Rolling Element Bearing Fault Detection using Statistical Features and Ensemble Classifiers

Chhaya Grover, Neelam Turk

Abstract: Rolling element bearing health condition is monitored by analysing its vibration signature. Raw vibration signal picked up through suitably placed accelerometers is difficult to analyse hence many signal processing techniques have been proposed and developed by researchers to process the data for suitably extracting an effective signal feature set. Various machine learning techniques have been used for interpretation and accurate fault diagnosis using this extracted feature set. In this study "Empirical mode decomposition" is used for pre-processing the raw vibration data. Six "Statistical features" are extracted from the best Intrinsic mode function obtained through EMD and "Ensemble machine learning classifiers" are used for bearing fault diagnosis. A stacked ensemble of five classifiers is proposed for accurate fault diagnosis and results are compared with conventional ensemble classifiers to prove its effectiveness.

Keywords: Empirical mode decomposition, Ensemble classifiers, Statistical features, Vibration signature analysis

I. INTRODUCTION

In the 'inner race', 'outer race' or 'rolling element' of a rolling element bearing, occurrence of localized faults may take place due to the factors such as high temperature, high electrical discharge and corrosion. Whenever a damaged rolling element strikes the inner or outer race or the healthy rolling elements strike a fault on the inner or outer race, high frequency vibrations are generated. These characteristic vibration signatures can be analysed to get an insight into bearing health. Researchers have proposed many schemes for vibration signature analysis in the past. Adequate Fault sensitive features have been extracted from the vibration signals and analysis has been done using intelligent decision making techniques. A review of various vibration feature extraction techniques in time domain, frequency domain, and joint time frequency domain for fault diagnosis in rotating machines is presented by Yang et al.[1] and Gulez & Badi [2]. Statistical time domain features have been effectively used by Delgado et al [3], Nayana B.R. and Geethanjali P. [4] and Kankar, Sharma & Harsha [5] in their extracted feature set along with various other features. In addition to investigating various fault sensitive features for vibration signal analysis, researchers have been exploring various artificial intelligence methods for accurate fault diagnosis. A comprehensive review of "Artificial algorithms" used in rotating machinery fault diagnosis has been done by Liu et al. [6].

   Chhaya Grover*, Department of Electronics Engineering, J.C. Bose University of Science and Technology, YMCA, Faridabad, India.
   E-mail: chhayagrover@jssaten.ac.in
   Dr. Neelam Turk, J. C. Bose University of Science and Technology, YMCA, Faridabad, India.

They have discussed advantages and limitations of KNN, Naïve Bayes, SVM, ANN and Deep learning algorithms. Zhang et al. [7] have presented a systematic review of "Machine learning and deep Learning algorithms" for bearing fault diagnostics. A review on "Meta Classification Algorithms" using WEKA has been presented by Bal and Sharma [8] in their research paper.

Literature shows a growing research interest in combining a set of learning algorithms to generate ensembles for investigating complex problems. These ensemble algorithms tend to exploit the strengths of the base classifiers to enhance the overall accuracy. Kotsiantis et al. [9] in their paper, titled "Machine learning: A review of classification and combining techniques", have described ensembles of classifiers for improving classifier accuracy. Dietterich T. [10] have reviewed Ensemble Methods in Machine Learning and explained why ensemble classifiers often perform better than any single individual classifier.

Ensembles methods have been applied to a wide range of industrial problems in the area of condition monitoring. There is empirical evidence of the effectiveness of this approach in fault diagnosis of rotating machines. An ensemble of rule-based classifiers for fault diagnosis of rotating machinery was proposed by Dou et al. [11] to predict potential faults and subsequent breakdown of rotating machinery. A classifier ensemble was constructed and validated on the vibration data of two types of bearings: SKF6203 and NU205. Zio, Baraldi and Gola [12] proposed feature-based classifier ensembles for multiple fault diagnosis in rotating machinery. In this work, a multi-objective genetic algorithm is used for feature selection and ensembles of classifiers is developed to achieve higher accuracies. They used voting technique to effectively combine the predictions of the base classifiers. Sikder et al. [13] pre-processed vibration data using FFT and then applied an ensemble learning method , Random Forest for bearing fault diagnosis. The validation for the proposed scheme was done on the Case Western Reserve University (CWRU) dataset. Sharma, Amarnath and Kankar[14] used 15 time domain, frequency domain and wavelet-based features in feature vector and applied ensemble techniques namely rotation forest and random subspace for fault diagnosis.

Gaowei Xu et al. [15] proposed a bearing fault diagnosis method based on deep convolutional neural network (CNN) and random forest (RF) ensemble learning. They generated two dimensional gray-scale images from one dimensional time domain vibration signals, extracted multi-level features using convolution neural network and used ensemble of multiple Random Forest classifiers for classification of faults.

Karimi and Jazayeri-Rad applied boosting methods to compare the fault diagnosis performances of single neural networks with two ensemble neural networks [16].

In this paper an Ensemble classification approach for rolling element bearing fault diagnosis is proposed using six simple and conventional statistical features. First, Empirical Mode Decomposition of the vibration signals is performed to obtain Intrinsic Mode Functions (IMFs). Second, six statistical parameters are extracted from representative IMFs. Third, proposed ensemble classifiers and conventional Ensemble classifiers are used to diagnose three types of conditions in rolling element bearings: normal (faultless), fault in inner race and fault in outer race and performance evaluation is done using 7 evaluation metrics.

The rest of the paper is arranged as follows: Section II provides a theoretical framework of EMD, six statistical Parameters used in this study and Ensemble classifiers. Section III deals with the methodology, section IV reports the obtained results and section V deals with conclusion and scope for future work.

II. THEORITICAL FRAMEWORK

A. Overview of Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) was proposed by Huang et al. as a mathematical tool to analyse a non-stationary and non-linear signal by decomposing it into different Intrinsic Mode Functions (IMF)[17].

Steps of EMD Algorithm are listed below:

Step1: All of the local maxima and local minima points for the input vibration x(n) signal are calculated.

Step2: Upper envelope is derived by connecting all local maxima points using cubic spline function.

Step3: Lower envelope is derived by connecting all local minima points using cubic spline function.

Step4: Mean value of upper and lower envelopes is calculated.

Step5: Updated signal is derived by subtracting the mean calculated in above step i.e. $x(n)_{new} = x(n)$ - mean ,

and steps 1 to 5 are repeated on the updated signal until it fulfils the conditions to be considered as an IMF i.e. the maximum difference between the number of extrema and the number of zero crossings is 1 and the mean value of the both of the envelopes is zero.

Step6: Residue (r_m) is calculated by subtracting first IMF from x(n) i.e. $r_m = x(n)-IMF_1$

Step7: Steps 1-6 are iterated on the residue to find all the IMFs of the vibration signal

The algorithm will terminate when the residue becomes a monotonic function. Thus, after Empirical mode decomposition, the vibration signal x(n) can be represented as a sum of "m" IMFs (IMF_i, where i=1 to m) and residue (r_m)

$$x(n) = \sum_{i=1}^{m} IMF_i + r_m \quad (1)$$

B. Overview of Statistical Features

Statistical techniques have been used for alarm purposes in industrial plants in case of failures. K. Tom prepared a detailed "Primer on Vibrational Ball Bearing Feature Generation for Prognostics and Diagnostics Algorithms".[18] The report gives an overview of various techniques for feature generation for prognosis and diagnosis of bearing faults.

In this work following Histogram features (Histogram Upper Bound and Histogram Lower Bound) and Moments (1st, 2nd, 3rd and 4th) have been used as features for bearing fault diagnosis:

Histogram: A Histogram or a Discrete Probability Density Function provides a visualization profile for vibration data and can be used as a tool to characterize it. Vibration data from healthy bearing has Normal Gaussian distribution whereas there is proportional increase in the number of high levels of acceleration in the vibration data collected from a damaged or faulty bearing, This results in non-Gaussian PDF[18].

Histogram upper bound (HL) is defined as

$$HU = \left(x(n)\right) + \frac{\Delta}{2} \quad (2)$$

Histogram lower bound (HL) is defined as

$$HL = \left(x(n)\right) - \frac{\Delta}{2}$$
$$(3)$$

where $\Delta = \left(x(n)\right) - Min\left(\frac{x(n)}{N-1}\right)$

Moments: Four central statistical moments (i.e. moments calculated about the mean) have been used in the feature vector[18].

1st Statistical Moment (Mean): The first statistical moment is the average or mean value of a signal. This is the DC value of the signal. Mean of sampled vibration signal $x(n) = [x(1), x(2), x(3) \dots \dots \dots x(N)]$ having N Samples is given as:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x(n) \quad (4)$$

2nd Statistical Moment (Variance): Second statistical moment is the variance of the vibration signal i.e. $Variance = \sigma^2$ where σ is the standard deviation of the discrete vibration signal $x(n)$ given as:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N} [x(n) - \mu]^2} \quad (5)$$

3rd Statistical Moment (Skewness): Third statistical moment is called skewness. It captures asymmetry in data distribution. It can be positive or negative i.e. data can be skewed left or right. Skewness of vibration data can change significantly in the presence of a fault in the bearing.

$$SK = \frac{\sum_{n=1}^{N} [x(n)-\mu]^3}{(N-1)\sigma^3} \quad (6)$$

4th Statistical Moment (Kurtosis): Fourth statistical moment is called Kurtosis. It is a measure of the steepness of the data distribution, a negative kurtosis value indicates a flat distribution relative to a normal distribution. Changes in kurtosis due to faulty bearings can be used to identify types of faults.

$$KU = \frac{\sum_{n=1}^{N} [x(n)-\mu]^4}{(N-1)\sigma^4} \quad (7)$$

C. Overview of Ensemble Classifiers

As enormous amounts of vibration data are collected for bearing health monitoring, it becomes important to integrate different concepts for intelligent decision making for accurate fault diagnosis. There have been numerous studies in the past decade for combining machine learning classifiers into an ensemble. There are many ways by which base classifiers can be combined together to generate Ensemble Classifiers, which are proven to outperform any single classifier within the ensemble. Ensemble methods combine the predictions from multiple models and have emerged as a powerful machine learning technique for improving the accuracy and robustness of classification.

A detailed description of ensemble classifiers is given in the book titled "Combining Pattern Classifiers: Methods and Algorithms" by L.Kuncheva [19]

In this work, five ensemble machine learning algorithms are applied on the MFPT dataset which is described in detail in the next section. The five algorithms are:

1. **Bagging:** Bagging (Bootstrap Aggregation) is based on estimation of a statistical parameter like mean from numerous random samples of data. According to L. Breiman [20], from training data set, numerous random samples are drawn (and later replaced) to train multiple machine learning models. Prediction is made by each model and the results are averaged to reduce the variance of predictions.

2. **Random Forest:** It is an extension of bagging ensemble classifier given by L. Breiman [21]. Bagged decision trees have a shortcoming that greedy algorithm is used to select the best split point for building trees. Because of this, generated trees look quite similar and the variance of the predictions from different bags gets reduced, which ultimately effects the robustness of the predictions. Random Forest decreases the similarity between the bagged trees by disrupting the greedy algorithm during tree generation. This leads to the use of random subset of the input attributes to generate split points.

3. **Boosting:** Boosting ensemble method uses machine learning models in succession to boost the prediction outcomes by removing the errors in predicted outcomes by previous models.

   a. AdaBoost: Y. Freund and Robert E. Schapire [22] discusses the use of decision tree models having a single decision point. The construction of first model is done by weighing each instance in the training dataset and continuously updating the weights based on the overall accuracy of the model. Next models are trained and added in succession. The process continues until no further improvements are possible.

   b. LogiBoost: It performs classification using additive logistic regression scheme and can handle multi-class problems.

   c. MultiBoost: It is a combination of AdaBoost and an improved version of bagging called wagging. MultiBoost uses base learning algorithm C4.5 to generate decision tree. It combines the high bias and variance reduction properties of AdaBoost with excellent variance reduction property of wagging.

4. **Voting:** Voting works by taking two or more sub-models for making predictions. Finally, the predictions are combined based on some criteria e.g. by taking the average of the predictions. J. Kittler [23] and L. Kuncheva [19] have discussed Voting algorithm in detail.

5. **Stacking:** An extension to voting ensembles is stacking. In this ensemble method, multiple sub-models are selected and instead of taking the average of predictions, another supervisor model is trained to combine the predictions from the sub-models to give best outcome.

III.   PROPOSED METHODOLOGY

For validation of the proposed scheme and comparative evaluation of various ensemble classifiers, the dataset provided by the Machinery Failure Prevention Technology (MFPT) Society [24] has been used in this study. The MFPT data was acquired from a NICE bearing [25] having roller diameter of 5.97 mm, pitch diameter of 31.62 mm, contact angle of 00 and an input shaft rate of 25 Hz. A single radial accelerometer has been used to obtain the data. The acquired data is stored in a MATLAB® double-precision, binary format .mat file. In addition to acceleration data, the data files also include sampling rate, shaft rate and load. The dataset consists of data files collected from a bearing test rig and also from real machines. Data collected from test rig includes 3 files of healthy bearing under fixed load, 3 files with vibration signals from bearing having outer race faults under fixed load, 7 files with vibration signals from bearing having outer race faults under seven types of loads and 7 files with vibration signals from bearing having inner race faults under seven types of loads. Data related to following three conditions has been used for this study:

1. Baseline (No Fault): 3 baseline or healthy conditions with a sample rate of 97,656 Hz and 270 lbs of load recorded for 6 sec.

2. Fault in Outer Race: 7 outer race fault conditions with a sample rate of 48,828 Hz and various loads of 25, 50, 100, 150, 200, 250, 300 lbs recorded for 3 sec.

3. Fault in Inner Race: 7 inner race fault conditions with sample rate of 48,828 Hz. and various loads of 0, 50, 100, 150, 200, 250, 300 lbs recorded for 3 sec.

In order to match the sample rate of other fault sets, we down sampled the baseline data set to 48,828 Hz. The original vibration signals were split into pieces each having 2048 points.

Table-I: Classes for the MFPT dataset.

Class	Bearing Condition	No. of samples
Normal	No fault	430
IR	Inner Race	498
OR	Outer Race	498

A novel approach for rolling element bearing fault diagnosis is presented in this study utilizing- Empirical Mode decomposition, Statistical features and Stacked ensemble classifier.

The methodology of this approach can be explained in following steps:

1. Collecting the raw bearing vibration signals from MFPT repository.
2. Performing EMD on the vibration signals to generate Intrinsic mode functions.
3. Finding best IMF for each of the vibration signal based on correlation coefficient.
4. Calculating six statistical parameters (HL, HU, μ, VAR, SK, KU) from best IMFs of vibration signal to generate feature vectors.
5. Divide the dataset into training and test data in 80%:20% ratio. Input feature vector from training data set to the ensemble classifier and train the model.
6. Input test dataset features to obtain rolling element fault classification results as normal (faultless), outer race fault and inner race fault.

The process steps of the proposed fault diagnosis scheme are shown in the form of a flow chart in Fig. 1.

Empirical Mode Decomposition, best IMF selection and Statistical parameter extraction are performed using MATLAB R2018a. The Ensemble classifiers used are: Bagging, Random Forest AdaBoost, Logiboost, Stacking and Voting. For experimentation, these classifiers have been trained and tested in Weka, a software containing java packages for machine learning algorithms. The MATLAB scripts and Weka models have been run on a system with Intel Core i5-7200U CPU @ 2.50 GHz and 8 GB RAM.

A. Data Pre-processing

Data samples of size 2048 points are extracted from the vibration signals obtained from MFPT data set, giving a total of 1423 samples consisting of 429 normal (healthy) and 994 (faulty). Out of 994 samples obtained through faulty bearing, 497 samples are with fault in inner race of bearing and 497 samples are with fault in outer race of bearing. The 1423 vibration signals are first individually decomposed into a set of IMFs (Intrinsic mode functions) by applying EMD algorithm. The best IMF i.e. the IMF having the highest correlation with the parent signal is referred as the representative signal. Only representative signals are considered for the remaining steps.

To increase the number of instances in our dataset in a well-balanced manner, Synthetic Minority Oversampling Technique (SMOTE) developed by N. Chawla, Bowyer K.W. et al.[26] is used. This statistical technique increased the number of instances in our dataset in a balanced way by generating new instances from originally existing minority cases in the training dataset. The amount of SMOTE percentage taken is 25% and number of nearest neighbours is taken as 5. The algorithm took samples and its 5 nearest neighbours for each fault class, and combined features of the target case with features of its neighbours to generate new cases. After applying SMOTE a total of 1205 instances (407 IR, 392 OR and 406 Normal) are obtained from the original 1138 instances of training dataset.
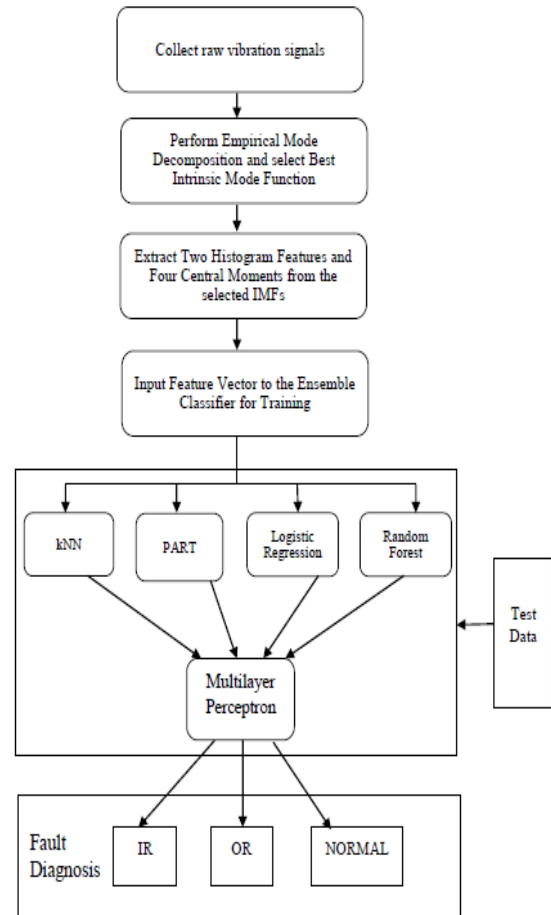

Fig. 1. Proposed Fault Diagnosis Scheme

B. Feature Extraction

The six parameters are extracted from the representative signals using equations (2), (3), (4), (5), (6) and (7). These parameters are then used to create the feature vector. The obtained data matrix has 1424 rows and 7 columns. This data is then normalized by rescaling each attribute to a range of 0 to 1 using the equation (8) for every numeric attribute 'x'.

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (8)$$

The data is split into two parts in 80:20 ratio; training data- 80% and test data-20%.

C. Fault Classification and Performance Evaluation

The training data is fed to Weka's package for Meta classifiers. In bagging, REPTree - a standard decision tree, is configured as the model being bagged. REPTree builds a decision tree using either information gain or variance and prunes it by reduced-error pruning. The size of each bag is taken the same as that of the training dataset, to generate a new sample of different composition. A total of 100 iterations are performed on the dataset. Keeping all these parameters same for Random Forest classifier, the model is trained on the training data set. In all of three boosting models (AdaBoost, Logiboost and Multiboost) weak learner is chosen as REPTree algorithm and 10 numbers of iterations are performed.

Rolling Element Bearing Fault Detection using Statistical Features and Ensemble Classifiers

In Voting ensemble, four classification sub-models that can make uncorrelated predictions are selected. The selected sub-models are KNN, PART, Logistic Regression and Random Forest. For combining the predictions of the sub models the parameter chosen is average of probabilities. In the proposed stacked ensemble classifier, four sub models are chosen; one lazy classifier-KNN, one rule based classifier-PART, one function classifier -Logistic Regression and Random Forest which is a tree based classifier. The supervisor model taken is Multilayer Perceptron which is trained to combine the predictions from the sub model in the best possible way. It uses backpropagation to classify instances.

The classification accuracy on Training dataset is estimated by stratified 10-fold cross validation. In every trial, a classifier is trained on any 9 folds and validated on the remaining fold. For each classifier, Training and Testing accuracy & time required to build classifier model are recorded in Table-II. Five additional evaluation metrics (MAE, MCC, F1-Score, Entropy, AUROC) are calculated on Test dataset and are recorded in Table-III. The mathematical expressions of these metrics are listed as below:

Accuracy: It is defined as the ratio of correctly classified instances (i.e. sum of True positive and True negative instances) to the total number of instances.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

where, TP-True positive, FP-False positive, FN-False negative and TN-True negative

Mean Absolute Error (MAE): Sum of absolute errors for all instances divided by the number of instances is called mean absolute error.

$$MAE = \frac{1}{N}\sum_i |\hat{x}_i - x_i| \qquad (10)$$

where $\hat{x}_i$: predicted label, x_i: true label, N: number of instances.

Matthews Correlation Coefficient (MCC): It is a correlation coefficient that indicates the correlation between predicted class and actual class. It can be calculated mathematically using TP, FP, TN and FN values as:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TN+FN)(TN+FP)(TP+FN)}} \qquad (11)$$

F1 Score: It is calculated by taking the harmonic mean of the precision and recall. Given a threshold value the F1 score provides a measure for goodness of classifier.

$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (12)$$

Mean Scheme Entropy: This is the entropy per instance for the classification scheme. The cross-entropy for the classification model across the entire training dataset is required to be minimized. So this is calculated by calculating the average cross-entropy across all training examples. Cross-entropy for classification of multi class problem can be calculated as:

$$-\sum_{l=1}^{C} y_{o,l}log(p_{o,l}) \qquad (13)$$

Where C: number of classes, l: class label, o: observation and p: predicted probability for observation o of class l and y: binary indicator

Area under the ROC curve (AUROC): This metric calculates the area under the ROC curve, which is a graph plotted between sensitivity and (1-specificity) of a classifier.

Time taken to build Model: This is the time elapsed to train the classifier model.

A bar plot showing the accuracy values for the training and test data is shown in Fig. 3. Bar charts for all other metrics are shown in Fig. 4.

IV. RESULT AND DISCUSSION

The statistical significance of ensemble classifiers is validated by training and test data. Seven metrics are used to compare the efficacy of the models. These metric values are recorded in Table-II and Table-III.

From Table-II and Fig.2., it is observed that the performance of proposed stacked ensemble classifier is the best with respect to accuracy although the time taken for building this ensemble classifier model is highest. The 10-fold cross validation accuracy obtained with training dataset is 95.1867 % while accuracy obtained with test dataset is 92.6316 %. Although with reference to MAE, Multiboost gives better performance but its accuracy is very low as compared to proposed classifier.

It is also inferred that proposed ensemble classifier performs decently well in this metric too. It outperforms other ensembles in MCC with the highest value of 0.927. As MCC value of 1 indicates a perfect classifier, it can be said that proposed classifier is performing very well.

Moreover, as F1-score value of 1 indicates perfect precision and recall and our proposed model gives an F1-score of 0.923, it is verified that our model is a good model that correctly distinguishes between three bearing fault classes.

Another metric that has been used for evaluation of the performance of proposed classifier is the area under the Receiver Operating Characteristics(ROC) curve, or AUROC. AUROC is a good indicator of model's performance and captures both sensitivity and specificity. In addition to that AUROC score considers the rank of each prediction instead of its absolute value, hence it is independent of the threshold set for classification. This metric value for the proposed classifier is 0.978, highest amongst all classifiers. The proposed classifier also shows remarkable improvement in mean entropy.

The pictorial representation of performance evaluation of all classifiers is shown in the form of bar charts in Fig. 3 and 4. The good performance of proposed stack ensemble on the test set can also be seen from the confusion matrix plotted in Fig.5. and metric values by individual class reported in Table -IV.

Table-II: Accuracy and Time taken for model building for the different classifiers.

Classifier	Accuracy		Time taken for model building
	Training	Testing	
Random Forest	94.9378 %	91.5789 %	0.15 seconds
Bagging	94.9378 %	91.9298 %	0.13 seconds
AdaBoost	93.7759 %	90.8772 %	0.06 seconds
Logiboost	94.1909 %	90.5263 %	0.57 seconds
Multiboost	94.2739 %	91.5789 %	0.05 seconds
Voting	94.8548 %	92.2807 %	0.26 seconds
Proposed Stacked Ensemble	95.1867 %	92.6316 %	4.20 seconds

Fig.2. Comparison of Training and Testing Accuracies of classifiers
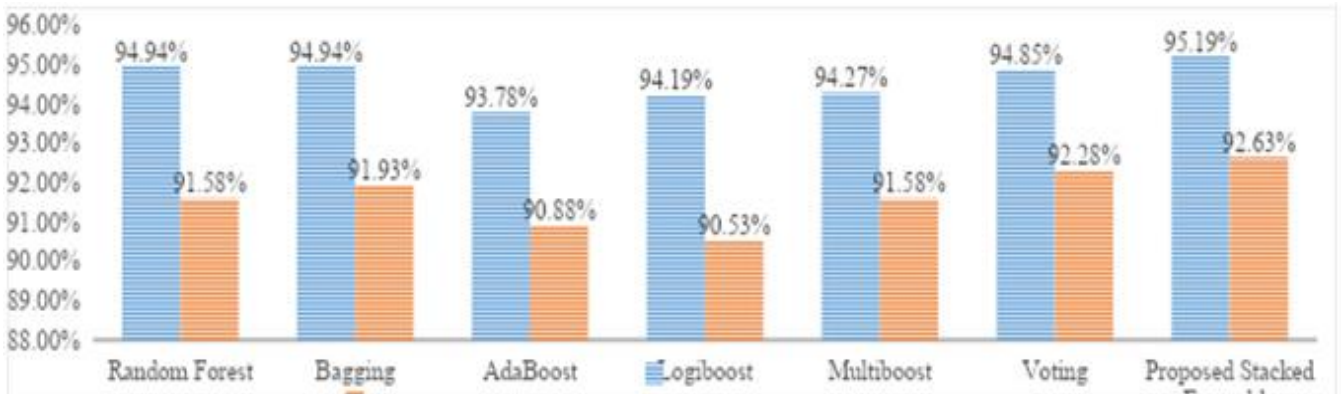

Fig.3. Comparison of Five Metric Values of classifiers

Table –III: Metric values obtained for the different classifiers on Test dataset.

Classifier	MAE	MCC	F1 Score	Mean Entropy	AUROC
Random Forest	0.0704	0.874	0.916	1.53	0.974
Bagging	0.0738	0.879	0.920	0.23	0.968
AdaBoost	0.0645	0.862	0.909	0.42	0.955
Logiboost	0.0817	0.858	0.906	0.27	0.965
Multiboost	0.0557	0.874	0.916	0.84	0.963
Voting	0.0737	0.884	0.916	0.22	0.978
Proposed Stacked Ensemble	0.0656	0.927	0.923	0.25	0.978

Rolling Element Bearing Fault Detection using Statistical Features and Ensemble Classifiers

(a)

(b)

(c)

(d)

(e)

(f)

Fig.4. Bar Charts for (a) MAE (b) Accuracy (c) AUROC (d) MCC (e) Mean entropy and (f) F1-Score for different classifiers

Table-IV: Metric values by class for proposed ensemble classifier

Class / Metrics	F1-Score	MCC	AUROC	Precision	Recall	True Positive Rate	False Positive Rate
Normal	0.971	0.96	0.995	1	0.944	0.944	0
Inner Race	0.913	0.872	0.965	0.894	0.933	0.933	0.051
Outer Race	0.9	0.842	0.962	0.896	0.905	0.905	0.061

Fig. 5. Confusion Matrix for the proposed Stacked Ensemble Classifier

V.   CONCLUSION AND FUTURE SCOPE

Statistical parameters extracted from time domain vibration signals can serve as adequate fault sensitive features for vibration signature analysis. These features can be used with machine learning classifiers for efficient diagnosis of bearing faults. The aim of this paper is to explore the application of stacked ensemble classifiers in vibration signature analysis of rolling element bearings. A stacked ensemble of five classifiers is proposed for accurate fault diagnosis and results are compared with conventional ensemble classifiers to prove its effectiveness and robustness.

In this study two histogram parameters and four statistical moments are used in the feature vector. The features extracted from training dataset are used to train a stacked ensemble of K-nearest neighbour, Logistic regression, PART and Random Forest classifiers along with Multilayer Perceptron as learning classifier. The performance of proposed ensemble classifier using stacking is compared with regular ensemble classifiers on the basis of seven evaluation metrics. It is observed that proposed ensemble classifier outweighs the other classifiers, producing an MCC value of 0.927 and an F1-score of 0.923 which are much better than those of other classifiers.

This work can be extended in future in two areas. First, new fault sensitive features can be added to these six features and their comparative effectiveness can be studied to obtain a more optimal subset of features by dimensionality reduction. Second, other state-of-the-art stacked ensemble classifiers can be developed to increase the classification accuracy for bearing faults. The optimization of results in these two areas will help in developing a more reliable and accurate bearing fault diagnosis scheme.

REFERENCES

1. H. Yang, J. Mathew, and L. Ma, "Vibration Feature Extraction Techniques for Fault Diagnosis of Rotating Machinery -A Literature Survey," Asia Pacific Vib. Conf. Aust., no. November, pp. 1–7, 2003.
2. S. N. Engin, K. Gulez, and M. N. M. Badi, "Advanced signal processing techniques for fault diagnostics - a review," Mathematical and Computational Applications, vol. 4, no. 1. pp. 121–136, 1999.
3. M. Delgado et al., "Bearing Faults Detection by a Novel Condition Monitoring Scheme based on Statistical-Time Features and Neural Networks," IEEE, no. c, 2011.
4. B. R. Nayana and P. Geethanjali, "Identification of Bearing Faults Using Statistical Time Domain Features and Fused Time-Domain Descriptor Features," 2018.
5. P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods," Expert Syst. Appl., vol. 38, no. 3, pp. 1876–1886, 2011.
6. R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery : A review," Mech. Syst. Signal Process., vol. 108, pp. 33–47, Aug. 2018.
7. S. S. Zhang, S. S. Zhang, B. Wang, and T. G. Habetler, "Machine Learning and Deep Learning Algorithms for Bearing Fault Diagnostics - A Review," Proc. 2019 IEEE 12th Int. Symp. Diagnostics Electr. Mach. Power Electron. Drives, SDEMPED 2019, pp. 257–263, 2019.
8. R. Bal and S. Sharma, "Review on Meta Classification Algorithms using WEKA," Int. J. Comput. Trends Technol., vol. 35, no. 1, pp. 38–47, 2016.
9. S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," Artif. Intell. Rev., vol. 26, no. 3, pp. 159–190, 2006.
10. T. G. Dietterich, "Ensemble methods in machine learning," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell.), vol. 1857 LNCS, pp. 1–15, 2000.
11. D. Dou, J. Jiang, Y. Wang, and Y. Zhang, "A rule-based classifier ensemble for fault diagnosis of rotating machinery," J. Mech. Sci. Technol., vol. 32, no. 6, pp. 2509–2515, 2018.
12. E. Zio, P. Baraldi, and G. Gola, "Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery," Appl. Soft Comput. J., vol. 8, no. 4, pp. 1365–1380, 2008.
13. N. Sikder, K. Bhakta, A. Al Nahid, and M. M. M. Islam, "Fault diagnosis of motor bearing using ensemble learning algorithm with FFT-based preprocessing," 1st Int. Conf. Robot. Electr. Signal Process. Tech. ICREST 2019, no. July, pp. 564–569, 2019.
14. A. Sharma, M. Amarnath, and P. K. Kankar, "Novel ensemble techniques for classification of rolling element bearing faults," J. Brazilian Soc. Mech. Sci. Eng., vol. 39, no. 3, pp. 709–724, 2017.
15. G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," Sensors (Switzerland), vol. 19, no. 5, 2019.
16. P. Karimi and H. Jazayeri-Rad, "Comparing the Fault Diagnosis Performances of Single Neural Networks and Two Ensemble Neural Networks Based on the Boosting Methods," J. Autom. Control, vol. 2, no. 1, pp. 21–32, 2014.
17. N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc., vol. 454, no. 1971, pp. 903–995, Mar. 1998.
18. K. F. Tom, "A Primer on Vibrational Ball Bearing Feature Generation for Prognostics and Diagnostics Algorithms," Army Res. Lab, no. March, 2015.
19. L. I. Kuncheva and E. Alpaydin, Combining Pattern Classifiers: Methods and Algorithms, vol. 18, no. 3. 2007.
20. L. Breiman, "Bagging Predictors," Mach. Learn. 24, 123-140, vol. 24, no. 421, pp. 123–140, 1996.
21. L. Breiman, "Random forests," Mach. Learn. 45(1)5-32., pp. 1–122, 2001.
22. Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," Proc. 13th Int. Conf. Mach. Learn., pp. 148–156, 1996.

23. J. Kittler, M. Hater, and R. P. W. Duin, "Combining classifiers," Proc. - Int. Conf. Pattern Recognit., vol. 2, no. 3, pp. 897–901, 1996.
24. "Fault Data Sets - Society for Machinery Failure Prevention Technology." [Online]. Available: https://mfpt.org/fault-data-sets/. [Accessed: 15-June-2019].
25. "Nice Ball Bearings." [Online]. Available: http://rbcbearings.com/ballbearings/index.htm. [Accessed: 17-Oct-2019].
26. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

AUTHORS PROFILE

Chhaya Grover is a Ph.D. research scholar in Department of Electronics Engineering in J. C. Bose University of Science and Technology, Faridabad, Haryana, India. She holds a Bachelor of Technology degree in Electronics and Communication from Institute of Engineering and Technology, Lucknow, India and Master of Technology degree in Electronics and Communication from National Institute of Technology, Kurukshetra, India. She has more than twenty years of teaching experience and currently working in the Department of Electronics Engineering in JSS Academy of Technical Education, Noida, India. Her research area is vibration signature analysis using advance signal processing for condition monitoring and fault diagnosis.

Dr. Neelam Turk received B.E degree from North Maharashtra University Jalgoan, India, M.Tech. in Electronics and Communication Engineering from National Institute of Technology, Kurukshetra, India and Ph.D. degree in Electrical Engineering from National Institute of Technology, Kurukshetra, India. Dr. Neelam Turk has authored and co-authored multiple peer-reviewed scientific papers and presented her work at many national and International conferences. She has more than 20 years of teaching experience. Currently she is Chairperson & Associate Professor in the Department of Electronics Engineering in J. C. Bose University of Science and Technology, Faridabad, Haryana, India. Dr. Neelam Turk's research interests include Signal Processing, Speech processing, Wireless Communication and Communication Systems.