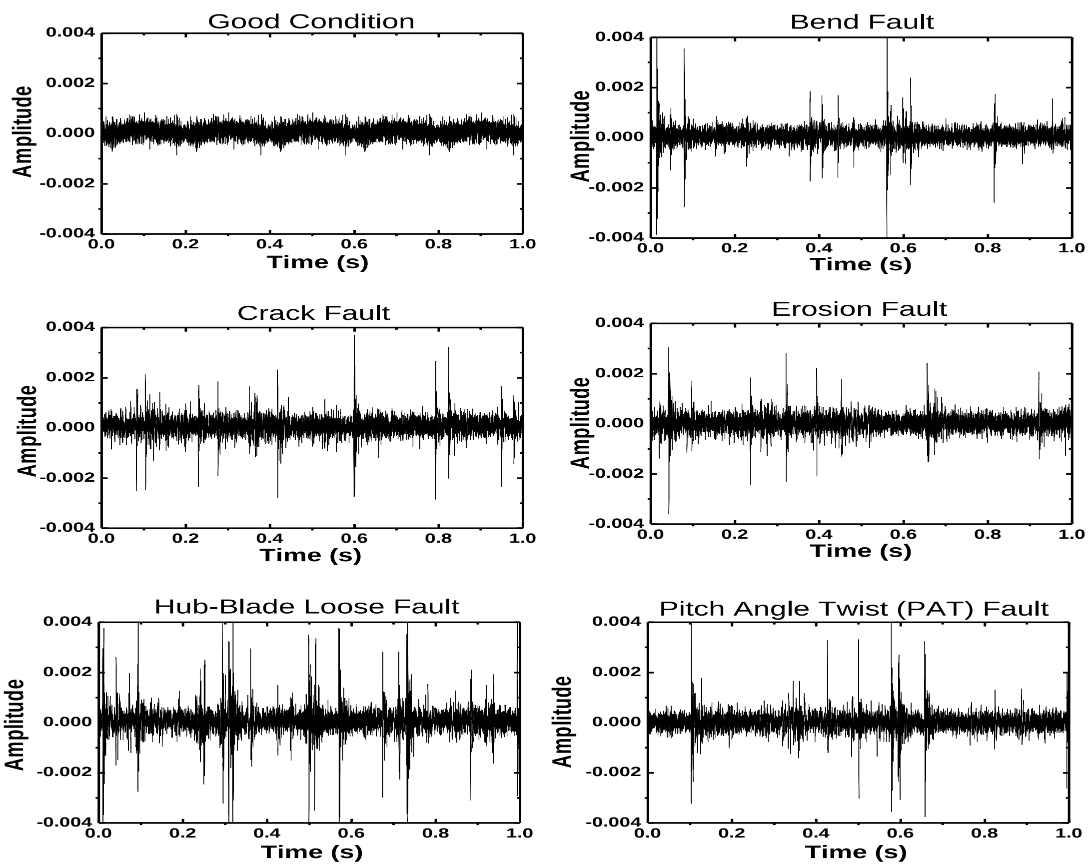


Figure 4. Time-domain signal plot.

3. Feature Extraction and Feature Selection Process

3.1. Feature Extraction

In this study, vibration data for various blade fault conditions were collected from data acquisition system (DAQ). Directly vibration signals cannot be used as input to the classifier. Hence, features need to be extracted using statistical methods. The process of computing some measures which will represent the signal is called feature extraction. Statistical information for vibration signals yields different parameters such as sum, mean, median, mode, minimum, maximum, range, skewness, kurtosis, standard error, standard deviation and sample variance [11]. Once the statistical feature extraction was completed, the features were taken and the feature selection method was implemented. The most contributing features are selected from the obtained statistical features using the J48 algorithm.

3.2. Feature Selection

J48 decision tree algorithm is adapted from the C4.5 algorithm in WEKA. It 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 the root to a leaf, and each node involves one attribute. The occurrence of an attribute in a tree provides information about the importance of the associated attribute.

A decision tree is a tree based knowledge representation methodology used to represent classification rules. J48 decision tree algorithm is a widely used one to construct decision trees [16].

The J48 decision tree algorithm has been applied to the problem for feature selection. The input to the algorithm is the set of statistical features described above and output of the decision tree shown in Figure 5. It is clearly shown that the top node is the best node for classification. The other features in the nodes of decision tree perform in descending order of significance.

It is to be mentioned here that only features that contribute to the classification appear in the decision tree and other features do not contribute much. The features which have the less discriminating capability can be consciously discarded by deciding on the threshold. This concept is made use for selecting good features. The algorithm identifies the good features for the purpose of classification of the given training data set, and thus reduces the domain knowledge required to select good features for pattern classification problem. J48 decision tree algorithm select the features using information gain and entropy reduction [12]. The most dominating features to represent the blade conditions sum, standard deviation, kurtosis, mode, skewness and range.

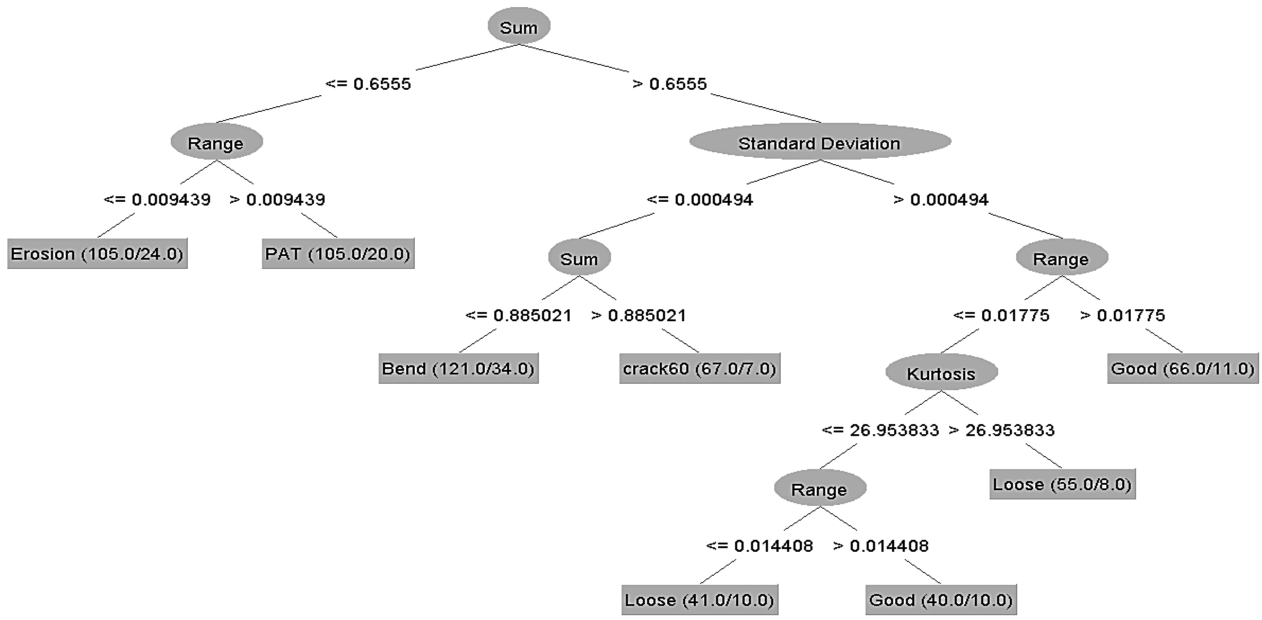


Figure 5: J48 Tree classification for feature selection.

4. Feature Classification – Random forest (RF) algorithm

Random forest (RF) [13] is an idea of the general system of random decision forests that are a group learning technique for characterization, regression and other different errands. It works by building a large number of decision trees at while execution and yields the class (classification) or mean prognosis (regression) of the individual trees [14]. Random decision forestsperfect thedecision trees propensity for overfitting to their training set.

A random forest is a classifier comprising of a group of tree structured classifiers $\{h(\mathrm{x},\Theta_k), k = 1, ...\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and every tree makes a unit choice for the most famous class at input x [15]. The random forest converges is given an ensemble of classifiers $h_1(\mathrm{x}),h_2(\mathrm{x}), ... ,h_k(\mathrm{x})$, and with the training set drawn at random from the distribution of the random vectorsY,X. The margin functionisdefined as

$$mg(X,Y)=av_kI(h_k(X)=Y)-max_{j\neq Y}av_kI(h_k(X)=j) \text{—(1)}$$

Where $I(\bullet)$is the indicator function. The margin measures the extent to whichthe average number of votes at $X,Y$for the right class exceeds the average votefor any other class. The larger the margin, the more confidence in theclassification. The generalization error is given by

$$PE^*=P_{X,Y}(mg(X,Y)<0) \qquad \text{—(2)}$$

Where, the subscripts X,Yindicate that the probability is over the X,Y space.

5. Results and discussion

The vibration signals were recorded for good condition blade and other conditions of wind turbine blade. 600 samples were gathered; out of which 100 samples were from the blade in good condition. 100 samples from every faulty condition were gathered. The statistical features were considered as features and serves as input to the algorithm. The classified data will be the required outcome of the algorithm.

From twelve features, only six features were selected as explained in section 3. The six features are sum, standard deviation, kurtosis, mode, skewness and range. Figure 6 shows the number of features vs classification accuracy. These six features were given as the input to the random forest algorithm to carry out the classification of faults on blade.

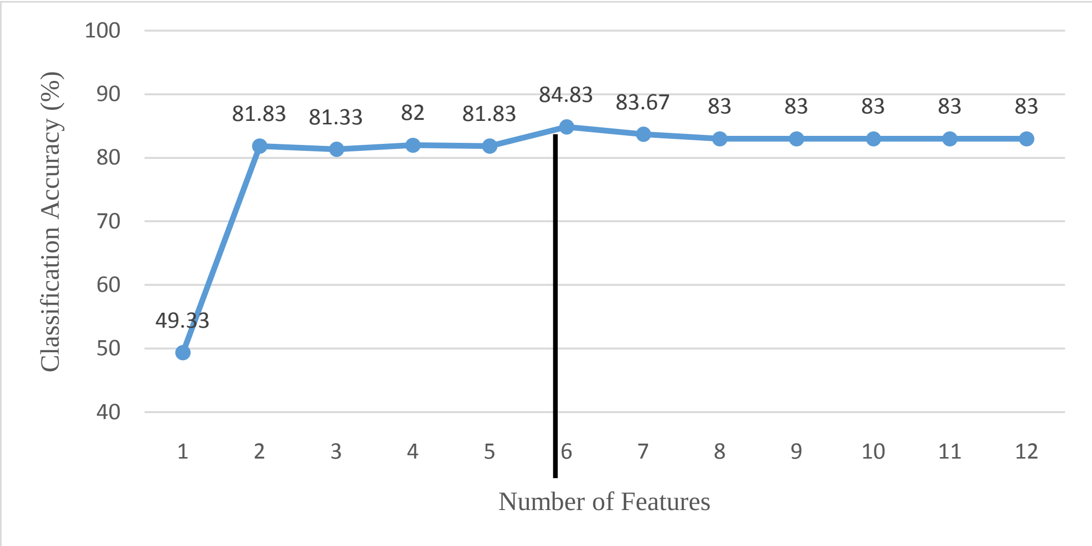


Figure 6: Number of features Vs Classification Accuracy.

After the feature selection, these features were given to the random forest algorithm as an input. The number of features are taken in random by the algorithm. In this study, the random selection (K) is set from one to six since, only six features were given as input to the classifier. In the classifier, the number of trees to be built is kept constant to all the random selection (I=30) and the maximum depth of the tree also kept constant (depth=10). Table 2 shows the random selection of features from one to six. From Figure 7, K=3 has the maximum classification accuracy of 89.50%

Table 2: Classification accuracy of random forest.

Number of attributes taken in the random selection (K)	Classification Accuracy (%)
1	87.00
2	87.33
3	**89.50**
4	87.67
5	89.17
6	88.17

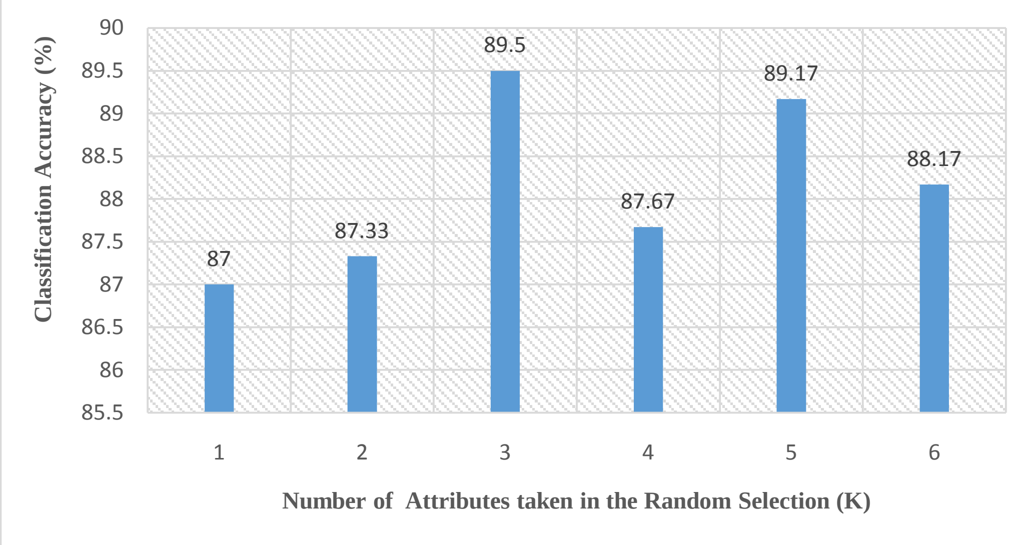


Figure 7: Classification Accuracy for number of random selection.

The classifier performance is verified by 10-fold cross validation [16]. In this confusion matrix, the diagonal element represents the correctly classified instance and the others are misclassified. Table 3 represents the confusion matrix of the random forest classifier. In confusion matrix, the diagonal elements represent the correctly classified instances and the others are misclassified instances. Also one can observe more misclassifications between good and loose conditions. For the loose condition, the bolts between the hub and the blade were made loose (please note that the blade was in good condition). However, at high wind speed, the blade can stick to the hub and behave like a good condition during operation. Because of this, the signature of the loose condition sometimes resembles good condition and the classifier finds difficult to distinguish between them; hence, more misclassifications.

Table 3: Confusion matrix of random forest algorithm.

Blade conditions	Good	Bend	Crack	Loose	Pitch twist	Erosion
Good	**84**	0	1	15	0	0
Bend	0	**91**	5	0	0	4
Crack	0	7	**89**	4	0	0
Loose	13	0	5	**82**	0	0
Pitch twist	0	0	0	0	**97**	3
Erosion	0	2	0	0	4	**94**

Table 4: Classwise accuracy of random forest algorithm.

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC area
Good	0.84	0.026	0.866	0.84	0.853	0.977
Bend	0.91	0.018	0.91	0.91	0.91	0.955
Crack	0.89	0.022	0.89	0.89	0.89	0.993

Loose	0.82	0.038	0.812	0.82	0.816	0.96
Pitch twist	0.97	0.008	0.96	0.97	0.965	0.998
Erosion	0.94	0.014	0.931	0.94	0.935	0.991

The kappa statistics are used to measure the arrangement of likelihood with the true class and was found to be 0.874. The mean absolute error is a measure of how close forecasts or predictions are to the ultimate result and was found to be 0.0589. The root mean square error is a quadratic scoring rule which processes the average size of the error. The value was found to be 0.1654.

The detailed classwise accuracy is shown in Table 4. The classwise accuracy is expressed in terms of the true positive rate (TP), false positive rate (FP), precision, recall, F-Measure and receiver operating characteristics (ROC) area. TP and FP are very much important in the classification accuracy. The true positive (TP) rate should be close to ‘1’ and the false positive (FP) rate should be close to ‘0’ to propose the classifier is a better classifier for the classification. In the random forest algorithm, it has the TP near to ‘1’ and FP close to ‘0’ then it can be predicted that the classifier which build for the particular problem is very much effective for the fault diagnosis problem.

6. Conclusion

The wind turbine is important in the use of wind energy. This paper described an algorithmic based classification of vibration signals for the evaluation of the wind turbine blade condition using machine learning approach.From the acquired vibration data, a model was set up using data modelling technique. A random forest algorithm was used to learn and classify the different conditions of the blade. The model was tested under 10-fold cross validation and correctly classified instances was found to be 89.50%. The error rate is relatively less and may be considered for the blade fault diagnosis. Hence, the random forest algorithm can be practically used for the condition monitoring of wind turbine blade to reduce the downtime and to use more wind energy.

Reference:

1. Yogesh Kumar, Jordan Ringenberg, Soma ShekaraDepuru, Vijay K. Devabhaktuni, Jin Woo Lee, EfstratiosNikolaidis, Brett Andersen, AbdollahAfjeh, Wind Energy: Trends And Enabling Technologies, Renewable And Sustainable Energy Reviews, Volume 53, January 2016, Pages 209-224.
2. AbolfazlPourrajabian, Peyman Amir NazmiAfshar, Mehdi Ahmadizadeh, David Wood, Aero-Structural Design And Optimization Of A Small Wind Turbine Blade, Renewable Energy, Volume 87, Part 2, March 2016, Pages 837-848.

3. Kusiak A, Verma A. A data-driven approach for monitoring blade pitch faults in wind turbines. Sustainable Energy, IEEE Transactions on. 2011 Jan;2(1):87-96.
4. Jamie L. Godwin, Peter Matthews, Classification and Detection of Wind Turbine Pitch Faults Through SCADA Data Analysis, International Journal of Prognostics and Health Management, ISSN 2153-2648, 2013 016.
5. Chen B, Matthews PC, Tavner PJ. Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS. Expert Systems with Applications. 2013 Dec 1;40(17):6863-76.
6. ShahaboddinShamshirband, DaliborPetković, HadiSaboohi, Nor BadrulAnuar, IrumInayat, ShatirahAkib, ŽarkoĆojbašić, VlastimirNikolić, Miss Laiha Mat Kiah, Abdullah Gani, Wind turbine power coefficient estimation by soft computing methodologies: Comparative study, Energy Conversion and Management, Volume 81, May 2014, Pages 520-526.
7. Omar MabrokBouzid, Gui Yun Tian, KanapathippillaiCumanan, Davidmoore, Structural Health Monitoring Of Wind Turbine Blades: Acoustic Source Localization Using Wireless Sensor Networks, Hindawi Publishing Corporation, Journal Of Sensors Article ID 139695, 2014.
8. Pratumnopharat P, Leung PS, Court RS. Wavelet transform-based stress-time history editing of horizontal axis wind turbine blades. Renewable Energy. 2014 Mar 31;63:558-75.
9. Simon Hoell, Piotr Omenzetter, Structural damage detection in wind turbine blades based on time series representations of dynamic responses, Smart Materials and Nondestructive Evaluation for Energy Systems 2015, Conference Paper - March 2015, Volume 9439, pp. 1-11.
10. Joshuva, A., Sugumaran, V. Fault diagnostic methods for wind turbine: A review (2016) ARPN Journal of Engineering and Applied Sciences, 11 (7), pp. 4654-4668.
11. M. Amarnath, V. Sugumaran, Hemantha Kumar, Exploiting Sound Signals For Fault Diagnosis Of Bearings Using Decision Tree, Measurement, Volume 46, Issue 3, April 2013, Pages 1250-1256.
12. V. Sugumaran, V. Muralidharan, K.I. Ramachandran, Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, Mechanical Systems and Signal Processing, Volume 21, Issue 2, February 2007, Pages 930-942.
13. Livingston. F, Implementation of Breiman’s Random Forest Machine Learning Algorithm, ECE591Q Machine Learning Journal Paper, 2005.

14. Kellie J. Archer, Ryan V. Kimes, Empirical Characterization Of Random Forest Variable Importance Measures, Computational Statistics & Data Analysis, Volume 52, Issue 4, 10 January 2008, Pages 2249-2260.
15. Svetnik. V, Liaw. A, Tong. C, Culberson. J.C, Sheridan. R.P, Feuston. B.P, Random Forest: A Classification And Regression Tool For Compound Classification And QSAR Modelling, Journal Of Chemical Information And Computer Sciences, 2003 Nov Volume 43, Issue 6, November 2003, Pages 1947-1958.
16. V. Muralidharan, V. Sugumaran, A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis, Applied Soft Computing, Volume 12, Issue 8, August 2012, Pages 2023-2029.