

Identification of bearing fault in induction motor using random forest algorithm

KrishnaVeni KaliappanSivakumar^{1*}, *Gowsalya* Ganesan², *Gnana Soundari* Chermakani², *Darsani*Muthukumar²

¹ Assistant Professor (Sr. Gr.), Department of Electrical and Electronics Engineering, Mepco Schlenk Engineering College, Sivakasi, India.

² UG Student, Department of Electrical and Electronics Engineering, Mepco Schlenk Engineering College, Sivakasi, India.

Abstract. In day-to-day life 90% of industries use induction motors due to less maintenance, high efficiency, good Power factor and low cost. Maintenance of the induction motor is important for continuous operation in industries. 40-60% of the fault in Induction motors is due to bearing failure. Unexpected bearing failures could cause industries to spend money on repairing and replacing the bearing, along with that other nearby components might be damaged. Failure in bearing, decrease the plant's operating efficiency, increases downtime, raises operating costs and in the worst case, it may cause injuries to workers. The proposed method detects and diagnoses the bearing fault using vibration signals. The fault gets detected by using the Machine learning classifier. The proposed method achieves high accuracy in detecting and diagnosing the bearing fault. The proposed work is implemented using Google Colab (collaboratory) software. The result demonstrates the usefulness of the suggested strategy of enhancing the maintenance of bearing in good condition and safe operation in the induction motor.

1 Introduction

Induction motors play a vital role in industries. Industries like sugar and textile use Induction motor for ginning, blowing and cording process because of its simple construction, robust and rugged, reliability and durability [1]. Not only industries but also E Vehicle also uses Induction motors [2]. There are three kinds of fault that occur in Induction motor such as Mechanical fault, Electrical fault and Environment fault. Mechanical fault occurs in the internal housing of the motor. It is further divided into bearing fault, broken rotor ball fault and rotor mass unbalance fault. Among them the most common fault is bearing fault [3]. Bearings are the essential components of an electric motor that play a critical role in reducing friction and ensuring smoother operation. Bearings are designed to reduce the friction in the motor when two surfaces are in contact. By reducing friction, there will be minimum wear and tear in the motor, so the lifespan of the motor is increased. It allows the motor to run freely and reduces the energy spent to rotate the motor. Using bearings in the

*Corresponding author: krishnaveni.ks@mepcoeng.ac.in

motor increase the overall performance by reducing the vibration and noise. Bearings are cost effective component. Damage in bearing may occur due to dust contamination (47 %), disassembly (14 %)misalignment(13%),insufficient lubrication (11%),overloading (5%),corrosion(5%)and others (5 %) which will directly affect the performance and lifetime of the Induction motor[4,5].Damage in the bearing leads to increased friction,increase temperature,increase wear,decrease efficiency,vibration produced transmitted to the system and damage other components, reduce the performance of the motor including speed, torque and accuracy and may cause catastrophic failure. This can be avoided by monitoring the bearing condition.

2 Literature survey

There are several signals used for identifying the bearing fault. They are vibration signals, acoustic signals, motor current signals, noise signals and temperature signals [6].Vibration and temperature signals are collected from the motor and based on both the signal behavior the fault in bearing is identified, whereas the main drawback in that method is measured. Temperature signal provides inaccurate measurement and gives slow response. [7] The fault in the bearing using vibration, temperature and noise signals are analyzed. In this paper, it is concluded that temperature and noise signals has limited applicability and sensitive to external application. In [8,9] comparison is done for both vibration and motor current signals are used for finding the fault in motor. In this paper, it is concluded that the motor current signal is limited towards diagnostic capability. On the other hand, vibration signals are the best choice for detecting the fault in motors in ceit provid equantitative data and cost effective technique. In the proposed work vibration signals were collected. Various features like statistical and frequency domain features extracted from the collected vibration signals.[10] The performance of statistical features and frequency domain features extracted from vibration signals are compared and concluded that statistical features are much better in identifying the fault in bearing since it requires low computational equation and can provide valuable insights into complex data. In article [11]60featuresare extracted to the train the model, it will increase the computational complexity, increases computation altime and decrease accuracy of the model. In order to sort out this problem, feature selection technique is used. In [12] principle compound analysis (PCA) is used for feature selection. Using PCA needs data normalization and more sensitive to outliers. The article addressed in [13] use Genetic algorithm for feature selection. From 16 statistical parameters, 9 features are selected for training the model. The major drawback in GA is slow convergence and lack of guarantee of finding the global optimum features. The alternate method for feature selection is correlation lot. The advantage of using correlation lot is to selects highly related parameters and is easy to compute. The selected feature is computed in the classifying technique such as Ensemble Empirical Mode, Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), Hilbert transform, Artificial Neural Network (ANN), K – Nearest Neighbor (KNN), Support Vector Machine(SVM),Decision Tree, Random Forest, Optimized Elman Adaboost, Fuzzy Neural Network etc. The bearing fault detection using Complete Ensemble Empirical Mode Decomposition(CEEMD) is proposed in article [14].The CEEMD technique requires some tuning parameters such as ensemble members, stopping criterion and maximum number of iterations. The detection of bearing failure using Discrete Wavelet Transform (DWT) is discussed in [15].The major drawback in using DWT is it only supports stationary signals, discretization error occurs and small shift in input sample create huge change in output coefficient. In article [16] FFT is used for detecting fault. The drawback of using FFTis, it only supports non- periodic signals and is sensitive to outliers and noise. In article [17] described about Hilbert transform for detecting bearing fault. Due to

Non causality, it is difficult to work in real time application. All the above methods such as DWT, FFT and Hilbert transform all are time consuming and need solving of complex mathematical expressions. This drawback will be overcome by using Machine Learning Techniques. The advantage of using Machine Learning Techniques are, if deals with huge amount of data, took less time to give results and fault can be identified at very early stage. In [18] auto encoder method is used for detecting the bearing fault. Auto encoder comes under unsupervised learning so there will be no labelled data this makes quality less predictions. CNN is the best fit for image processing and not suitable for signal analyzing [19]. Elamada boost is used in [20] for detecting the fault. It is a time-consuming process and cannot deal with large dataset and over fitting occurs. Various methods such as KNN, naïve bayes, SVM and ANN are compared and following observation are made in article [21]. Using KNN is not efficient since it is the weakest algorithm, sensitive to outliers and accuracy is dependent on choice of k . In naïve bayes it involves solving of more equation and time consuming algorithm. The drawback of using SVM is it work well if there a clear margin between two classes and the algorithm need to solve the quadratic optimization problem to find optimal hyper plane that separate two classes. ANN involves intensive computation and time consuming process. It also requires tuning parameters such as learning rate, number of nodes and nodes per layer. The work addressed in [22] suggested that Random Forest has best accuracy than ANN. Since random forest has high accuracy, low over fitting, less sensitive towards outliers and no assumption is required [23,24]. For collecting vibration signal from the induction motor various sensor have been used among that Micro electro mechanical system (MEMS) accelerometer is the most suitable one [25] since it is low power consumption, consumes low power and used in variety of applications.

3 Work adopted

The paper targets to classify bearing faults using supervised machine learning algorithm. Pitting and bruising in the bearing inner race, outer race and rolling element happen due to over loading, improper lubrication and misalignment. For monitoring, vibration signals from the induction motor get collected under different operating condition with various bearing sets such as healthy bearing, ball defect bearing, inner race fault in the bearing and outer race fault in the bearing. The vibration signals are collected using an accelerometer. In addition to the real time samples, data are also taken from Case Western Reserve University (CWRU) database [26,27]. The vibration signal appears to be similar in all four conditions except, variation in vibration magnitude. To obtain necessary information, statistical features are extracted from vibration signals. The process of reducing the dimension and transforming raw samples into numerical features is known as feature extraction. 9 statistical features are extracted from the collected sample signals in the proposed system. They are minimum, maximum, mean, standard deviation (SD), root meansquare (RMS), skewness, kurtosis, crest factor and form factor. Training and testing the model with 9 features may increase the computing time and also give rise to over fitting. So, the inappropriate or irrelevant feature gets eliminated using feature selection technique. Correlation plot is used for selecting feature which are more relevant and results in rejection of irrelevant data from the extracted ones. Finally, the selected features from the correlation plot is used to train the Random forest algorithm to perform the fault classification task. There are totally 3360 samples. Among them 840 sample are considered for each case. The entire sample is divided into 65:35 portions for training and testing the model i.e among 3360 samples, 2184 samples are used for training and 1176 samples are used for testing. To ensure the consistency in training the model, the google colab code is run for several times.

Once the code has been successfully compiled, the model is tested using evaluation metrics to determine its performance. These metrics include accuracy, precision, recall and F1 score. Once the raw sample is given as input to the model, intermediate process such feature extraction and feature selection are performed and the model predict status of the new sample and give the status of the bearing as output. Mail will be sent to the user using the Simple Mail Transfer Protocol (SMTP) once the prediction is in fault condition. The block diagram of the proposed system is shown in Fig.1. The above process is implemented in online web based notebook environment named Google colab notebook

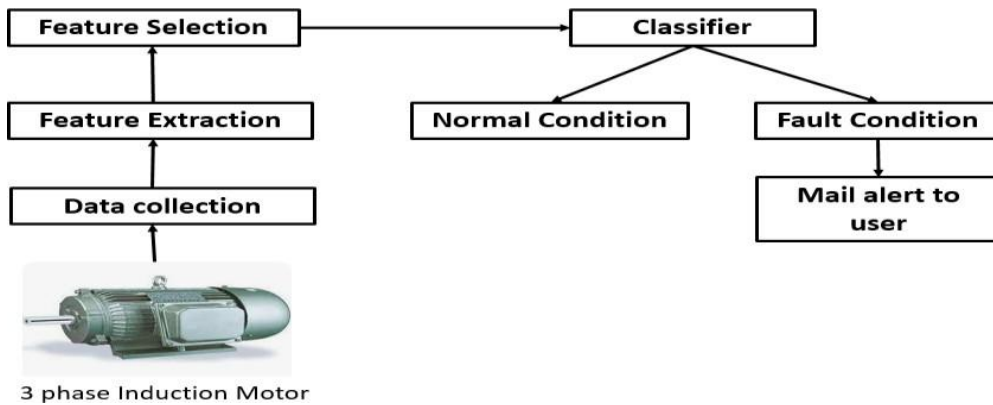


Fig.1. Block diagram for the proposed work

4 Real time data acquisition

The experimental setup for collecting the vibration signals from the 3-phase induction motor is shown in Fig. 3. And the motor specification is tabulated in Table 1. Vibration signals get collected using ADXL335 accelerometer. It is easy to use, more accurate and less cost. Accelerometer connected with Arduino UNO for analog interface. Where the accelero meter (sensitivity 270mV/g to 330mV/g) is mounted on the axial position of the Induction motor to collect vibration signals. In the experimental set up bearing which has 8 balls in the bearing is used. The vibration signals from the motor are measured by using vibration sensor interfaced with Arduino UNO board. Using the cool term software, the data from the Arduino were read into the excel sheet. Data are collected for healthy bearing, inner race damaged bearing, out errace damaged bearing and rolling element damaged bearing. The bearing used to collect the vibration signals are shown in Fig. 2. The damage in the bearing is created manually.

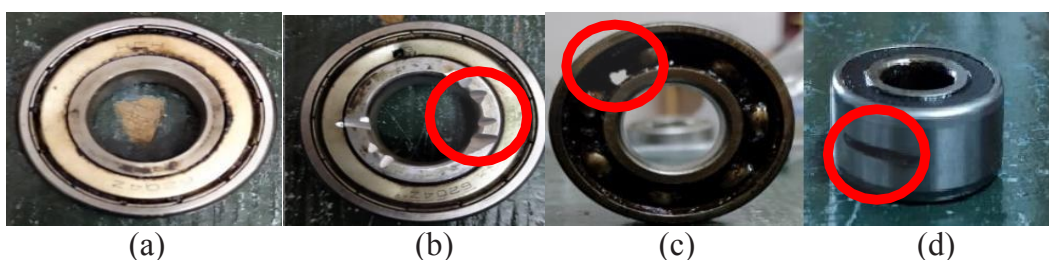


Fig. 2. Ball bearings. (a) healthy bearing (b) inner fault bearing, (c) ball fault bearing (d) outer bearing

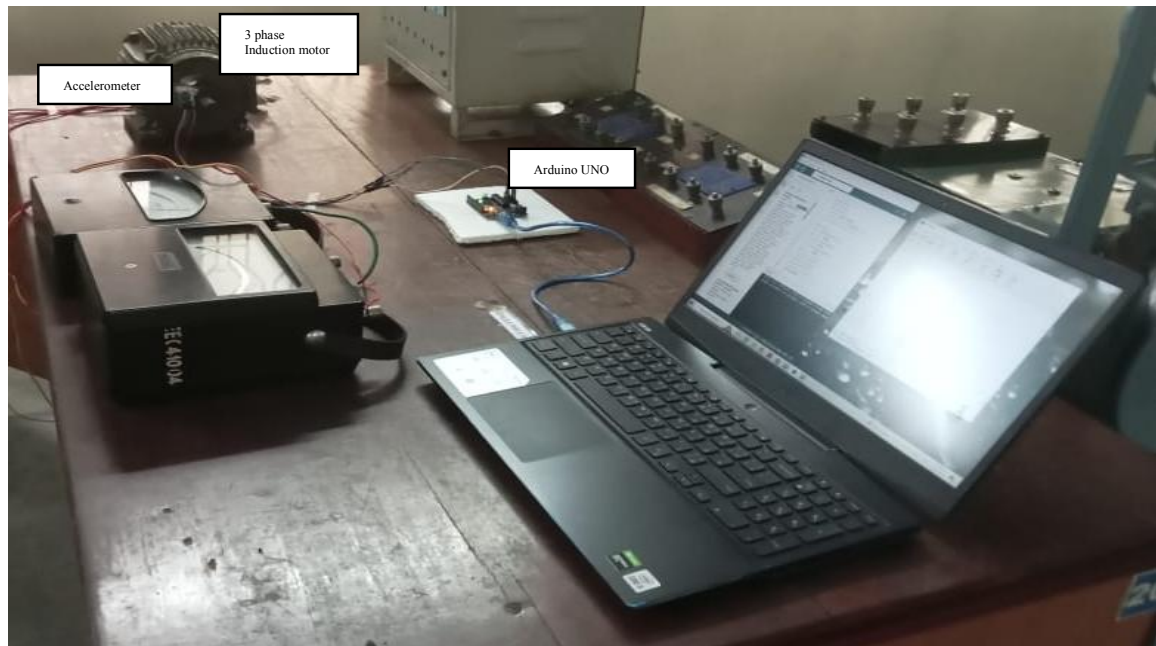


Fig.3.Experimental set up of the proposed work

Table 1.MotorSpecification

Parameter	Range
MotorRating	1HP
Voltage	415V
Current	1.5A
Connection	Delta-star
Ratedrpm	2880 rpm

5 Implementation of the proposed work

5.1 Feature Extraction

From the raw signals, nine standard statistical features are calculated. Compared to time domain features and frequency domain features, statistical features are easy to calculate and efficient in fault diagnosis. Time domain feature has low precision and frequency domain feature need of transfer function and hard to extract the frequency domain parameters from steady state response. The statistical features are given in the equation (1)to(7).

$$Mean = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$\text{Standard deviation} = \sqrt{\sum_{i=1}^n \frac{(x_i - \mu)^2}{N-1}} \quad (2)$$

$$\text{RMS} = \sqrt{\sum_{i=0}^n \frac{x^2}{n}} \quad (3)$$

$$\text{Skewness} = \frac{1}{n} \sum_{i=0}^n \frac{(x_i - \bar{x})^3}{(\text{standard deviation})^3} \quad (4)$$

$$\text{Kurtosis} = \sum_{i=0}^n \frac{(x_i - \bar{x})^4}{(\text{standard deviation})^4} \quad (5)$$

$$\text{Crest factor} = \frac{\text{RMS}_{\text{maximum}}}{\text{RMS}} \quad (6)$$

$$\text{Form factor} = \frac{\text{RMS}}{\text{mean}} \quad (7)$$

5.2 Feature selection

The extracted nine statistical features used to train the model, accuracy decreases and time of computation will get increases. To improve the accuracy, the classifier trained with highly related features. So, feature selection is included in the proposed work. Because, all the extracted features may have different characteristics. The features which have same characteristics are chosen for classifying fault. Correlation plot is emerging technique used for features election. It is easy to implement and shows highly correlated parameter. A correlation lot is a type of visualization that gives the relationship between variables in the dataset. It is a matrix with heat map or scatter lot that display correlation between all pairs of variables in the dataset. In this project heat map is used. In a heat map, a color gradient is used to represent the magnitude of the correlation coefficient. The value of coefficient ranges between -1 and 1. A perfect positive correlation is represented by 1 and zero means there is no correlation among the variables. The Pearson's correlation coefficient (r) is computed between the two continuous variables. Using the equation (8),

$$r = \frac{(n \sum_{i=1}^n xy - \sum_{i=1}^n x \sum_{i=1}^n y)}{\sqrt{[\sum_{i=1}^n x^2 - (\sum_{i=1}^n x)^2 / n] [\sum_{i=1}^n y^2 - (\sum_{i=1}^n y)^2 / n]}} \quad (8)$$

where,

n is a number of observations

$\sum_{i=1}^n xy$ is the sum of product of each pair of corresponding observations

$\sum_{i=1}^n x$ is the sum of X observations

$\sum_{i=1}^n y$ is the sum of Y observations

$\sum_{i=1}^n x^2$ is the sum of squared X observations

$\sum_{i=1}^n y^2$ is the sum of squared Y observations

From the Fig. 4. it is inferred that RMS & SD and kurtosis & crest factor are highly correlated. Since the above stated features alone have very high correlation coefficient, plot features such as SD, RMS, kurtosis and crest factor are selected. All other parameters were eliminated from the collected data sample.

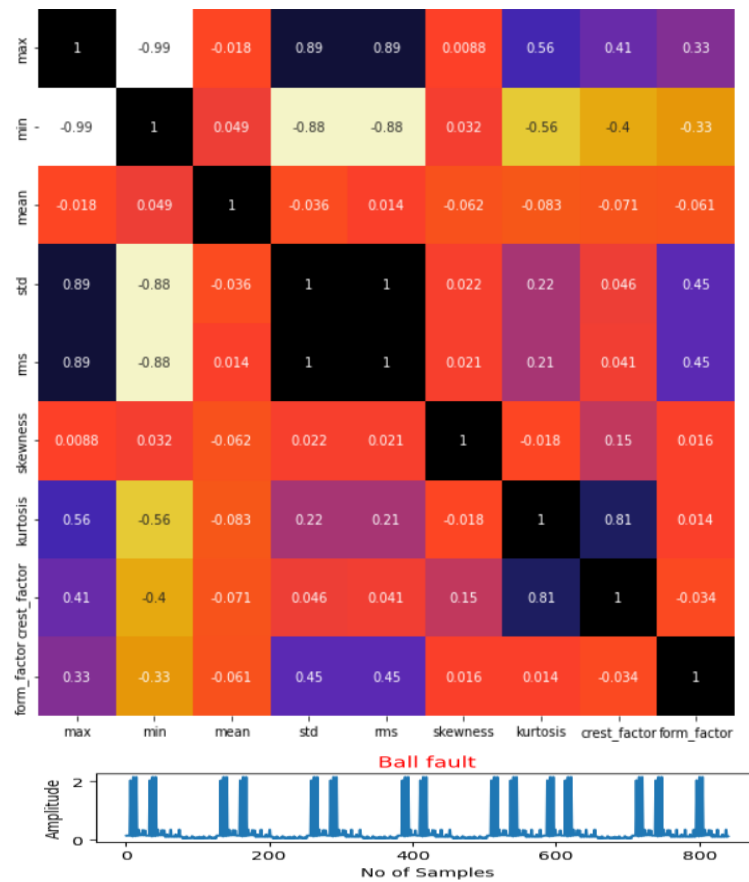


Fig.4.CorrelationPlot

Waveforms under 3 faulty conditions are shown below for the selected features from the correlation lot. It is observed that 3 faulty conditions have different wave form pattern and has variation in amplitude.

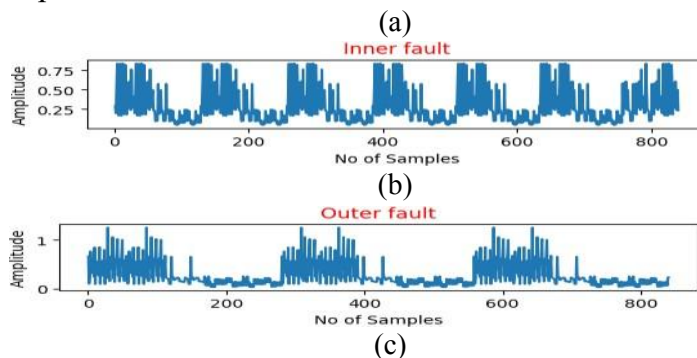
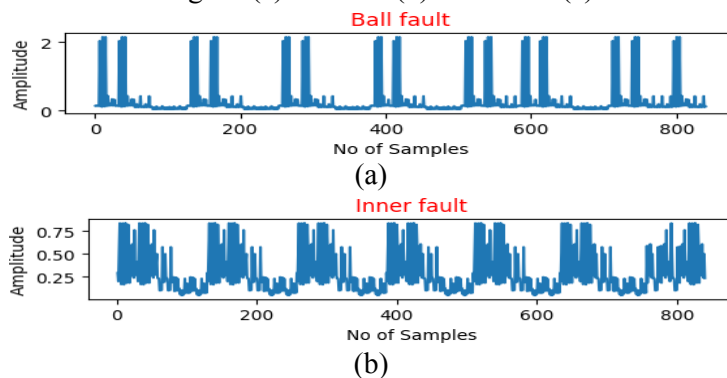


Fig.5. SD wave form of vibration signals(a)Ball fault(b)Inner fault (c)Outer fault



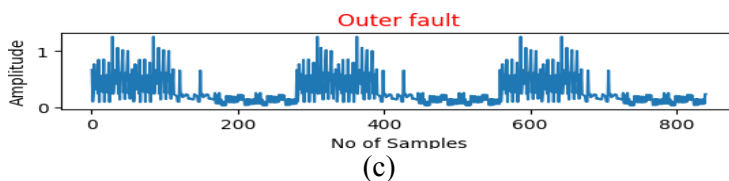


Fig.6.RMS wave form of vibration signals(a)Ball fault(b)Inner fault(c) Outer fault

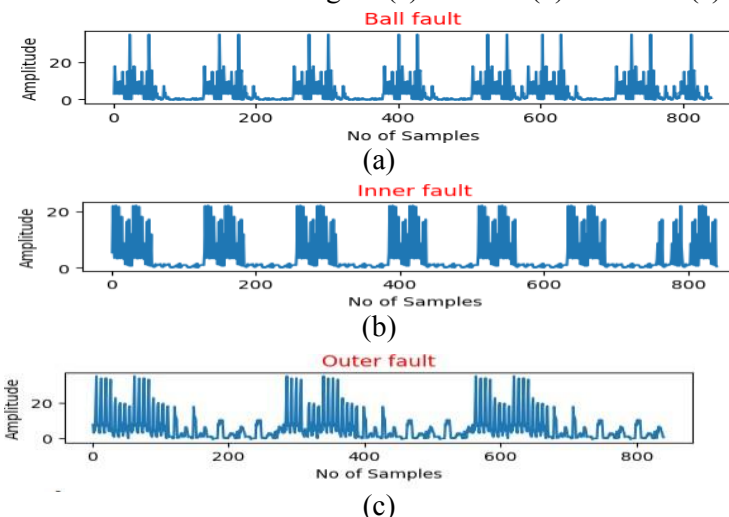


Fig.7.Kurtosis wave form of vibration signals(a)Ball fault(b)Inner fault(c)Outer fault

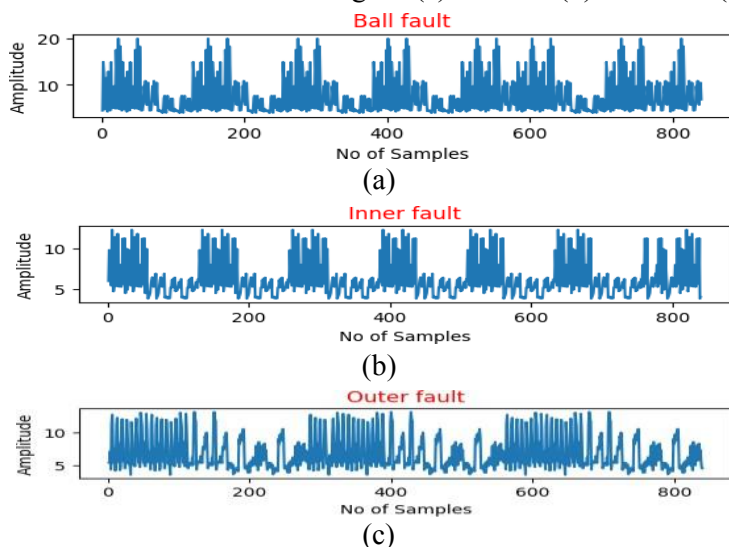


Fig.8.Crest factor wave form of vibration signals(a)Ball fault(b)Inner fault(c)Outer fault

Both SD and RMS have significant changes during any type of bearing fault. For RMS and SD, during Ball fault the amplitude varies from 0 to 2, inner fault the values get reduced, and the amplitude varies from 0.25 to 0.075 and outer fault the amplitude varies from 0 to 1 inferred from Fig. 5. And Fig. 6. The feature plot for kurtosis feature remains same for all type of fault. For crest factor feature during ball fault amplitude varies from 0 to 20. Inner and outer fault amplitude varies from 0 to 10 shown in Fig. 8.

5.3 Random forest

After choosing the highly related features from the correlation lot, random forest algorithm is used to build the classification model to detect the fault in bearing. Random forest is the supervised machine learning technique. It is a popular and reliable algorithm used for

regression and classification tasks. Since it is a supervised learning algorithm, it uses labeled data sets for predicting results. Numerous decision trees are combined to predict the output. It is also known as ensemble learning technique.

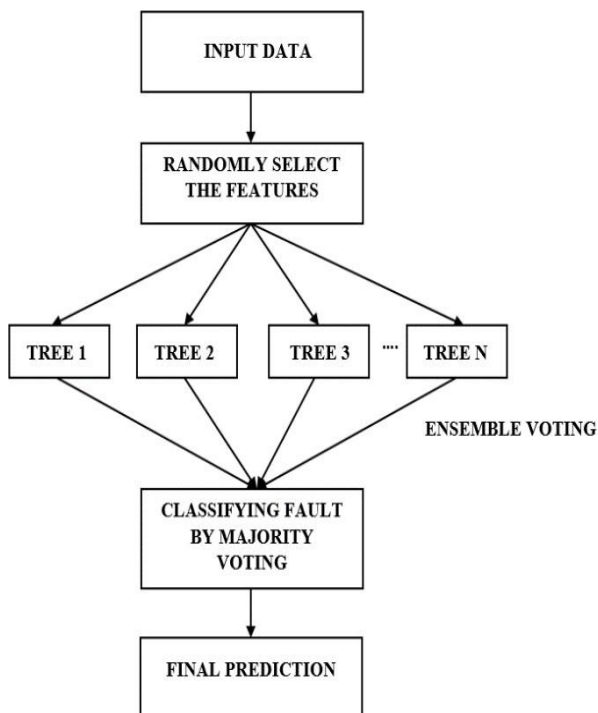


Fig.9.Workflow of Random Forest Technique

The work flow for classifying the fault in the bearing data set follow these steps,

1. Randomly choose the sub set of the data from the input dataset.
2. Choose a portion of the data at random and create a tree.
3. Built the decision tree for the selected data.
4. Repeat the step1to3for a predefined number of times to create multiple decision tree
5. For classification problems, each decision tree independently predicts the output and final prediction is determined by majority voting method.

The work flow model of random forest is shown in Fig.9.The random forest uses multiple decision tree to classify the fault. As, it combines the output of many decision tree, the accuracy gets increased and over fitting get reduced. Random forest is relatively robust to noisy data and outliers, as it can learn the underlying structure of the data by aggregating the information from many trees. Random forest can be easily parallelized to make it well-suited for large datasets. It takes less time to predict the output compared to the other algorithm. It maintains the accuracy of the output even the huge amount of data is missing.

5.4 Performance evaluation of the proposed system

A machine learning model's performance must be assessed in order to figure out whether it

Is generating predictions that are reliable and accurate. There are several other important reasons for model evaluation they are,

1. To verify the model's accuracy
2. To determine a model's weakness
3. To avoid over fitting

By using confusion matrix, accuracy, precision, recall are the model evaluation metrics and it is calculated using equation (9) to (12). By using the confusion matrix terms like True Positive(TP), True Negative(TN), False Positive(FP) and False Negative(FN) are calculated.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

$$F1\ score = 2 * \frac{precision * recall}{precision + recall} \tag{12}$$

Actual labels	Ball Fault	293	0	0	0
	Inner Fault	5	291	0	0
	Normal	0	0	280	0
	Outer Fault	0	0	0	307
		Ball Fault	Inner Fault	Normal	Outer Fault
		Predicted labels			

Fig.10. Confusion matrix of Random Forest

Actual labels	Ball Fault	293	0	0	0
	Inner Fault	5	291	0	0
	Normal	0	0	280	0
	Outer Fault	13	8	0	286
		Ball Fault	Inner Fault	Normal	Outer Fault
		Predicted labels			

Fig.11. Confusion matrix of KNN

For evaluating the performance of the proposed system, it is compared with conventional classification algorithm such as KNN. Model accuracy defines the measure of how well it can categories or predict a task out come. The accuracy obtained from random forest is 99.57% and KNN is 97.78%. The confusion matrix of random forest and KNN for testing samples are shown in Fig.10. and Fig.11. of the proposed system and conventional system. The precision shows the quality of positive predictions made by the model. The random forest has a precision scoreof0.99and KNN has 0.98, so the chances for positive prediction in random forest are highly accurate. Recall measures the correctly identified TP cases and model recall value for random forest is 0.99 and KNN is 0.97. Both precision and recall for the model is high for random forest

Table2.Comparisonbetween randomforestand KNN

Parameters	Randomforest	KNN
Accuracy	99.57%	97.78%
Precision	0.99	0.98
Recall	0.99	0.97
F1Score	0.99	0.98

6 Conclusion

In this project vibration signals of both healthy and faulted bearings are collected from induction motor,statistical features are extracted and correlation lot is used for feature selection. Selected features are used to build the classifier model. Using feature selection techniqueine extracted are features reduced to four features. Therefore, accuracy gets increased and computational time decreases. By detecting the bearing fault at the initial stage, replacing of bearing can be done in the initial. The accuracy obtained from the Random Forest is 99.57 % whereas the accuracy of KNN is 97%Random Forest prediction is more accurate when compared to KNN. Traditional fault diagnosing method sometimes generate false alarm, but the proposed technique reduces the false positive alarm.

7 Reference

1. BekirGecer,OzturkTosun,HasbiApaydin,N.fusunOymanSerteller,IECCME,**6**(2021)
2. AkeshiTakahashi,ShinjiSugimoto,KazuoNishihama,SatoshiSumita,ShunTaniguchi, ICEM, **22**(2021)

3. A.Naha, A.K.Samanta,A.Routray,A.K.Deb, *IEEETrans.Instrum.Meas.*,**65**, **July**(2016).
4. K.S.KrishnaVeni,*Stator IntJInnovateResSciEngTech*,**8**(2019)
5. V.ErickAlviento,B.JefersonCabais,M.JuanMiguelFajardo,G.RamonGarcia,H.Alejandro Ballado, *CSPA*,**17**(2021)
6. ArchitUpadhyay,*International J Of CoreEngi&Management*,IJCEM,**1**(2014)
7. TanerGoktas, *IEEE Sens.J.*, **22**(2022)
8. SulochanaWadhvani,S.P.Gupta,VinodKumar, *IEEE SDEMPED*, **40** (2014)
9. H.PrashantJain,P.SantoshBhosle, *IOP Conf.Ser.:Mater.Sci.Eng.*, **30** (2020)
10. AllanAlvesPinheiro,IagoModestoBrandao,CesardaCosta, *Eur.J.Eng.*,**4**,February(2019)
11. AmandeepSharma,RajvardhanJigyasu,LiniMathew,ShantanuChatterji, *Int.Conf.Electr.Eng. Inform*,**23**(2018)
12. RafiaNishatToma,E. AlexanderProsvirin,Jong-MyonKim, *J.Sens.*,**20**(2020)
13. PauloAntonioDelgadoArredondo,DanielMorinigoSotelo,RoqueAlfredoOsornioRios, Juan Gabriel Avina Cervantes, Horacio Rostro Gonzalez , Renede JesusRomero ElseveirSci.,**83**(2016)
14. XinshengLou,KennethA.Loparo, *MSSP*,**18**(2015)
15. NiloySikder,KangkanBhakta,AbdullahAlNahid,MMManjurulIslam, *ICREST*,**19**(2019)
16. MinaAbd-el-Malek,K.Ahmed AbdelSalam,E.OlaHassan, **93** (2017)
17. XinHuang,Guangrui Wen ,ShuzhiDong, Haoxuan Zhou,ZihaoLei ,ZhifenZhang,XuefengChen, *IEEE TransInstrumMeas*, **70**(2021)
18. DhirajNeupane,YunsuKim, JongwonSeok, *IEEEAccess*,**9**(2021)
19. QiangFu,BoJing,PengjuHe,Shuhao Si,andYunWang,*Fault Feature Selection and IEEE Sens.J.*,**18**(2018)
20. BodiCui,YangWeng,NingZhang, *PERC*,**19** (2022)
21. Raj Kumar Patel,V.K.Giri, *J.Mech.Sci.*,**8**(2016)
22. JitendraKumarJaiswal,RitaSamikannu, *WCCT*,**25**(2019)
23. LanjunWan,KunGong,GenZhang,XinpanYunan,ChangyunLiandXiaojunDeng, *IEEEAccess*,**9**(2021)
24. Dr.S.KaviPriya,Mrs.N.Saranya, *IETE*,**15**(2022)
25. MohamadHazwan,Mohd Ghazali,WanRahiman, *A Systematic Review*,*Schok.Vib.*, **21**(2021)
26. DhirajNeupane,JongwonSeok,*Bearing IEEEAccess.*,**8** (2020)
27. A.WadeSmith,B.RobertRandall, *A benchmark study*,*MSSP*,**64**(2015)