International Journal of Engineering

Journal Homepage: www.ije.ir

# Fault Detection of Anti-friction Bearing using Ensemble Machine Learning Methods

S. Patil*[a], V. Phalle[b]

[a] *Centre of Excellence in Complex and Nonlinear Dynamical Systems (CoE-CNDS), Veermata Jijabai Technological Institute, Mumbai, India*
[b] *Mechanical Engineering Department, Veermata Jijabai Technological Institute, Mumbai, India*

*P A P E R  I N F O*

*Paper history:*
Received 30 June 2018
Received in revised form 26 August 2018
Accepted 26 October 2018

*Keywords*:
*Anti-friction Bearing,*
*Ensemble Learning,*
*Vibration Signal*
*Fault Detection*

*A B S T R A C T*

Anti-Friction Bearing (AFB) is a very important machine component and its unscheduled failure leads to cause of malfunction in wide range of rotating machinery which results in unexpected downtime and economic loss. In this paper, ensemble machine learning techniques are demonstrated for the detection of different AFB faults. Initially, statistical features were extracted from temporal vibration signals and are collected using experimental test rig for different input parameters like load, speed and bearing conditions. These features are ranked using two techniques, namely Decision Tree (DT) and Randomized Lasso (R Lasso), which are further used to form training and testing input feature sets to machine learning techniques. It uses three ensemble machine learning techniques for AFB fault classification namely Random Forest (RF), Gradient Boosting Classifier (GBC) and Extra Tree Classifier (ETC). The impact of number of ranked features and estimators have been studied for ensemble techniques. The result showed that the classification efficiency is significantly influenced by the number of features but the effect of number of estimators is minor. The demonstrated ensemble techniques give more accuracy in classification as compared to tuned SVM with same experimental input data. The highest AFB fault classification accuracy 98.12% is obtained with ETC and DT feature ranking.

***doi***: *10.5829/ije.2018.31.11b.22*

## 1. INTRODUCTION

AFB is common element in most of the rotating machinery. It is used to support a load and enables relative motion between machine components. Unexpected failure due to faults developed in such important components causes a machinery to shut down, which results in production and economic loss. Therefore, early detection of incipient faults which are developed in bearing is utmost important. Condition-based maintenance strategies are been used for the diagnosis of faults occurring in rolling element bearing (REB). Numerous condition monitoring techniques such as vibration analysis, acoustic emission, oil debris analysis, temperature trend analysis, electrostatic and ultrasound were developed for fault diagnosis in REB. Vibration analysis based condition monitoring is one of the most-reported techniques in fault diagnosis of REB [1].

Cerrada et al. [1] extensively reviewed various approaches in fault detection and severity analysis of REB using ML techniques. Authors concluded that the Support Vector Machine (SVM) and Nearest Neighbor clusters are the most studied classifiers for fault detection in REB with vibration signals. Feature extraction, proper feature selection and hyper-parameter selection are the key pillars for accurate artificial intelligent fault diagnosis of AFB. This is due to non-stationary and nonlinear nature of vibration signals caused by operational conditions of machinery, such as variable load causes shaft speed fluctuations which leads to the difficulty in feature extraction and degradation in prediction accuracy of diagnosis method [2, 3]. Sugumaran and Ramachandran [4] had demonstrated SVM and Proximal SVM for fault detection of REB. In this, histogram features were extracted from time domain signal and results are compared with different features present in the feature

*Corresponding Author Email: *sangram.patil1989@gmail.com* (S. Patil)

set. SVM model using different kernel parameters were studied to select appropriate activation function. Decision Tree (DT) was used to select prominent features and study shows the Radial Basis Function (RBF) which gives good classification accuracy [5].

Sakthivel et al. [6] have presented soft computing techniques like Gene Expression Programming (GEP), Wavelet-GEP, SVM and Proximal SVM for eight different conditions in centrifugal pump. Time domain features were extracted from vibration signals and ranked based on DT with information gain and entropy reduction. Wang et al. [7] have employed Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) techniques for health value prediction of slewing bearing using a multiple sensor data such as temperature and torque. In this, ANFIS based model gives better health value prediction than ANN based models. Bansal et al. [8] used multi-class SVM based gear fault diagnosis technique with frequency domain data and analyze types of kernel functions utilized in training. Training datasets were obtained by interpolating and extrapolating angular speeds which are near to the speed of test data and study for optimization of a permissible range of speed variations for successful prediction of faults. SVM classifier is tuned by particle swarm optimization for hyper-parameter selection. Ensemble Empirical Mode Decomposition and optimized SVM based REB fault detection procedure explained by Zang and Zhou [9]. The optimization of SVM was carried out by grid search algorithm for proper section of $\gamma$ and $C$. Yin and Hou [10] surveyed the advancement in SVM for fault diagnosis and process monitoring in industrial systems.

Omar and Lior [11] have provided an exhaustive review of ensemble learning technique used for various applications. Hu and Min [12] used Gradient boosting decision tree model for automated detection of driver fatigue using EEG signals and results are found to be more efficient as compared to traditional machine learning algorithms like KNN, SVM and NN. Saleh and Farsi [13] classified the polarimetric synthetic aperture radar (PolSAR) data using ensemble classifier. In this, ensemble classifier with majority voting, Naïve Bays and multi objective heuristic combination rule were used. ETC ensemble technique was utilized for genre detection of music with numerical features [14] also used in biomedical application for image classification [15]. Kumar and Sahoo [16] proposed GA-RF hybrid classification system for diagnosis of cardiovascular diseases. The GA-RF system was compared with other combinations such as PCA, Relief-F etc. Several researchers are currently experimenting on the performance of ensemble ML methods for detection of ball bearing faults. Batista et al. [17] shows the application of ensemble SVM on simulated data with various level of noise. Each SVM is designed for a specific noise, after which all the SVMs are combined by an iterative Boolean combination technique which leads to degradation of error rate in the presence of high noise data. Zang et al. [18] proposed an ensemble based incremental SVM for fault diagnosis of REB. Multivariable features extracted from time domain and frequency domain vibration data are used as an input to SVM for ensemble learning. More desirable results in terms of classification accuracy and statistical analysis of classifier were obtained with ensemble-based SVM and compared with traditional SVM. Random Forest (RF) and ANN classifier performance are compared for multiple fault detection in electrical induction motor with time domain statistical features, and results show that RF gives precise classification accuracy [19]. Haideri [20] implemented a rule-based ensemble classifier technique for bearing fault detection. Genetic algorithm is used for feature selection. Three base classifiers are used to create an ensemble technique and the results were compared with an individual classifier. The literature review emphasizes on the importance of feature extraction, selection, hyper-parameter optimization and effectiveness of ensemble classifiers. It has been also reported in literature that SVM has been extensively utilized to classify REB fault detection over the past decade. Since, ensemble classifiers are implemented for AFB fault diagnosis, selection of number of estimators for training with the number of ranked features remained unattended by the researchers.

Therefore, in this paper, three ensemble ML techniques i.e. RF, GBC and ETC are utilized for fault detection of AFB. Time domain features extracted from temporal vibration signal from experimental test rig for different bearing conditions. These extracted features were ranked using two different feature selection techniques like DT and RLasso, based on this different rank feature sets were created. Results were obtained for ranked feature sets and different estimators with three ensemble classifiers and compared with often used SVM with RBF activation function.

## 2. ENSEMBLE TECHNIQUES

In ensemble classification, instead of one, a set of classifiers are used in order to make a prediction. Statistically, this helps in reducing the variance of classifiers and gives a better empirical performance. Random Forest, Gradient Boosting and Extra Tree Classifiers are the types of ensemble technique explained as follows:

**2. 1. Random Forest (RF)** is an adequate modification of bagging that builds a large set of unrelated trees and then averages them for prediction. Consider a total number of cases *'M'* to generate random forest. Firstly, for each tree, select a subset from the dataset using bootstrap sampling. For each subset grow a tree $R_i$ *(i= 1,2,3,…,M)* by randomly selecting *'n'* out of *'N'* total variables, out of those '*n*' variables pick a node *'d'* using Gini Impurity $(I_G)$ for best split point.

$$I_{G(p)} = 1 - \sum_{a=1}^{A} {P_a}^2 \tag{1}$$

Where, $P_a$is the fraction of instances tagged with class *'a'*. *'a'* is the number of classes *(a= 1,2,3…A)*. Then split the node into two daughter nodes and repeat the above procedure for growing trees for *'M'* number of cases. Collect the outcomes of each tree $\{C_m\}$ to predict a class label. For, each instance, the majority voting is run over $\{C_m\}_1^m$ and class prediction is achieved [21].

**2. 2. Gradient Boosting Classifier (GBC)** is a type of Ensemble technique which is used to develop the prediction model. It was proposed by Friedman [21] and typically uses DT classifier as a base classifier.

Consider a training set $(x_i, y_i)$ *(where,* $i = 1,2,3,......N$*)* and a differential loss function (i.e. 'deviance' in this case) $L(y_i, z)$ , where $z$ is the predicted value. Initialize the model as

$$f_o(x) = arg\ min_z \sum_{i=1}^{i=N} L(y_i, z) \tag{2}$$

Now, consider $m^{th}$ tree out of $M$ number of trees (*where* $m = 1,2,3,....M$). For each $m^{th}$ tree, the residual, i.e. the negative gradient is given by

$$g(x_i) = -\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \tag{3}$$

Let the size of each $m^{th}$ tree be $J_m$ and the tree region be given by $R_{jm}$ **(***where* $j = 1,2,3,.....J_m$.*)* For each observation in each tree calculate the gradient which is given by

$$\Theta_{jm} = arg\ min_z \sum_{i=1}^{i=N} L(y, z + f(x_i)) \tag{4}$$

Now, update the model of the succeeding tree with the gradient of previous tree. This is given by:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{j=J_m} \ell_{jm} * I \tag{5}$$

for each x which belongs to the region of that tree (i.e. $R_{jm}$). The overall output is given by $f_M(x)$ [21].

**2. 3. Extra Tree Classifier (ETC)** is more or less similar to RF classifier, except the top-down approach of splitting is replaced by a randomized process of splitting, which helps in decreasing the variance by increasing the bias of the tree. This is because the choice of optimal cut-point is responsible for a large amount of variance of induced tree. Unlike RF approach, ETC drops the idea of using bootstrap copies. Instead, it uses the whole learning sample. From the statistical point of view, this idea leads to an advantage in terms of split increasing bias, whereas the split-point randomization often has an excellent variance reduction. For example, if there are *N* total attributes in our training class and if *k* attributes are selected, the number of split-points is equal to *k*. Let these split points be denoted by *S (i.e.* $S_1$, $S_2$, $S_3$ *…*$S_k$*)*. These splits are chosen at random. A decision tree is created from every split. Each split returns a score in the form of the probability of selecting each class. Hence, for class A, the probabilities are given by $P_A$ *(i.e.* $P_{A1}$ ,$P_{A2}$ ,$P_{A3}$ ,...... $P_{Ak}$*)*. For finding the prediction, the probabilities of all classes are averaged and the class with the highest probability is chosen. This is also called majority voting. This complexity reduction helps the Extra Tree Classifier to produce better results in several high-dimensional complex problems and reduces the computational burden. [22].

**2. 4. Tuned SVM** is a supervised machine learning method derived as per statistical learning theory and mostly reported for bearing fault detection. RBF Kernel based Multi-class SVM [5] is used in this paper. Parameter tuning is done by grid search method [9] to obtain best values of γ and *C*. The tuned SVM is further used for computing classification accuracy for the same data and Results are used to compare with proposed Ensemble techniques.

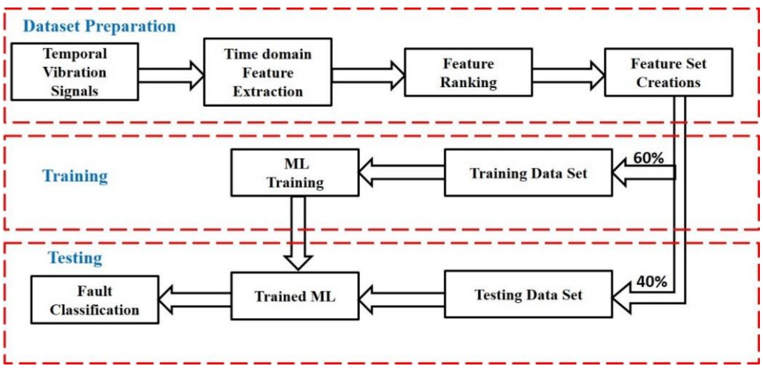



**Figure 1.** Fault diagnosis Methodology

## 3. METHODOLOGY

This fault diagnosis methodology has mainly three stages i.e. Dataset preparation, Training ML, and Testing as shown in Figure 1.

**3. 1. Test Rig** As shown in Figure 2, an experimental test rig has utilized to capture of raw vibration time domain signals for all specimen bearing conditions (Figure 3) with various speed and load combinations. The uni-axial accelerometer is mounted vertically on the bearing casing in a radial direction and plugged into the OROS data acquisition system. The detailed specifications of experimentations are shown in Table 1. In aadition 5 bearing conditions tabulated in Table 2.

**3. 2. Feature Extraction** Features represent the characteristic information present in the signal. They are represented as a k-dimensional dataset derived from the original m-dimensional dataset. Twelve statistical features are extracted from captured raw time domain signal a(n) for various combinations of bearing condition, speed and load. Table 3 list the selected statistical features with their mathematical formulations.


**Figure 2.** An Experimental Test Rig The labels are 1. Laptop; 2. Data acquisition system; 3. Variable frequency drive;4. Uniaxial Accelerometer; 5. Specimen bearing with housing; 6. Radial Load; 7. Dead weights

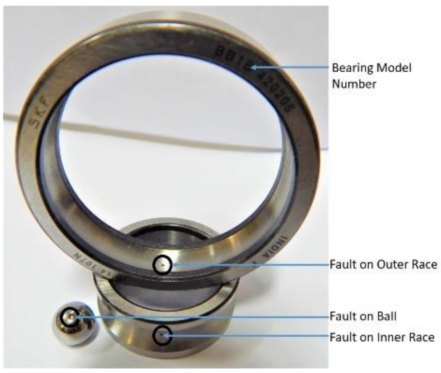



**Figure 3.** Specimen Bearing with faults on different components

**TABLE 1.** Experimental Parameters

| **Aspects** | **Specifications** |
|---|---|
| Bearing Make and Model | SKF BB1B4202015 |
| Speeds (rpm) (8 speeds) | 600 to 2700 (step size: 300) |
| Loads (N) (6 loads) | 15 to 40 (step size: 5) |
| Transducer | Uniaxial PCB made Piezoelectric ICP type Accelerometer (frequency range: 1Hz to 10kHz) |
| Sampling rate (samples/s) | 25600 (By Nyquist criteria) |
| Sampling time (s) | 5 |
| Number of Repetitions | 5 |

**TABLE 2.** Bearing Conditions

| | | |
|---|---|---|
| Types of bearing conditions (Classes) | 1. | Healthy bearing |
| | 2. | Inner race fault in bearing |
| | 3. | Outer race fault in bearing |
| | 4. | Fault in Ball of bearing |
| | 5. | Combine faults in bearing |

**TABLE 3.** Statistical Features [3,4]

| **Sr. No.** | **Feature** |
|---|---|
| 1. | $Root\ Mean\ Square(a_{rms}) = \sqrt{\frac{\sum_{n=1}^{N}(a(n))^2}{N-1}}$ |
| 2. | $Kurtosis(K_a) = \frac{\sum_{n=1}^{N}(a(n)-\mu_a)^4}{(N-1)\sigma_a{}^4}$ |
| 3. | $Mean\ (\mu_a) = \frac{1}{N}\sum_{n=1}^{N} a(n)$ |
| 4. | $Maximum(a_{max}) = \max\|a(n)\|$ |
| 5. | $Standard\ Deviation(\sigma_a) = \sqrt{\frac{\sum_{n=1}^{N}(a(n)-\mu_a)^2}{N-1}}$ |
| 6. | $Varience(\sigma_a{}^2) = \frac{\sum_{n=1}^{N}(a(n)-\mu_a)^2}{N-1}$ |
| 7. | $Skewness(SK_a) = \frac{\sum_{n=1}^{N}(a(n)-\mu_a)^3}{(N-1)\sigma_a{}^3}$ |
| 8. | $Peak\ to\ Peak(a_{p-p}) = \max(a(n))\ \text{-}\min(a(n))$ |
| 9. | $Crest\ Factor(CF_a) = \frac{a_p}{a_{rms}}$ |
| 10. | $Impulse\ Factor(IF_a) = \frac{a_p}{\frac{1}{N}\sum_{n=1}^{N}\|a(n)\|}$ |
| 11. | $Clearance\ Factor(CLF_a) = \frac{a_p}{(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\|a(n)\|}\ )^2}$ |
| 12. | $Shape\ Factor\ (SF_a) = \frac{a_{rms}}{\frac{1}{N}\sum_{n=1}^{N}\|a(n)\|}$ |

**3. 3. Feature Set Preparation** Time domain features express the various properties of vibration signals which plays vital role in classification. Some features have a higher weightage than others, which makes the latter redundant for classification. Feature ranking processes are carried out for measuring the contribution of each feature towards classification. Proper feature selection prevents overfitting and optimizes execution time. In this study, DT and RLasso techniques are used for feature ranking, the details of which are as follows:

**A. Decision Tree based Ranking**

DT works on the principle of simple decision-making rules worked out in a flow chart form to get the desired output. In DT, many subsets of the existing dataset are created to form decision nodes and leaf nodes. Decision nodes represent features and leaf nodes represent a decision. Here, it may be noted that in regression, the target is in the form of numerical values. The input is the feature set and output is the formation of decision tree with a decision called as root node which represents top most decision node of the tree corresponding to the best predictor. The working of the DT for feature ranking is given below [5]

*I).* The input i.e. a feature set X is given to the model. *II).* This input set is split into branches based on different attributes or features. *III).* Information Gain (IG) represents the power of respective feature and differentiate to its target class. IG is calculated at each split while growing a DT. IG (S,A) represents the relativity of feature (A) to collected examples (S) and is defined as:

$$IG(S,A) = Entropy(S) - \sum_{v=Value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (6)$$

Where the value (A) be the set of all feasible values and $S_v$ is a subset of S for which feature A has value $v$ and entropy (S) is a given by,

$$Entropy(S) = \sum_{i=1}^{n} -P_i \log_2 P_i \quad (7)$$

where *'n'* is the number of classes and $P_i$ is the ratio of *'S'* to the specific class *'i'*. *IV.* This process is repeated several times. After the last split, the average of the output of all existing branches is given to the related leaf node.

**B. Randomized Lasso (RLasso) Based Ranking**

The Lasso model is used to calculate the sparse coefficients by applying the weights repeatedly to calculate the Lasso score which represents the importance of features by minimizing the least square penalty. RLasso is an advancement in Lasso, in which the feature set is sub-sampled and the Lasso score is calculated based on random re-weighting of several times and scores used for feature ranking [23].

$$\beta^{\lambda,W} = argmin_{\beta \epsilon R^{P}} ||Y - X\beta||_2^2 + \lambda \quad (8)$$

Let, $W^k$ be independent distributed random variable in $(\alpha, 1)$ for $k= 1,2,...,p$. The randomized estimators $\beta\lambda, W$ for regularization parameter $\lambda \epsilon R$ is then,

## 4. RESULTS AND DISCUSSION

In this study, three ensemble techniques namely RF, GBC and ETC are used for fault detection of AFB. Characteristic information in terms of statistical features are extracted from time domain vibration signals, produced by simulating various conditions of AFB with variable speed and load on machinery fault simulator. Firstly, twelve time domain statistical features were measured by the vibration signals which are ranked using two ranking algorithms viz. DT and R-Lasso. Table 4 demonstrates the ranked features obtained by employing both the algorithms. SF and RMS are the top-ranked features in feature set using DT and RLasso, respectively. This is because DT uses the information gain and RLasso uses the linear model of regression for ranking. Using ranked features, twelve feature sets were created such that the first feature set contains the first top-ranked feature, then the second feature set contains the top two ranked features and so on. It enables to test the importance of features and the overall accuracy of the system. The impact of number of estimators is also tested by increasing the number of boosting stages from 10 to 100. This process allows us to find the best combination of estimators with a number of ranked features in classification accuracy which is presented in Tables 5 and 6.

Table 5 represents the testing accuracy and Figure 4 shows the execution time of machine learning techniques based on DT feature ranking.

**TABLE 4.** Feature Ranking

| Feature Ranking Method | Rank | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **DT** | *SF* | *p-p* | *K* | *SK* | *rms* | $\sigma^2$ | *max* | $\sigma$ | *CLF* | *CF* | μ | *IF* |
| **RLasso** | *rms* | *p-p* | $\sigma$ | *max* | $\sigma^2$ | *K* | *SF* | μ | *SK* | *CLF* | *CF* | *IF* |

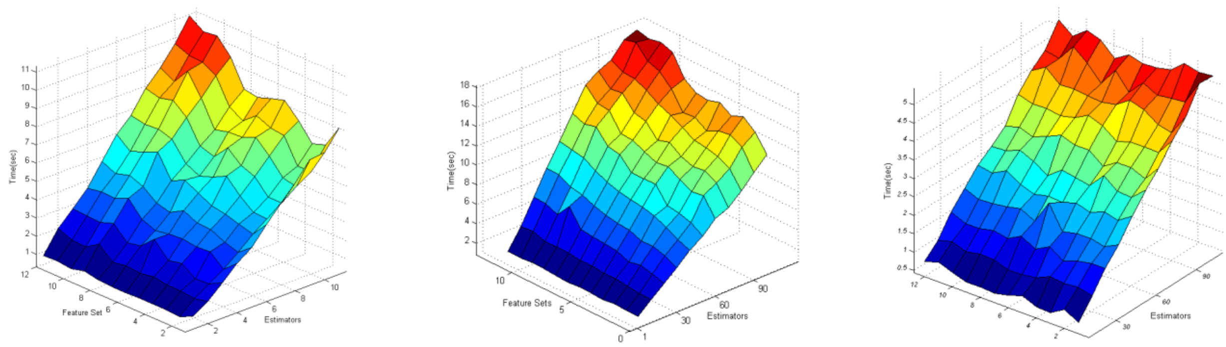
**TABLE 5.** Numeric Prediction of ML with DT Ranking Technique

| | | Feature sets | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Estimators** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **DT_RF** | **10** | 61.46 | 88.54 | 91.25 | 92.5 | 93.54 | 94.37 | 95 | 95.21 | 95 | 95.42 | 95.42 | 95.42 |
| | **20** | 61.42 | 88.75 | 92.71 | 92.7 | 93.54 | 93.96 | 95.2 | 95.42 | 95.63 | 96.04 | 96.46 | 95.83 |
| | **30** | 61.04 | 89.79 | 92.92 | 95.00 | 96.25 | 96.46 | 96.88 | 96.46 | 96.25 | 96.25 | 96.04 | 95.83 |
| | **40** | 60.00 | 89.38 | 91.88 | 94.58 | 96.25 | 96.25 | 96.25 | 96.25 | 96.25 | 95.83 | 96.25 | 96.46 |
| | **50** | 61.25 | 88.96 | 92.5 | 94.79 | 96.04 | 96.67 | 96.67 | 95.42 | 96.67 | 96.04 | 96.88 | 96.46 |
| | **60** | 60.63 | 89.38 | 92.5 | 95.21 | 96.04 | 96.88 | 96.88 | 96.46 | 96.46 | 96.25 | 95.83 | 96.46 |
| | **70** | 60.42 | 89.17 | 91.88 | 94.79 | 96.04 | 97.29 | 97.29 | 96.04 | 96.04 | 96.67 | 96.67 | 96.04 |
| | **80** | 60.63 | 89.79 | 92.71 | 94.58 | 96.04 | 96.25 | 96.46 | 96.25 | 96.04 | 95.83 | 96.67 | 96.46 |
| | **90** | 60.63 | 90 | 92.71 | 94.58 | 96.25 | 96.67 | 96.46 | 96.67 | 96.25 | 96.46 | 95.83 | 96.46 |
| | **100** | 60.21 | 88.75 | 92.5 | 94.79 | 96.46 | 96.88 | 96.46 | 96.67 | 96.04 | 96.67 | 96.25 | 96.04 |
| **DT_GBC** | **10** | 62.50 | 88.33 | 88.54 | 91.04 | 93.54 | 93.54 | 93.75 | 93.96 | 94.37 | 94.17 | 94.17 | 94.37 |
| | **20** | 63.75 | 87.71 | 90.63 | 93.33 | 94.37 | 94.79 | 94.37 | 94.79 | 94.79 | 95.21 | 95.21 | 95.21 |
| | **30** | 63.96 | 87.92 | 91.04 | 93.75 | 95.00 | 95.00 | 95.00 | 95.42 | 94.37 | 94.58 | 95.21 | 94.79 |
| | **40** | 61.46 | 89.17 | 91.25 | 93.96 | 95.00 | 95.21 | 95.00 | 95.00 | 94.79 | 94.79 | 95.00 | 94.79 |
| | **50** | 60.62 | 88.75 | 91.67 | 94.17 | 95.00 | 95.00 | 95.21 | 94.79 | 94.79 | 95.00 | 95.21 | 95.21 |
| | **60** | 61.04 | 88.96 | 91.67 | 94.58 | 95.21 | 95.42 | 95.21 | 94.58 | 95.00 | 95.42 | 95.21 | 95.00 |
| | **70** | 60.42 | 89.17 | 91.87 | 94.17 | 95.83 | 95.00 | 95.63 | 94.37 | 95.63 | 94.58 | 94.79 | 95.21 |
| | **80** | 60.21 | 89.38 | 91.87 | 94.58 | 95.63 | 95.21 | 95.42 | 94.58 | 95.21 | 94.79 | 94.79 | 95.21 |
| | **90** | 60.62 | 88.96 | 91.87 | 94.17 | 95.42 | 95.42 | 95.83 | 94.58 | 95.21 | 95.00 | 95.00 | 95.21 |
| | **100** | 60.42 | 89.38 | 91.87 | 94.79 | 95.00 | 95.21 | 95.63 | 94.79 | 95.00 | 95.42 | 95.00 | 95.42 |
| **DT_ETC** | **10** | 61.88 | 88.75 | 89.38 | 94.58 | 96.88 | 96.88 | 96.88 | 96.67 | 96.25 | 96.88 | 96.46 | 95.63 |
| | **20** | 60.42 | 89.38 | 90.83 | 95.63 | 97.08 | 97.29 | 97.29 | 97.08 | 96.46 | 97.08 | 96.25 | 96.88 |
| | **30** | 61.67 | 88.96 | 91.46 | 94.79 | 97.92 | 98.12 | 98.12 | 97.5 | 96.88 | 97.5 | 97.5 | 96.67 |
| | **40** | 61.67 | 91.04 | 93.13 | 95.83 | 97.71 | 97.92 | 97.71 | 97.08 | 97.29 | 97.29 | 97.71 | 96.67 |
| | **50** | 60.42 | 90.83 | 91.46 | 95.83 | 97.71 | 97.92 | 97.92 | 97.71 | 96.88 | 97.29 | 97.5 | 97.5 |
| | **60** | 60.83 | 90.63 | 92.29 | 96.04 | 97.71 | 97.71 | 97.5 | 97.5 | 97.71 | 97.71 | 97.29 | 97.08 |
| | **70** | 60.42 | 89.79 | 91.87 | 95.42 | 97.5 | 97.71 | 97.5 | 97.08 | 97.29 | 97.5 | 97.71 | 97.5 |
| | **80** | 60.83 | 89.79 | 92.71 | 96.25 | 97.71 | 97.92 | 98.12 | 97.29 | 97.5 | 96.88 | 97.29 | 97.08 |
| | **90** | 60.62 | 90.83 | 92.08 | 95.42 | 97.92 | 97.92 | 97.71 | 97.08 | 97.71 | 97.29 | 97.71 | 97.5 |
| | **100** | 60.83 | 90.63 | 92.29 | 95.83 | 97.92 | 97.71 | 97.92 | 97.29 | 97.5. | 97.92 | 97.08 | 96.88 |
| **DT_SVM** | | 58.33 | 72.7 | 79.16 | 84.34 | 85.62 | 84.99 | 83.75 | 83.54 | 85.20 | 84.58 | 84.16 | 83.12 |

Using this the accuracy of the Shape Factor is tested, which is the top-ranked feature and it is seen that the accuracy is nearly about 61% for all three ensemble techniques and 58% for tuned SVM. A sudden change is observed when the feature number is increased in the second feature set and from the second to the fourth feature set, no significant change is noted in testing accuracy. The peak accuracy for DT based classification is observed between the fifth to eighth feature set. After that, no significant change is observed in the testing accuracy. A testing accuracy of 97.3% is achieved for sixth and seventh feature set with 70 estimators for DT_RF. This indicates that first 6 features and 70 estimators are sufficient to obtain the highest testing accuracy. Also, Figure 4(a) clearly shows that the execution time increases as the feature set and estimators increases. For DT_GBC, highest accuracy of 95.8% is observed for 5$^{th}$ and 7$^{th}$ feature set which shows that 5 features and 70 estimators are sufficient to achieve highest testing accuracy using DT_GBC.

**TABLE 6.** Numeric Prediction of ML with RLasso Ranking Technique

| | | Feature sets | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Estimators** | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **RLasso_RF** | **10** | 70.63 | 87.29 | 86.46 | 85.00 | 87.08 | 92.29 | 94.58 | 93.75 | 95.83 | 95.42 | 94.37 | 95.63 |
| | **20** | 69.79 | 88.96 | 87.29 | 86.04 | 86.67 | 93.96 | 94.37 | 93.96 | 95.42 | 96.46 | 96.04 | 95.00 |
| | **30** | 69.79 | 88.75 | 88.75 | 86.67 | 87.29 | 93.54 | 95.42 | 92.92 | 96.67 | 95.83 | 96.46 | 95.63 |
| | **40** | 69.37 | 89.58 | 86.88 | 85.00 | 87.92 | 93.33 | 95.00 | 95.42 | 95.83 | 96.04 | 96.04 | 95.83 |
| | **50** | 69.37 | 87.71 | 87.71 | 86.25 | 87.71 | 93.75 | 94.37 | 94.37 | 95.83 | 96.25 | 95.83 | 96.04 |
| | **60** | 69.37 | 89.79 | 87.92 | 86.67 | 88.33 | 93.54 | 94.37 | 95.21 | 96.25 | 96.25 | 96.25 | 95.83 |
| | **70** | 69.58 | 88.12 | 88.12 | 87.08 | 87.50 | 93.54 | 94.58 | 94.79 | 96.04 | 96.67 | 96.04 | 96.25 |
| | **80** | 69.37 | 89.38 | 88.75 | 87.08 | 88.75 | 94.17 | 94.58 | 95.00 | 96.46 | 96.46 | 96.04 | 96.25 |
| | **90** | 69.58 | 89.79 | 87.92 | 86.46 | 87.92 | 93.33 | 95.00 | 95.21 | 96.67 | 96.04 | 95.83 | 96.46 |
| | **100** | 69.37 | 89.79 | 87.92 | 86.88 | 88.33 | 94.17 | 94.17 | 95.00 | 96.25 | 96.46 | 96.04 | 96.25 |
| **Rlasso_GBC** | **10** | 68.54 | 86.04 | 84.58 | 83.54 | 83.54 | 91.67 | 92.08 | 91.67 | 93.13 | 93.33 | 94.37 | 94.37 |
| | **20** | 71.67 | 88.75 | 87.50 | 87.08 | 87.08 | 90.42 | 93.54 | 92.92 | 94.58 | 95.21 | 95.21 | 95.00 |
| | **30** | 72.71 | 87.50 | 87.92 | 86.88 | 86.88 | 91.25 | 94.37 | 94.17 | 95.42 | 94.58 | 95.21 | 94.79 |
| | **40** | 72.71 | 86.25 | 88.33 | 87.08 | 87.29 | 91.87 | 93.75 | 94.17 | 95.00 | 94.58 | 95.00 | 94.79 |
| | **50** | 71.88 | 86.46 | 87.92 | 86.67 | 87.50 | 92.08 | 93.75 | 94.17 | 95.42 | 95.00 | 94.79 | 95.00 |
| | **60** | 70.00 | 87.29 | 88.33 | 87.29 | 87.29 | 92.08 | 93.96 | 94.37 | 95.21 | 95.00 | 94.79 | 95.21 |
| | **70** | 69.79 | 87.71 | 87.29 | 87.29 | 87.50 | 91.67 | 93.96 | 93.75 | 95.21 | 95.00 | 95.00 | 95.21 |
| | **80** | 69.79 | 88.12 | 87.92 | 87.29 | 87.92 | 92.71 | 93.54 | 94.17 | 95.63 | 95.21 | 95.42 | 95.00 |
| | **90** | 69.37 | 88.33 | 87.71 | 87.92 | 88.12 | 93.13 | 93.96 | 93.75 | 95.83 | 95.21 | 95.00 | 95.42 |
| | **100** | 69.37 | 89.17 | 87.71 | 87.29 | 87.29 | 93.13 | 93.96 | 94.17 | 95.42 | 95.00 | 95.63 | 95.42 |
| **RLasso_ETC** | **10** | 68.96 | 89.17 | 87.50 | 85.42 | 88.75 | 93.75 | 94.58 | 94.17 | 97.29 | 96.25 | 96.46 | 95.63 |
| | **20** | 69.17 | 88.75 | 86.25 | 86.25 | 88.54 | 94.58 | 96.25 | 94.79 | 96.67 | 96.04 | 96.88 | 96.25 |
| | **30** | 69.37 | 88.96 | 87.08 | 86.88 | 87.50 | 94.58 | 95.63 | 94.17 | 97.08 | 96.88 | 97.71 | 96.67 |
| | **40** | 68.54 | 88.54 | 86.88 | 87.71 | 88.33 | 94.79 | 95.83 | 95.42 | 96.88 | 97.71 | 97.08 | 97.50 |
| | **50** | 69.79 | 88.54 | 86.67 | 87.71 | 88.54 | 95.21 | 95.83 | 96.25 | 97.29 | 97.71 | 96.88 | 97.50 |
| | **60** | 69.37 | 88.75 | 86.46 | 87.29 | 88.33 | 94.79 | 95.42 | 96.04 | 97.08 | 97.50 | 96.88 | 97.29 |
| | **70** | 68.54 | 89.17 | 87.50 | 86.67 | 88.54 | 94.79 | 94.79 | 94.79 | 97.29 | 97.71 | 97.50 | 97.50 |
| | **80** | 69.37 | 88.75 | 87.08 | 86.88 | 88.54 | 94.58 | 95.00 | 95.21 | 97.50 | 97.29 | 97.71 | 97.50 |
| | **90** | 70.00 | 89.79 | 87.29 | 86.88 | 88.33 | 95.21 | 95.63 | 95.00 | 97.29 | 97.50 | 97.92 | 97.29 |
| | **100** | 69.58 | 88.96 | 87.92 | 87.50 | 88.12 | 94.58 | 95.42 | 95.42 | 97.29 | 97.71 | 97.29 | 97.50 |
| **RLasso_SVM** | | 61.46 | 71.87 | 71.66 | 72.08 | 72.08 | 77.5 | 79.16 | 82.29 | 83.12 | 83.75 | 84.16 | 83.12 |




(a) (b) (c)

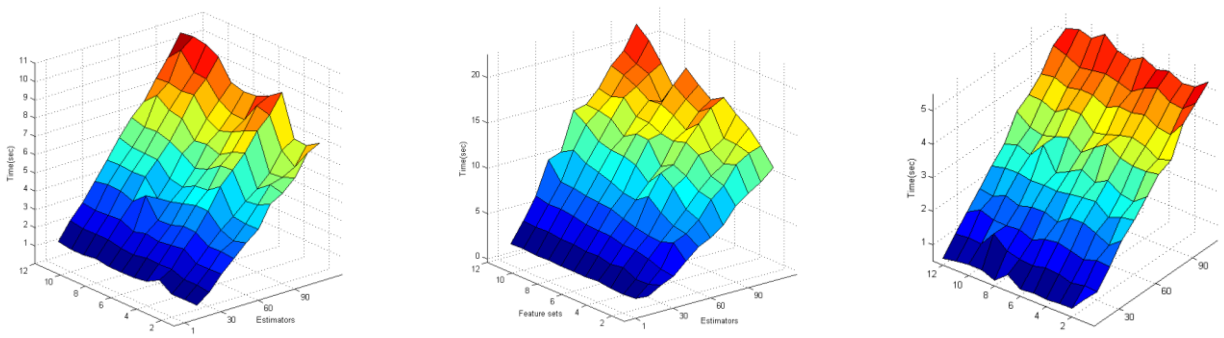
**Figure 4.** DT based feature ranking 3D plot for execution time vs feature set vs estimators(a) RF, (b)GBC, (c)ETC

For execution time it follows the same trend as DT_RF as shown in Figure 4(b), but overall execution time is more than for DT_RF. Peak testing accuracy of 98% is obtained with DT_ETC for 6$^{th}$ and 7$^{th}$ feature set with 30 estimators which proves that top 6 features with 30 estimators are enough to give peak testing accuracy. Figure 4(c) shows the execution time that increases as the number of estimators increases but less significant change is observed for the increased feature set. Also, the overall execution time is very less to RF and GBC. Moreover, tuned DT-SVM gives the highest classification accuracy of 85.6% with 5$^{th}$ feature set which is differentially lower than three reported ensemble techniques.

Similarly, Table 6 represents the testing accuracy of machine learning techniques with R-Lasso feature ranking technique. RMS is a top-ranked feature with R-Lasso and its individual significance with respect to testing accuracy is nearly around 70%. This accuracy is higher than the shape factor's accuracy, which is the top-ranked feature according to DT ranking. There is a sudden increase in the testing accuracy from 2$^{nd}$ feature set (i.e. top two features), which remains constant till the 5$^{th}$ feature set. From the 6$^{th}$ feature set, accuracy increases till the 9$^{th}$ feature set and above this there is less significant change in the testing accuracy. Also, for a particular feature set, no significant change is observed with increase in estimators. RLasso_RF gives the highest classification accuracy of 96.7% at the 9$^{th}$ and 10$^{th}$ feature set. It means that the top 9 features and 30 estimators are adequate for highest testing accuracy.

The time required for execution is depicted in Figure 5(a) and it follows the same trend as DT_RF. In line with this, RLasso_GBC based classification accuracy of 95.8% at the 9$^{th}$ feature set is obtained and its execution time trend is same as that of RLasso_RF but the overall time required for execution is more than RLasso_RF shown in Figure 5(b). Further, RLasso _ETC gives the highest accuracy of 97.9% among RLasso_RF and R_GBC with lower execution time. The execution time pattern is similar to that of DT_ETC as per Figure 5(c). Although RLasso_SVM has a classification accuracy of 84.16% at the 11$^{th}$ feature set, it is still lesser than the three ensemble techniques with the Rlasso feature ranking.

All the machine learning techniques which have been used, along with the two feature ranking methods according to their individual highest classification accuracies with their classification measures such as Cross Validation (CV), Kappa, Mean absolute error, F1 score and execution time are summarized in Table 7.




(a) (b) (c)

**Figure 5.** RLasso based feature ranking 3D plot for execution time vs feature set vs estimators(a) RF, (b)GBC, (c)ETC

**TABLE 7.** Summary of classifier

| Feature Ranking Method | Classifier | Number of features | Estimators | Classification accuracy (%) | CV accuracy (%) | kappa | R2 | Mean Absolute Error | F1 score | Time (s) |
|---|---|---|---|---|---|---|---|---|---|---|
| DT | RF | 6 | 70 | 97.29 | 96.40 | 0.9661 | 0.8800 | 0.07 | 0.97 | 6.52 |
| | GBC | 5 | 70 | 95.83 | 95.30 | 0.9479 | 0.8260 | 0.11 | 0.96 | 11.08 |
| | ETC | 6 | 30 | **98.12** | **97.77** | **0.9700** | **0.9100** | **0.05** | **0.98** | **1.95** |
| | SVM | 5 | - | 85.62 | 85.36 | 0.8046 | 0.2968 | 0.47 | 0.84 | 3.08 |
| Rlasso | RF | 9 | 30 | 96.67 | 95.37 | 0.9583 | 0.8656 | 0.09 | 0.97 | 4.40 |
| | GBC | 9 | 90 | 95.83 | 94.48 | 0.9479 | 0.8177 | 0.12 | 0.96 | 15.05 |
| | ETC | 11 | 90 | **97.92** | **96.65** | **0.9739** | **0.9062** | **0.06** | **0.98** | **4.64** |
| | SVM | 11 | - | 84.16 | 86.37 | 0.8020 | 0.2927 | 0.47 | 0.84 | 4.86 |

ETC is the superior classifier with 98.12 and 97.9% classification accuracy for both DT and RLasso Feature ranking techniques, respectively. Table 8 gives classification accuracy of each respective fault of top scorer DT_ETC amongst all reported machine learning techniques with best CV accuracy of 97.8% and least execution time of 1.95 s.

**TABLE 8.** Confusion Matrix for DT_ETC with 98.12% accuracy

| | | Predicted class | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| True class | 1 | 94<br>97.91% | 0 | 0 | 2<br>2.09% | 0 |
| | 2 | 0 | 95<br>98.95% | 0 | 0 | 1<br>1.05% |
| | 3 | 0 | 0 | 95<br>98.95% | 0 | 1<br>(1.05%) |
| | 4 | 2<br>2.09% | 0 | 0 | 94<br>97.91% | 0 |
| | 5 | 0 | 2<br>2.09% | 1<br>1.05% | 0 | 93<br>96.87% |

## 5. CONCLUSION

This paper presents, AFB diagnosis methods with three ensemble machine learning techniques viz. RF, GBC and ETC. An experimental temporal vibration signals were used as an input to machine learning techniques by extracting twelve-time domain statistical features. For training and testing, twelve feature sets were created based on the ranked features using DT and RLasso techniques. Experimental result shows that DT based feature ranking technique is more efficient than RLasso in terms of the minimum number of features required for training, which has considerable impact on computational time. Result also indicates, that the number of estimators has little impact in terms of the testing accuracy for DT based machine learning technique and has less significant impact for RLasso based machine learning technique. The Ensemble techniques gives better classification accuracy than the tuned SVM, which is the most employed classification technique. ETC gives best classification accuracy for both the ranking techniques, the highest classification accuracy 98.12% obtained by ETC with DT feature ranking. Though ETC gives best results, the other two ensemble classifier also outperforms tuned SVM with the same experimental data. The study shows that the possible implementation of the ensemble techniques can be applied for bearing fault detection.

## 6. ACKNOWLEDGMENT

Authors would like to acknowledge Larsen & Toubro Infotech Ltd. funding under CSR-1step initiative.

## 7. REFERENCES

1. Cerrada, M., Sánchez, R. V., Li, C., Pacheco, F., Cabrera, D., de Oliveira, J. V., and Vásquez, R. E., "A review on data-driven fault severity assessment in rolling bearings", ***Mechanical Systems and Signal Processing***, Vol. 99, (2018),169-196.
2. Ming, A. B., Zhang, W., Qin, Z. Y., and Chu, F. L., "Fault feature extraction and enhancement of rolling element bearing in varying speed condition", ***Mechanical Systems and Signal Processing***, Vol. 76, (2016), 367-379.
3. Chalouli, M., Berrached, N. E., and Denai, M., "Intelligent Health Monitoring of Machine Bearings Based on Feature Extraction", ***Journal of Failure Analysis and Prevention***, Vol. 17(5),(2017), 1053-1066.
4. Sugumaran, V., & Ramachandran, K. I., "Effect of number of features on classification of roller bearing faults using SVM and PSVM", ***Expert Systems with Applications***, Vol. 38(4), (2011), 4088-4096.
5. Saimurugan, M., Ramachandran, K. I., Sugumaran, V., and Sakthivel, N. R., "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine", ***Expert Systems with Applications***, Vol. 38 No. 4, (2011), 3819-3826.
6. Sakthivel, N. R., Nair, B. B., and Sugumaran, V., "Soft computing approach to fault diagnosis of centrifugal pump", ***Applied Soft Computing***, Vol. 12 No.5, (2012), 1574-1581.
7. Wang, H., Hong, R., Chen, J. and Tang, M., "Intelligent health evaluation method of slewing bearing adopting multiple types of signals from monitoring system", ***International Journal of Engineering, Transactions A: Basics***, Vol. 28, No. 4, (2015), 573-581.
8. Bansal, S., Sahoo, S., Tiwari, R., and Bordoloi, D. J., "Multiclass fault diagnosis in gears using support vector machine algorithms based on frequency domain data", ***Measurement***, Vol. 46(9), (2013), 3469-3481.
9. Zhang, X., and Zhou, J. "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines", ***Mechanical Systems and Signal Processing***, Vol. 41 No.1-2,(2013), 127-140.
10. Yin, Z., and Hou, J. "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes", ***Neurocomputing***, Vol. *174*,(2016), 643-650.
11. Sagi, O. and Rokach, L., Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 8 No. 4, (2018) p.e1249.
12. Hu, J. and Min, J., "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model". ***Cognitive Neurodynamics***, (2018), 1-10.
13. Saleh, R. and Farsi, H., "Optimum Ensemble Classification for Fully Polarimetric Synthetic Aperture Radar Data Using Global-local Classification Approach". ***International Journal of Engineering-Transactions B: Applications***, Vol. 31 No. 2, (2018), 331-338.
14. Murauer, B. and Specht, G., "Detecting Music Genre Using Extreme Gradient Boosting". In Companion of the The Web

Conference 2018. International World Wide Web Conferences Steering Committee., 1923-1927, (2018).

15. Maier, O., Wilms, M., von der Gablentz, J., Krämer, U.M., Münte, T.F. and Handels, H., “Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences”. ***Journal of neuroscience methods***, Vol. 240, (2015), 89-100.
16. Kumar, S. and Sahoo, G., “A Random Forest Classifier based on Genetic Algorithm for Cardiovascular Diseases Diagnosis”. ***International Journal of Engineering-Transactions B: Applications***, Vol. 30 No. 11, (2017), 1723-1729.
17. Batista, L., Badri, B., Sabourin, R., and Thomas, M, “A classifier fusion system for bearing fault diagnosis”, ***Expert Systems with Applications***, Vol. 40 No.17, (2013), 6788-6797
18. Zhang, X., Wang, B., and Chen, X., “Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine”, ***Knowledge-Based Systems***, Vol. 89, (2015), 56-85.
19. Patel, R. K., and Giri, V. K., “Feature selection and classification of mechanical fault of an induction motor using random forest classifier”, ***Perspectives in Science***, Vol. 8, (2016), 334-337.
20. Heidari, M., “Fault Detection of Bearings Using a Rule-based Classifier Ensemble and Genetic Algorithm”, ***International Journal of Engineering Transactions A: Basics***, Vol. 30 No. 4, (2017), 604-609.
21. Friedman, J., Hastie, T., and Tibshirani, R., “The elements of statistical learning”, New York: Springer series in statistics, (2001).
22. Geurts, P., Ernst, D., and Wehenkel, L. "Extremely randomized trees." ***Machine learning,*** Vol. 63, (2006), 3-42.
23. Meinshausen, N., and Bühlmann, P., "Stability selection." ***Journal of the Royal Statistical Society: Series B (Statistical Methodology)***, Vol. 72 No. 4, (2010), 417-473.

# Fault Detection of Anti-friction Bearing using Ensemble Machine Learning Methods

S. Patil[a], V. Phalle[b]

[a] *Centre of Excellence in Complex and Nonlinear Dynamical Systems (CoE-CNDS), Veermata Jijabai Technological Institute, Mumbai, India*
[b] *Mechanical Engineering Department, Veermata Jijabai Technological Institute, Mumbai, India*

*PAPER INFO*

*Paper history:*
Received 30 June 2018
Received in revised form 26 August 2018
Accepted 26 October 2018

*Keywords*:
*Anti-friction Bearing,*
*Ensemble Learning,*
*Vibration Signal*
*Fault Detection*

**چکیده**

AFB یکی از اجزای بسیار مهم ماشین است و شکست برنامه‌ریزی نشده آن منجر به سوء عملکرد در طیف وسیعی از ماشین آلات چرخشی می‌شود که ناشی از خرابی‌های غیرمنتظره و زیان‌های اقتصادی است. در این مقاله، تکنیک‌های یادگیری ماشین‌های گروه برای تشخیص گسل‌های AFB مختلف نشان داده شده است. ابتدا ویژگی‌های آماری از سیگنال‌های ارتعاش زمانی استخراج شده و با استفاده از تست‌های آزمایشگاهی برای پارامترهای مختلف ورودی مانند بار، سرعت و شرایط تحمل، جمع‌آوری شد. این ویژگی‌ها با استفاده از دو تکنیک، یعنی (DT) و (RLasso) رتبه‌بندی می‌شوند، که بیشتر برای ایجاد آموزش و آزمایش مجموعه‌های ورودی برای تکنیک‌های یادگیری ماشین استفاده می‌شود. از سه تکنیک یادگیری ماشین‌آلات برای طبقه‌بندی خطاهای AFB یعنی RF, GBC و طبقه‌بندی اضافی درخت (ETC) استفاده می‌شود. تاثیر تعدادی از ویژگی‌های رتبه‌بندی شده و برآوردگرها برای تکنیک‌های گروهی مورد مطالعه قرار گرفته است. نتیجه نشان می‌دهد که کارایی طبقه‌بندی به طور قابل ملاحظه‌ای از تعداد ویژگی‌ها تأثیر می‌پذیرد، اما تأثیر تعداد برآوردگرها جزئی است. تکنیک‌های نشان داده شده، دقت بیشتری در طبقه‌بندی در مقایسه با SVM تنظیم شده با داده‌های آزمایشی مشابه داده می‌شود. بالاترین دقت طبقه‌بندی خطای AFB با دقت ۹۸/۱۲٪ با رتبه‌بندی ETC و DT بدست می‌آید.

***doi**: 10.5829/ije.2018.31.11b.22*