 SDHM, vol.12, no.1, pp.1-16, 2017

Condition Monitoring of Roller Bearing by K-Star Classifier and *K*-Nearest Neighborhood Classifier Using Sound Signal.

Rahul Kumar Sharma[*1], V. Sugumaran[1], Hemantha Kumar[2], Amarnath M[3]

Abstract Most of the machineries in small or large scale industry have rotating element supported by bearings for rigid support and accurate movement. For proper functioning of machinery, condition monitoring of the bearing is very important. In present study sound signal is used to continuously monitor bearing health as sound signals of rotating machineries carry dynamic information of components. There are numerous studies in literature that are reporting superiority of vibration signal of bearing fault diagnosis. However, there are very few studies done using sound signal. The cost associated with condition monitoring using sound signal (Microphone) is less than the cost of transducer used to acquire vibration signal (Accelerometer). This paper employs sound signal for condition monitoring of roller bearing by K-star classifier and k-nearest neighborhood classifier. The statistical feature extraction is performed from acquired sound signals. Then two layer feature selection is done using J48 decision tree algorithm and random tree algorithm. These selected features were classified using K-star classifier and k-nearest neighborhood classifier and parametric optimization is performed to achieve the maximum classification accuracy. The classification results for both K-star classifier and k-nearest neighborhood classifier for condition monitoring of roller bearing using sound signals were compared.

Keywords: K-star, *k*-nearest neighborhood; *k*-NN, machine learning approach, condition monitoring, fault diagnosis, roller bearing, decision tree algorithm; J48, random tree algorithm, decision making, two layer feature selection, sound signal, statistical features.

1 Introduction

Most of the Industrial machinery has a rotating element supported by bearings for continuous and proper application. Continuous lubrication is done for its proper functioning; however the continuous load on the parts and interference of foreign material lead to the failure of bearing and hence the system. These failures generally lead to high

---

[1] V.I.T. university, Chennai, India.

[2] N. I. T. Karnataka, Srinivasanagar, Surathkal, Mangalore, Karnataka, India.

[3] I. I. I. T. D. & M. Jabalpur, Khamaria, Jabalpur, India.

* Corresponding author

Department of Mechanical Engineering, School of Mechanical and Building Sciences, VIT University, Vandalur – Kelambakkam Road, Chennai - 600127, Tamil Nadu, India.

E-mail: rahulkumar.sharma2011@vitalum.ac.in

losses of capital and time. The reason for a typical failure is fatigue crack initiation and its propagation due to cyclic contact stressing. It is difficult to estimate the severity level of localized faults when the bearing is operating under variable loads and speeds. Most common failure mechanisms initiates from mechanical wear, lack of lubrication and corrosion of elements. Localized faults in rolling bearing causes impact which results into transient excitations which can be observed in vibration and sound signals. Analyzing these signals becomes difficult in faulty operating conditions as these signals have non-stationary characteristics. Hence, it is highly recommended to continuously monitor the health of bearing to avoid any sudden failure.

Condition monitoring is a process of continuously monitoring the health of a system or subsystem. This method has several steps after the data is acquired from test rig. The acquired signals can be analyzed using trend analysis, octave analysis, FFT, wavelet analysis and machine learning techniques. Of late, machine learning techniques are widely used for analysis due to its learning capabilities. It is performed by feature extraction followed by feature selection and classification of extracted features. Different features like wavelet features, histogram features and statistical features are used for condition monitoring. Acquired sound signal data set for different bearing condition is used to extract a set of statistical features. The methods used for feature selection process are decision tree, genetic algorithm etc. In a study, Sugumaran (2003), employed J-48 decision tree algorithm for feature selection process. The importance of features were decided on the basis of their appearance in the tree. Feature at the top of tree was considered most prominent and selected as first feature for classification. Further selection of features were done using the same methodology. Classifier which can be operated in simplest way and classify features with high accuracy is optimal for feature selection process. J-48 decision tree algorithm satisfies all these conditions. In a study, Ye tian (2015) performed fault diagnosis of roller bearing using combination of LMD (local mean decomposition) and SVD (singular value decomposition) and ELM (extreme learning machine). LMD-SVD was applied for feature extraction and ELM was employed for improving accuracy of diagnosis and reducing human intervention. Zhiqi (2015) had employed vibration sensor for fault diagnosis of planetary bearings. Accelerometer was used for data acquisition and detection algorithm is developed using advanced digital signal processing techniques to detect faults in both inner and outer race.

Standard classifiers like decision tree, K-star, SVM, PSVM, Naïve Bayes (NB) and Bayes Net (BN), etc. were used widely for classifying features. In a study, Wang (2007) employed signals in frequency domain for condition monitoring of a centrifugal pump. For classification of features neural network and fuzzy classifiers were used. In a study, Sugumaran (2007b) performed fault diagnosis of roller bearing using J-48 decision tree classifier. From extracted features, J-48 decision tree algorithm was employed to select the best features in order of their importance. Yuan (2006, 2007) employed support vector machine for condition monitoring. However, pattern size, classifier training time and computational complexity will further increase for PSVM. Jin Yi (2012) performed condition monitoring of vehicle hydraulic brake system using virtual instrumentation technology for the online examination of brake system. In a study M. Elangovan (2010) performed condition monitoring of various components using machine learning approach using classifiers like Bayes net, Naïve Bayes and support vector machine. Feature extraction was done using feature subset method and extracted features were classified using Naïve Bayes classifier. Xiaoming Xue (2015) had proposed fast ensemble empirical mode decomposition for fault diagnosis of roller bearing. Adaptive parameter selection is used and effectiveness of proposed method is demonstrated. Method is demonstrated to reduce the computational cost significantly. In a study Xiaovue Chen (2014) had performed fault diagnosis of roller element bearing using dependent feature vector and probability neural network. A dependent feature vector was used to denote the fault symptom attribute of roller bearing which achieves an effective accuracy.

In a study, Ananthapadmanabhan (1983) performed a study to demonstrate the effect of surface irregularities of rolling and sliding contacts in noise generation. It is clear from the result that increasing the roughness will increase the overall noise of the system. Level of noise increases with increase in wear of fine surface; however, Level of noise decreases with increase in wear of rough surface. Heng (1997) employed vibration and acoustic signals for fault diagnosis of rolling elements. Statistical parameter estimation method was employed to obtain parameters such as crest factor, kurtosis, skewness etc. Bayder (2003) performed a study to demonstrate the suitability of vibration and sound signal in fault diagnosis process of two stage gearbox. After comparing the result of analysis done on both vibration and sound signal, it is found that sound signals were more efficient for early stage fault diagnosis in rotating machine elements. Shibata (2000) employed summarized dot pattern method for condition monitoring of fan bearing using sound signal. Results were represented in a diagrammatic manner which makes it further easier for maintenance person to identify the faulty bearing. Wang (2011) employed frequency domain vibration signal for condition monitoring of bearing using possibility theory and fuzzy neural network. Diego (2013) employed an automatic method of condition monitoring of roller bearing by signal processing and pattern recognition techniques. For early stage fault diagnosis, a combination of envelope analysis, MSVM and rule-based expert system was employed.

Amarnath (2013) employed sound signal for condition monitoring of roller bearing. Feature selection and classification is done using J-48 decision tree algorithm. Classification accuracy of 95.5% was reported with 10-fold cross validation [21]. R. K. Sharma (2015) had employed Bayes net and Naïve Bayes for roller bearing fault diagnosis using sound signal. J48 decision tree algorithm was used for feature selection. Classification accuracy of Bayes net was reported to be higher than Naïve Bayes. Sound and vibration signal acquired from rotating machineries often map the features of fault related signals. Condition monitoring of the roller bearing has gain high importance in the recent years. There are enormous studies done on condition monitoring of roller bearing using vibration signal; however the high cost of transducers used for acquiring vibration signals (accelerometer) restrict its usage from small and medium scale industries. The transducer required for acquiring sound signals data (microphone) is lesser in cost relatively. Hence, there is a need to conduct a study that demonstrates suitability importance of sound signal for condition monitoring of roller bearing.

Machine learning is a widely used approach for condition monitoring. In terms of feature selection, decision tree algorithm is simplest and efficient for various types of data sets in a supervised learning. Lazy is a set of algorithms with variants like *k*-NN (*k*-nearest neighborhood), K-star and LWL (Locally Weighted Learning). K-star is a classifier works on instance based learning technique, it differs from other instance-based learners in that it uses an entropy-based distance function. However, *k*-NN is only approximated locally and all computation is deferred until classification. It can be useful to weight the contributions of the neighborhood, so that the nearer neighbors contribute more to the average than the more distant ones. These algorithms are efficient and among the simplest in machine learning. However, no known research is done by using K-star classifier and *k*-nearest neighborhood classifier for condition monitoring of roller bearing using sound signal. Hence, there is a need to perform a study of fault diagnosis of bearing by K-star classifier and *k*-nearest neighborhood classifier. In present study, an attempt is made to effectively classify different faults occurring in roller bearing with K-star classifier and *k*-nearest neighborhood classifier using sound signals. In most of the studies only classification accuracy is compared and lesser importance is given to the false positives; however, false positives can also cause severe failures by reporting faulty conditions as healthy. In present study, false positive is also considered along with classification accuracy for comparison to arrive at the final meaningful result.

The contributions of the present study are:

- After extraction of statistical features from acquired data sets, feature selection was performed and prominent features were selected by two layer feature selection process using decision tree algorithm and random tree algorithm. Features selected by decision tree algorithm and random tree algorithm had shown better classification accuracy when provided to classifiers in an increasing order of their importance.
- The condition monitoring of the roller bearing using sound signal is studied with K-star classifier and *k*-nearest neighborhood classifier. Both classifiers belong to lazy family; however, they vary in their ability and have different advantages. This is comparative study of classifier performance after optimizing their parameters. Classification accuracy of K-star classifier is found higher than *k*-nearest neighborhood classifier with even lesser false positives.

2 Experimental setup

In present study, bearing of the motor pump (SKF R7 NB 62) was studied. Speed of the motor was kept at 1200 rpm. Major components of the roller bearing are rolling elements which rotate on inner race and outer race way. The test rig was prepared and bearing was fixed. The block diagram of experimental setup is shown in Fig. 1.

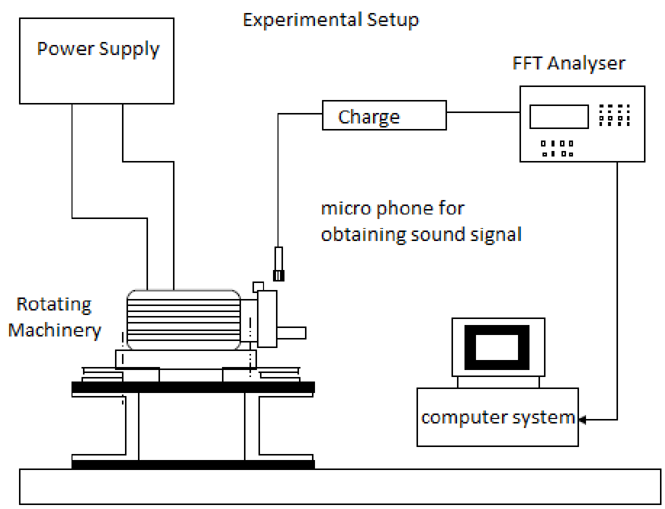

Figure 1: Experimental setup

Four below mentioned bearing conditions were studied and sound signal from the test setup was acquired using data acquisition system. The studied faults are as follows:

- Good condition bearing.
- Outer race fault bearing.
- Inner race fault bearing and
- Inner and outer race fault bearing.

EDM (Electric Discharge Machining) was employed to introduce pits in the inner and outer races of the roller bearing. The cylindrical pits created in both races had depth and diameter of 0.7 mm approximately. Sound signal from the test setup was acquired using a data acquisition system connected to a microphone mounted on the test setup. The good condition bearing was replaced by a faulty bearing and sound signals were acquired. The process was repeated for rest of two cases under same operating conditions. As stated earlier, there are four bearing conditions and these conditions are classes in which selected attributes should classify. The bearing that was used in the study had the following specification:

- Ball diameter = 4 mm.
- No. of rolling elements = 7
- Contact angle = 0 Deg.
- Average diameter = 14 mm.
- Inner ring speed = 0 rpm.
- Speed of outer ring = 1200 rpm.
- Frequency of outer race fault bearing= 84 Hz.
- Frequency of inner race fault bearing=149 Hz.
- Frequency of good condition bearing = 33.3Hz.

3 Feature extraction and feature selection process

Feature extraction is a process of extracting set of computational parameters from an acquired signal which represents characteristics of particular signal. In present study, a set of statistical parameters computed were *mean*, *mode*, *minimum*, *median*, *maximum*, *standard error*, *sample variance*, *kurtosis*, *standard deviation*, *skewness* and *count*. Feature selection was performed after feature extraction process. Feature selection process involves selection of features that are highly efficient and able to classify representing conditions in case of condition monitoring. In present study, two layer feature selection is performed using using decision tree and random tree. Decision tree represents knowledge with a tree structure comprised with a set of branches and nodes. Feature vectors were classified using decision tree. It is performed by moving from root to the leaf thorough nodes and branches. Each node consists of an attribute which provide information for classification. The tree structure ends at a leaf which represents the member of class. Random tree performs both classification and regression on the data. The input to classifier is feature which is classified with every tree in the forest and class label with majority will be output. Regression is performed by averaging all the responses over the trees in the forest. Decision tree algorithm is used as a first layer and random tree as second layer for selecting the richest features among all extracted features. The decision tree algorithm was applied to the statistical features and resultant tree is shown in Fig. 2. The random tree algorithm was applied to the statistical features and resultant tree is shown in Fig. 3.

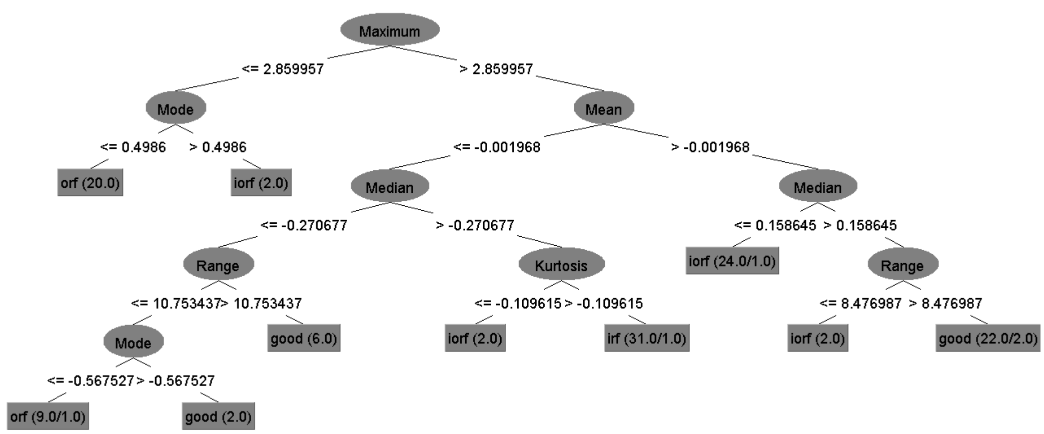

Figure 2: Decision tree using J-48 algorithm

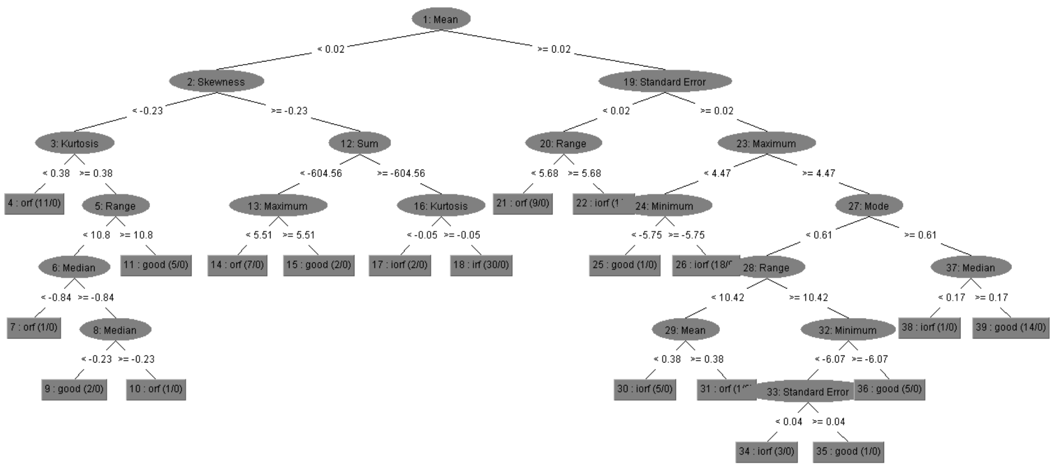

Figure 3: Decision tree using random tree algorithm

The feature selection method is as follows:

- Decision tree algorithm was used to select features from set of extracted features. The tree obtained is shown in Fig. 2. Only 6 features appeared in decision tree and were selected for feature classification. The selected features were *maximum, mean, median, mode, range* and *kurtosis*.
- The feature which appears on top of tree is richest feature and hence selected as best. Here, *maximum* was on the top of tree, hence, it was selected as the first feature for classification.
- Further feature selection was done by ranking the feature appearing on same level of branches on the basis of further branching size. *Mean* was selected as the 2nd feature for classification as *mean* had more branches than *mode*. Using same procedure 4th, 5th and 6th feature were found to be median, range and kurtosis.
- Following the above process with random tree algorithm, another tree was obtained (shown in Fig. 2). Features were selected using same process but the features already appeared in J- 48 decision tree were omitted. Rest of 6 features which were *standard error, skewness, sum* and *minimum* were ranked 7th, 8th, 9th and 10th respectively.

Even if decision tree has most prominent features, more than half of features are missing which may be important for other classifiers. Hence, same method was repeated with random tree algorithm and top most features are selected leaving those which already appeared in the decision tree algorithm. These features are placed after the features selected using decision tree algorithm. Hence, after merging the features from both the classifiers the selected features in the order of precedence were *maximum, mean, median, mode, range, kurtosis, standard error, skewness, sum* and *minimum*. Rest of the three features *count, standard deviation* and *sample variance* were ranked randomly.

4 K-star classifier

K-star is a classifier belongs to lazy family which performs good with both instance based and rule based learning. It solves the smoothness problem by taking the summation of all probability values in possible path. This technique is a great contribution to its performance. It allows integration of real and symbolic valued features and a method to deal with missing values. This instance based classifier uses employs entropic distance measure. For every other dimension a value should be chosen for parameters *x0* and *s* for real and symbolic features respectively. Suppose the value of symbolic feature *s* is approximately 1, data points with symbols other than one will have very lesser transformation probability; however, the data points with same symbol will show a higher transformation probability. Hence, the distance function will show prominent performance to nearest neighbors. When the value of symbolic feature *s* will approach zero, probability distribution of symbols will be reflect by transformation probability. This leads to frequent occurrence of favoring symbols. Effective number of data points can be computed for P* function by using the following expression:

$$n_0 \leq \frac{\left(\sum_b P^*(b|a)\right)^2}{\sum_b P^*(b|a)^2} \leq N \tag{1}$$

Probability of data point will be calculated by being in class *C* by summing the probabilities from 'a' to each data point that is a member of class *C*.

$$P^*(C|a) = \sum_{b \in C} P^*(b|a) \tag{2}$$

The probabilities for every class is calculated. Category distribution of the data point space 'a' at a given point can be estimated by relative probability. Typical techniques return result of classification as a single category; however, for simplifying the comparison maximum probability category is selected for classification of new data point.

5 k-nearest neighborhood classifier

The *k*-nearest neighborhood algorithm (*k*-NN) is a classifier where value of the function is approximated locally and each computation defers until classification is done. It works on instance based learning method. This is a non-parametric method which is used for both regression and classification as well. In either case, input data point is surrounded by *k* nearest training data point in attribute space. The output which is membership of a class depends on whether the classifier is used for regression or classification. An object is classified on the basis of majority vote of its neighbors. The object is assigned to the particular class with most common in its neighborhood. Suppose *k*=1, then object will definitely assigned to the class of single nearest neighbor. In case of regression property value of object is output. This output value is average value of its *k*-nearest neighborhoods. The *k*-NN algorithm is among the simplest of all machine learning algorithms; however, a limitation of the *k*-NN algorithm is its sensitivity to local structure of the data. Training of classifier is done by storing the attribute vectors and class labels of the training instances. However, classification is done by assigning a most frequent unlabeled vector among the *k* training instances nearest to the query point. Euclidean distance is most commonly used distance metric for continuous variables.

The K-nearest-neighbor (*k*-NN) algorithm works by computing the distance between a query scenario to its set in the data set. The distance between scenarios can be computed using *d(x,y) (*distance function), where *x and y* are the query scenarios composed of N features such that *x={x1,…..,xn}, y={y1,…..,yn}*. Two different types of distance function discussed in the summary are as follows:

- Absolute distance measuring:

$$d_A(x,y) = \sum_{i=1}^{N} |x_i - y_i| \quad (3)$$

- Euclidean distance measuring:

$$d_E(x,y) = \sum_{i=1}^{N} \sqrt{{x_i}^2 - {y_i}^2} \quad (4)$$

6 Results and discussion

The condition monitoring of roller bearing is performed by K-star classifier and *k*-nearest neighborhood classifier using machine learning approach. The results of the study are discussed.

6.1 Feature classification using K Star algorithm

For the K-star algorithm, the classification accuracy is maximum with 9 features. The classifier is able to classify the features with an accuracy of 87.5 % and has least false positive of 3 with the same number of features which means out of 120 data sets only 3 data sets will show a false positive. For further optimizing the classifier performance the value of global blend is varied from 0 to 100 with a step size of 5. The maximum classification accuracy is observed when the value of global blend was 45. The classification accuracy is increased from 87.5 % to 88.33 % without any change in false positive. The variation of classification accuracy is shown in Fig 4.

Table 1: Effect of number of features on classification accuracy and false positive for K-star classifier

No. of features	Classification accuracy (%)	False positive
1	58.33	11
2	74.17	6
3	77.5	7
4	77.5	6
5	80.83	6
6	79.17	4
7	80.83	5
8	85.83	4
9	87.5	3
10	85	3

11	85	3
12	87.5	4
13	87.5	3

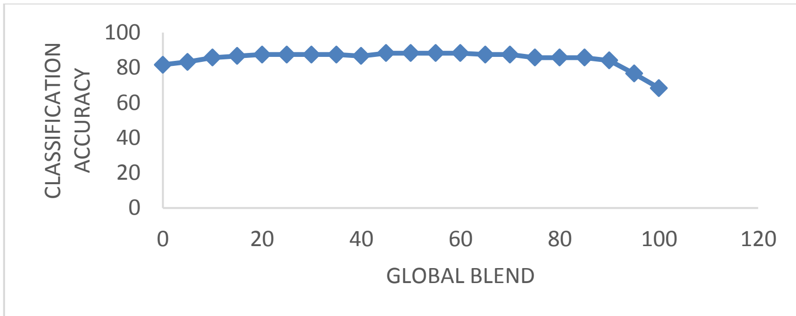

Figure 4: Effect of global blend on classification accuracy of k-star algorithm

Table 2: Confusion matrix for K-star classifier

	Good	IRF	ORF	IORF
Good	23	1	6	0
IRF	0	30	0	0
ORF	3	1	25	1
IORF	0	2	0	28

Table 2 is confusion matrix of K-star classifier with 9 features with global blend value of 45. Confusion matrix shows value of condition as class is classified to which value of the class. The conditions are Good, IRF (Inner race fault), ORF (Outer race fault) and IORF (Inner and outer race faults). This is best way to visualize the classification results. For each class of condition attribute 30 samples were considered. It looks in the form of a square matrix. Referring to Table 2, sum of all elements of first row represents the total number of instances which corresponds to good bearing condition (Good). The first column represents that instances which are classified as good bearing conditions from different classes of condition attribute. The values in row represents the real condition of data points and values in column represents the condition classes in which data points were classified by the classifier. Element in the first row and first column represents 23 data points with condition as good are classified as good. Hence, every diagonal element of confusion matrix represents correctly classified instances and non-diagonal elements represents incorrectly classified instances. Element in column 2 and row 1 represents 1 data point with good condition was classified incorrectly as IRF. Hence, sum of all elements in column 1 except the first row represents the false positive of the classifier, which in the case is 3. Last row signifies that out of 30 IORF bearing condition data points 28 data points were correctly classified as IORF bearing condition data points, 2 were incorrectly classified as IRF bearing condition data points; however, no data point is incorrectly classified as ORF or good bearing condition data point.

6.2 Feature classification using k-nearest neighborhood algorithm

For the *k*-NN algorithm, the classification accuracy is maximum with 5 features. The classifier is able to classify the features with an accuracy of 83.33 % and has least false positive of 5 with same number of features which means out of 120 data sets only 3 data sets will show a false positive. For further optimizing the classifier performance the value of *k*-NN is varied from 1 to 20 with a step size of 1. The maximum classification accuracy is observed when the value of *k*-NN is 1. The variation of classification accuracy and false positive with number of feature is shown by a plot in Fig 5.

Table 3: Effect of number of features on classification accuracy and false positive for k-NN classifier

No. of features	Classification accuracy (%)	False positive
1	45.83	16
2	64.17	13
3	71.67	10
4	72.5	8
5	83.33	5
6	75.83	8
7	78.33	6
8	80.83	6
9	81.67	5
10	81.67	5
11	81.67	5
12	80.83	5
13	79.17	5

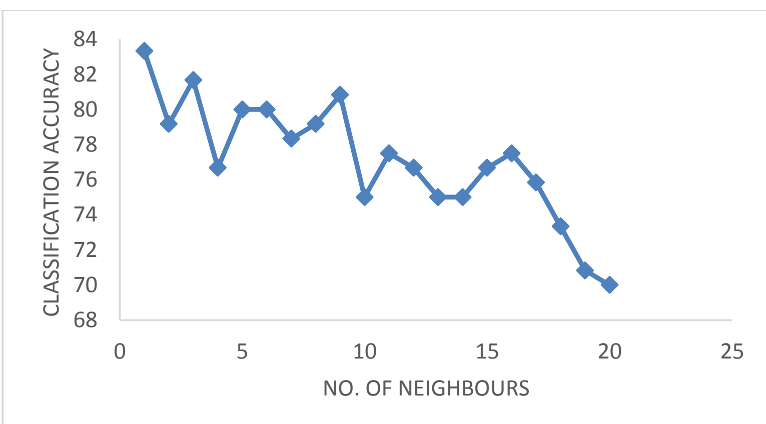

Figure 5: Effect of k on classification accuracy for k-nearest neighborhood classifier

Table 4: Confusion matrix for k-nearest neighborhood classifier

	GOOD	IRF	ORF	IORF
GOOD	23	3	3	1
IRF	1	26	0	3
ORF	4	1	24	1
IORF	0	3	0	27

Table 4 is confusion matrix of *k*-NN classifier with 5 features and value of *k*-NN as 1. For each class of condition attribute 30 samples were considered. Element in the first row and first column represents 23 data points with condition as good are classified as good. Hence, every diagonal element of confusion matrix represents correctly classified instances and non-diagonal elements represents incorrectly classified instances. Element in column 2 and row 1 represents 3 data point with good condition is classified incorrectly as IRF. Hence, sum of all elements in column 1 except the first row represents the false positive of the classifier which in the case is 5. Last row signifies that out of 30 IORF class data points 27 data points are correctly classified as IORF class data points, 3 are incorrectly classified as IRF class data points of condition attribute. However, no data point is incorrectly classified as ORF or good bearing condition data point.

6.3 Detailed accuracy by class

True positive rate is proportion of correctly classified data points in the total set and is also known as TP rate. False positive rate is proportion of misclassified data points in the total set and is also known as FP rate. *Precision* is the fraction of retrieved instances that are relevant. However, *recall* is the fraction of relevant instances that are retrieved and it is also known as sensitivity.

True Positive rate is the rate of classifying correct instances as correct by the corresponding classifier. *False positive rate* is the rate of classifying incorrect instances as correct by the corresponding classifier. *F-Measure* is harmonic mean of *precision* and *recall* as show below. F-measure can be seen as a compromise between *recall* and *precision*. It is high only when both *recall* and *precision* are high. *MCC* stands for Matthews Correlation Coefficient which accounts all combinations of true and false and provides a balanced measure which could be applied even with different size classes. *MCC* returns value between −1 and +1. Value of +1 represents a perfect classification, however, value of 0 represents random prediction and value of −1 indicates total misclassification. *Receiver Operating Characteristic* area is the area under *ROC* curve. *Precision-Recall* area is area under *PRC* Curve.

$$Precision = \frac{|\{\text{correct data points}\} \cap \{\text{all data points}\}|}{|\{\text{all data points}\}|}$$

$$Recall = \frac{|\{\text{correct data points}\} \cap \{\text{all data points}\}|}{|\{\text{correct data points}\}|}$$

$$\mathrm{F} = 2.\frac{precision.recall}{precision+recall} \tag{5}$$

The detailed accuracy by class is found. The detailed result for the *k*-star classifier and *k*-nearest neighborhood classifier for the maximum accuracy are shown in Fig. 3(a) and 3(b) respectively.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.767	0.033	0.885	0.767	0.821	0.771	0.967	0.849	good
	1.000	0.044	0.882	1.000	0.938	0.918	0.998	0.994	irf
	0.833	0.067	0.806	0.833	0.820	0.758	0.916	0.835	orf
	0.933	0.011	0.966	0.933	0.949	0.933	0.990	0.978	iorf
Weighted Avg.	0.883	0.039	0.885	0.883	0.882	0.845	0.967	0.914	

Figure 6: Detailed accuracy by class for K-star algorithm

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.767	0.056	0.821	0.767	0.793	0.728	0.856	0.688	good
	0.867	0.078	0.788	0.867	0.825	0.765	0.894	0.716	irf
	0.800	0.033	0.889	0.800	0.842	0.795	0.883	0.761	orf
	0.900	0.056	0.844	0.900	0.871	0.827	0.922	0.784	iorf
Weighted Avg.	0.833	0.056	0.835	0.833	0.833	0.779	0.889	0.737	

Figure 7: Detailed accuracy by class for K-NN algorithm

Fig. 6 shows the detailed accuracy of K-star classifier with 9 statistical features. In Table 2, first element of the confusion matrix represents 23 out of 30 instances are correctly classified as good bearing condition and 7 data points are misclassified as IRF and ORF bearing condition. From Fig. 6, *precision* is 0.885, true positive rate is 76.7 % (0.767) and false positive rate is 3.3 % (0.033) for the classification performed by the classifier. F-measure value is found to be 82.1% (0.821).

Fig. 7 shows the detailed accuracy of *k*-nearest neighborhood classifier with 5 statistical features. In Table 4, first element of the confusion matrix represents 23 instances are correctly classified as good bearing condition and 7 instances are misclassified as IRF, ORF and IORF bearing condition out of 30 instances. From Fig.7, *precision* is 0.821, true positive rate is 76.7 % (0.767) and false positive rate is 5.6 % (0.056) for the classification performed by the classifier. F-measure value is found to be 79.3% (0.793).

7.4 Classifier comparison

As shown in Fig. 8 the classification accuracy of K-star is higher than *k*-NN for 12 out of 13 features. The highest classification accuracy achieved by K-star is 88.33 % with 9 features. The highest classification accuracy achieved by *k*-NN is 83.33 with 5 features. As shown in Fig. 9, the false positive of K-star is lesser than *k*-NN for 12 out of 13 features. Least false positive for K-star is 3 with 9, 10 and 11 features; however, number of features will be considered as 9 because of highest classification accuracy at this point and lesser computational cost involves due to least possible number of features. Least false positive value for *k*-NN is 5 for 5, 9, 10, 11, 12 and 13 features; however, we will consider no. of features as 5 because of highest classification accuracy at this point and lesser computational cost involves due to least possible number of features the comparative study says that K-star classifier has a higher classification accuracy and lesser false positive than *k*-nearest neighborhood classifier. This clearly indicates that K-star is a better classifier from both measurements for the given setup. The results obtained in this study are specific to the dataset used and a similar performance in not assured for all feature datasets in other conditions.

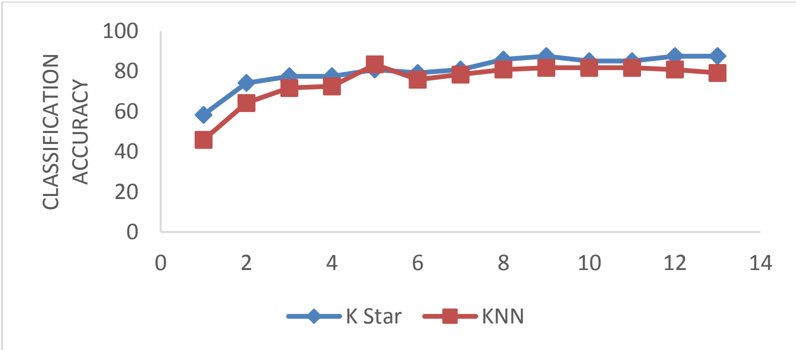

Figure 8: Classification accuracy vs. number of statistical features

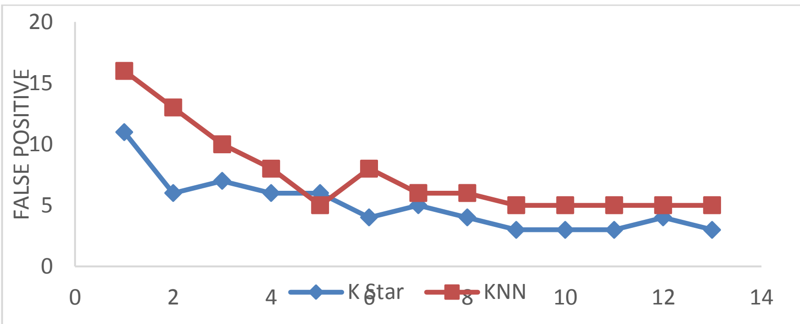

Figure 9: False positive vs. number of features

7 Conclusions

This comparative study is performed to find the performance of K-star and K-NN classifiers from lazy family in condition monitoring of roller bearing using sound signal. Four bearing condition were studied on the testing setup and sound signal is acquired. Statistical feature extraction process was performed on the data set and feature selection is performed using decision tree and random tree classifier. Feature classification is performed using K-star and *k*-NN classifiers. The results from the K-star and K-NN Net were compared on the basis of classification accuracy and number of false positive. From results, one can easily conclude that the classification accuracy using K-star classifier is higher than *k*-NN classifier for the same dataset. K-star classifier result shows a good performance with an accuracy of 88.33 % when tested with 9 statistical features and optimized value of global blend as 45. In a similar study, R. K. Sharma el al. [28] found a classification study of 86.67 % using decision tree algorithm with representative dataset having four conditions of bearing. The same dataset had achieved a classification accuracy of 89.16% using Bayes Net algorithm. Comparing the result, it is found that K-star classifier achieves higher classification accuracy than decision tree algorithm. The classification accuracy of K-star classifier is 88.33 % which is very similar to classification accuracy of 89.16 % by Bayes net classifier; however, K-star is showing a false positive of 3 whereas Bayes Net has a false positive of 4. This is superior as classification of a bad condition bearing as good condition bearing can lead to a complete shutdown and heavy loss. Hence, K-star is among the best algorithm in the list for condition monitoring of roller bearing using machine learning approach.

References

Diego Fernández-Francos; David Martínez-Rego; Oscar Fontenla-Romero; Amparo Alonso-Betanzos (2013): Automatic bearing fault diagnosis based on one-class m-SVM. Computers & Industrial Engineering 64, pp. 357–365.

F. Yuan; F.-L. Chu (2006): Support vector machines-based fault diagnosis for turbo-pump rotor. Mechanical Systems and Signal Processing 20, pp. 939–952.

H.Q. Wang; P. Chen (2007): Sequential condition diagnosis for centrifugal pump sys-tem using fuzzy neural network. Neural Information Processing: Letters and Reviews 2, pp. 41–50.

Huaqing Wang; Peng Chen (2011): Intelligent diagnosis method for rolling element bearing faults using possibility theory and neural network. Computers & Industrial Engineering 60, pp. 511–518.

J. Chen; H. Huang (2009): Feature selection for text classification with Naïve Bayes. Expert Systems with Applications 36, pp. 5432–5435.

Jin Yi (2012): Design of hydraulic Fault diagnosis system based on Lab VIEW. Advanced Materials research Vols., pp. 457-458.

K. Polat; S. Gunes (2009): A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Systems with Applications 36, pp. 1587–1592.

K. Shibata; A. Takahashi; T. Shirai (2000): Fault diagnosis of rotating machinery through visualization of sound signals. Mechanical Systems and Signal Processing 14, pp. 229–241.

L. Zhang; L.B. Jack; A.K. Nandi (2005): Fault detection using genetic programming. **Mechanical Systems and Signal Processing 19, pp. 271–289.**

M. Elangovan (2010): Studies on Bayes classifier for condition monitoring of single point carbide tipped tool based on statistical and histogram features. Expert Systems with Applications 37, pp. 2059–2065.

M. Amarnath; V. Sugumaran; Hemantha Kumar (2013): Exploiting sound signals for fault diagnosis of bearings using decision tree. Measurement 46, pp. 1250–1256.

N.R. Sakthivel; V. Indira; B.B. Nair; V. Sugumaran (2011): Use of histogram features for decision tree based fault diagnosis of monoblock centrifugal pump. Inter-national Journal of Granular Computing, Rough Sets and Intelligent Systems (IJGCRSIS) 2, pp. 23–36.

N. Baydar; A. Ball (2003): Detection and diagnosis of gear failure via vibration and acoustic signals using wavelet transform. Mechanical Systems and Signal Processing 17 (4), pp. 787–804.

R.B.W. Heng; M.J.M. Nor (1997): Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition. Applied Acoustics 53 (1–3), pp. 211–266.

S.-F. Yuan; F.-L. Chu (2007): Fault diagnostics based on particle swarm optimization and support vector machines. Mechanical Systems and Signal Processing 21, pp. 1787–1798.

Samanta B.; Al-balushi K.R.; Al-araim S.A. (2003): Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. Engineering Applications of Artificial Intelligence, Vol. 16, pp.657–665.

Sharma, R.K.; Sugumaran, V.; Kumar, H.; Amarnath, M. (2015): A comparative study of naive Bayes classifier and Bayes net classifier for fault diagnosis of roller bearing using sound signal. Int. J. Decision Support Systems, Vol. 1, No. 1, pp.115–129.

Sugumaran V.; Ramachandran K. I. (2007): Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. Mechanical Systems and Signal Processing 21, pp. 2237–2247.

Suykens J. A. K.; Van Gestel T.; Vandewalle J.; De Moor B. (2003): A support vector machine formulation to PCA analysis and its Kernel version. ESAT-SCD-SISTA **Technical Report.**

T. Ananthapadmanaban; V. Radhakrishnan (1983): An investigation of the role of surface irregularities in the noise spectrum of rolling and sliding contact. WEAR 83, pp. 300–409.

V. Muralidharan; V. Sugumaran (2012a): A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. Journal of Applied Soft Computing, pp. 1-7.

V. Sugumaran; K.I. Ramachandran (2007b): Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. Mechanical Systems and Signal Processing 21, pp. 2237–2247.

Ye tian; Jian Ma; Chen Lu; Zili Wang (2015): Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. Mechanism and Machine Theory, vol. 90, pp. 175-186.

Zhiqi Fan; Huaizhong Li (2015): A hybrid approach for fault diagnosis of planetary bearings using an internal vibration sensor. Measurement, vol. 64, pp.71-80.

Xiaoming Xue; Jianzhong Zhou; Yanhe Xu; Wenlong Zhu; Chaoshun Li (2015): An adaptively fast ensemble empirical mode decomposition method and its applications to rolling element bearing fault diagnosis. Mechanical systems and Signal Processing, Vol. 62-63, pp.444-459.

Xiaoyue Chen; Jianzhong Zhou; Jian Xiao; Xinxin Zhang; Han Xiao; Wenlong Zhu; Wenlong Fu (2014): Fault diagnosis based on dependent feature vector and probability neural network for rolling element bearings. Applied Mathematics and Computation, vol. 247, pp. 835-847.