Detection of Fault in Bearing of an Industrial Gearbox using Machine Learning Approaches

Ashish Kumar Srivastava[1], Abhishek Singh Gangwar[2]

[1]M.Tech. Scholar, Dept. of Mechanical Engineering, Goel Institute of Technology & Management, Lucknow, U.P., India

[2]Assistant Professor, Dept. of Mechanical Engineering, Goel Institute of Technology & Management, Lucknow, U.P., India

Abstract - In a gearbox, bearing plays a very vital role. Bearing is a mechanical component used to support shaft on which gears are mounted in a gearbox. It also reduces friction produced while torque is being transmitted. Rolling contact bearings used in an industrial gearbox. In a rolling contact bearing, a steel ball or steel roller is used to reduce the frictional resistance since rolling friction is always less than sliding friction. This helps to smoothly allow rotation of shaft. Since there is a surface contact between rolling element and races, defects may get induced in them. This has to be eradicated as major portion of loss in power results due to faulty bearing of any gearbox. Machine learning algorithms can be utilized to monitor health and condition of bearing while they are performing their work of power transmission. Results obtained from Machine Learning algorithms makes it easy to predict health and maintenance type for bearings.

All experiments have been performed on the bearing dataset. This work predicts the bearing failure in five different class named as Ball fault, Inner race fault, Outer race fault, surface fault, Healthy bearing. These experiments involve different machine learning algorithm and one deep learning algorithm. In machine learning algorithm, we compare and analyse the K-nearest neighbour, decision tree, random forest, support vector machine algorithm. In deep learning we analyses multilayer perceptron algorithm. Our experimental results have shown that random forest algorithm is performing best with 87.15% accuracy and 0.192 root mean square error.

Keywords: Bearing, Gearbox, Machine Learning, K-NN, SVM, Decision Tree, Random Forest

I. INTRODUCTION

Bearing (for this work rolling contact bearing) also known as anti-friction used in industrial gearbox are mostly made of steel and are prone to various kinds of factors acting on them while working such as speed, load, shocks, etc. These may be responsible for certain failure occurring in a bearing and its elements. Bearings are made of steel and proper heat treatment is done to avoid any kind of stresses in material. During operation they incur some amount of stress which may result in failure if working is continued in this condition.

To avoid such occurrence bearings, need to be examined for any such signs. This has to be done very intensely and for doing so gearbox has to be made ideal. This hampers productivity and loss is increased of time and money. Such efforts can be reduced by maximizing productivity, which involves use of Machine Learning concept. By using Machine Learning algorithms, we can predict the fault type and provide a precise simulation data to choose an appropriate type of maintenance strategy.

The main objective of this paper is finding the best performing algorithm for bearing dataset. This work also aims to compare different machine learning algorithms. To accomplish this, we have selected different type of algorithms like one algorithm is instance based (K-NN), two are tree-based algorithm named decision tree and random forest. Another one is kernel based algorithm SVM and last algorithm is deep learning based multilayer perceptron algorithm.

This paper is arranged as: Section 2 enumerates various machine learning algorithms used in this work. Section 3 explains proposed methodology for the work. Section 4 provides experimental results for the work. Section 5 enumerates about conclusion and last section explains about future scope of proposed work.

II. PREVIOUS WORK

Most of the researchers have worked on bearing fault detection problem [3], [8], [9]. They approached this problem using machine learning techniques. Machine learning techniques like decision tree, k-nn and many others. In this work, we will try to focus on different types of machine learning algorithms.

In this section a discussion about various machine learning algorithms utilized in this work will be done.

2.1 K-Nearest Neighbor algorithm

K-NN is Instance based algorithm also known as lazy learning algorithm. There is no training phase in this algorithm. At the time of inferencing, new instance finds the K nearest neighbors according to Euclidean distance. According to voting method, it will select the final class label. It is important to select the appropriate value of k. The value of k should be positive and odd number [4].

2.2 Decision tree algorithm

Decision tree is tree based one of the famous algorithms. This algorithm builds tree based on the features and its value. Information gain method is used to find the best feature. In this tree, nodes represent feature and edges represents feature value. At the time of inferencing, new instance traverses the full tree and predict the class label [5].

2.3 Random forest Algorithm

Random forest algorithm is updated version of decision tree algorithm. This algorithm gives equal priority to all feature and construct same number of trees as features available in dataset. At the time of inferencing, new instance goes through all tree and predict class labels. Majority voting method is used to find the final class prediction. This algorithm mostly outperforms as compare to decision tree algorithm because of giving equal opportunity to all feature values [1].

2.4 Support Vector Machine

Support vector machine is famous kernel-based algorithm. SVM algorithm separate two class instances with maximum margin. This algorithm follows principle of linear separability [6]. SVM algorithm can be of types:

SVM linear classifier: If dataset is linearly separable. SVM predicts the linear boundary.

SVM nonlinear classifier: If dataset is not linearly separable, SVM algorithm predicts the nonlinear hyperplane. Examples of nonlinear hyperplanes is redial basis kernel, polynomial kernel, puk kernel and normalized poly kernels, these can also be customized according to the dataset plot.

2.5 Multi-Layer Perceptron

Multilayer perceptron is part of deep learning algorithm. This algorithm is based on the biological nervous system. Also, simple perceptron works on concepts of linearly separable. If dataset is not linearly separable, multilayer perceptron algorithm is used. This algorithm has hidden layer in between input layer and output layer. This algorithm outperforms as compared to simple perceptron algorithm if dataset is not linearly separable.

III. PROPOSED METHODOLOGY

In this section we will discuss about the proposed methodology for work of this paper. It is as follows:

- Collecting dataset.
- Dataset is pre-processed for calculation.
- Above mentioned algorithms are analysed for obtaining the result.
- Collection of analysed result is done.
- Results from all the algorithms are compared for further analysis.
- On the basis of analysed results a best suited algorithm is proposed.

IV. SIMULATION/EXPERIMENTAL RESULTS

4.1 Experimental Setup

All experiments are performed on the WEKA tool [2]. WEKA is an open source and machine learning tool used for machine learning algorithm analysis. Figure 4.1 shows the screenshot of the tool where we can select the algorithm.

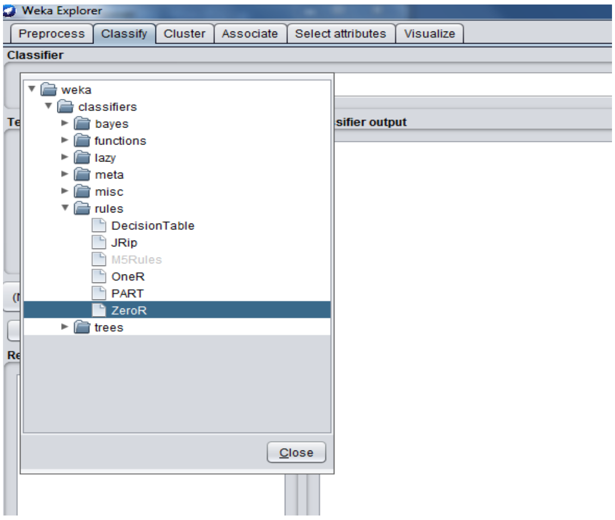

Fig. 2.1 WEKA Tool

4.2 Dataset

All experiments have been performed on the bearing dataset and it taken from [3]. Table 1 shows the feature set and Table 2 shows the class labels of the dataset. This dataset has five classes, four class are fault type and last one is for healthy dataset.

TABLE 1. FEATURE SET OF BEARING DATASET

1: Motor Vibration
2: Vibration of planetary gearbox in x direction
3: Vibration of planetary gearbox in y direction
4: Vibration of planetary gearbox in z direction
5: Motor torque
6: Vibration of parallel gear box in x direction
7: Vibration of parallel gear box in y direction
8: Vibration of parallel gear box in z direction
9: Rotating speed of motor spindle
10: Applied load

TABLE 2. CLASS LABEL OF THE BEARING DATASET

Class 1: Ball (crack on ball)
Class 2: Inner (crack on Inner ring)
Class 3: Outer (crack on outer ring)
Class 4: surface (crack in surface of gear)
Class 5: Healthy bearing

4.3 Performance Metrics

It is very important to validate the performance of machine learning algorithms. In our experiments, we have used accuracy and root mean square error (RMSE). Since, validating an algorithm based on accuracy is not good idea. It is necessary to add one more performance metric that can show how good classification is. So, we have included RMSE value.

4.4 Experimental Results

This subsection shows the experimental result of different machine learning algoritshms and deep learning algorithm. This work also compares the performance of all these algorithms. Algorithms named K-nearest neighbor, decision tree, random forest algorithm, support vector machine and multilayer perceptron are used for performance comparison and analysis.

In K-NN algorithm, where k is positive and odd value, we have selected the value of k as five based on some experiments. Decision tree algorithm uses J48 decision tree algorithm. MLP has twenty hidden layers.

Figure 1 shows the accuracy comparison of bearing dataset. Random forest algorithm is performing best with 87.15% accuracy. After that decision tree algorithm achieves 83.09% accuracy. MLP achieved 81.1% accuracy. KNN and SVM achieves least accuracy of 74.61% and 75.76% respectively.

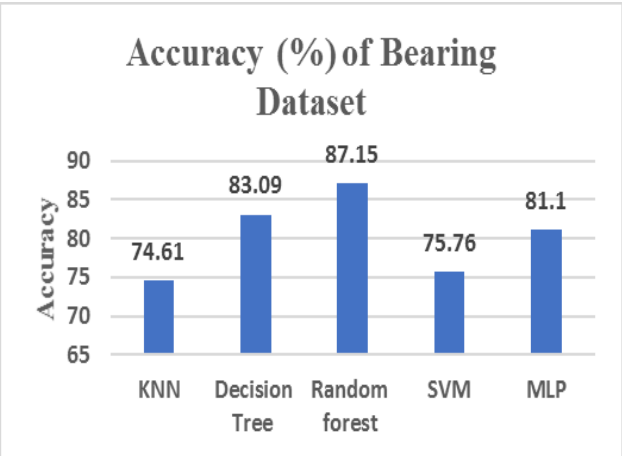

Fig. 5.1 Accuary of Bearing Dataset

Taking decision of best classifier using only accuary is not appropriate. It is also important to consider ho accuarte the algorithm is perfomring. So, in our experimnets we have included root mean square error (RMSE) to take the final decision.

Figure 2 shows the RMSE value of different algorithm on bearing dataset. Random forest algorithm achieves 0.192 error. Decsion tree achieved 0.233 error. KNN and SVM algorithm achieves almost same error of 0.26 and 0.227 error respectively.

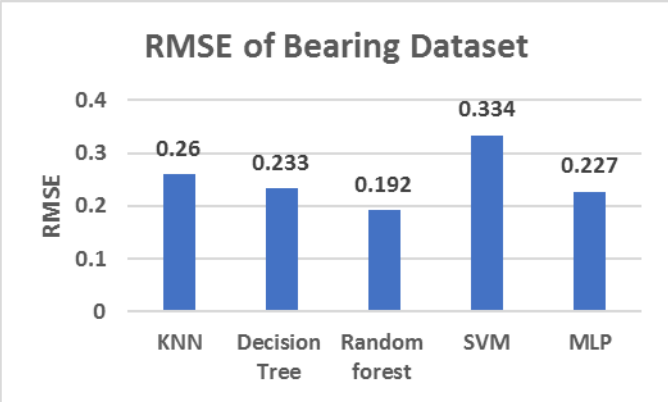

Fig. 5.2 RMSE of Bearing Dataset

Referring to Both Figure 1 and Figure 2, random forest algorithm is perfoming best for bearing dataset with 87.15% accuarcy and 0.192 RMSE. After that decision tree is performing well with 83.09% and 0.233 error.

Since, Random forest algorithm is perfoming the best, next section discuss the resukts of random forest algorithm.

TABLE 1: CONFUSION MATRIX OF RANDOM FOREST ALGORITHM ON BEARING DATASET

Actual/ predicted	Ball (B)	Ball (B)	Ball (B)	Ball (B)	Ball (B)
Ball (B)	1777	0	127	98	0
Inner (I)	0	1799	0	0	203
Outer (O)	244	0	1556	202	0
Surface (S)	74	0	134	1794	0
Healthy (H)	0	204	0	0	1798

Table 1 shows the confusion matrix of the random forest algorithm on bearing dataset where numbers in bold and in diagonal shows the correct classification and rest in row shows incorrect classification. For class label Ball, 1777 insatnces are correclty classified out of 2002. Similary, for class label Inner 1799 insatnces are correctly classified out of 2002 insatnce.

TABLE 2: CLASS LEVEL ACCURACY OF RANDOM FOREST ALGORITHM ON BEARING DATASET

Class Label	Class Label Accuracy
Ball (B)	88.76%
Inner (I)	89.86%
Outer (O)	77.72%
Surface (S)	89.61%
Healthy (H)	89.81%

It is also important to check the class level accuracy of the algorithm. Table 2 shows the class level accuracy of the random forest algorithm. Class label Inner, Surface and Healthy achieved almost similar accuracy of ~89%. Class label Ball achieved 88.76% accuracy and class label Outer achieved 77.72% accuracy.

V. CONCLUSION

This paper presents the machine learning approaches for bearing dataset. Experimental results have discussed the results of different machine learning algorithms named K-NN, decision tree, random forest, support vector machine, multilayer perceptron. Random forest algorithm is best performing algorithm with 87.15% accuracy and 0.192 error.

VI. FUTURE SCOPES

The proposed work can be extended to develop a simulation model which can be used for predicting fault of bearing of an industrial gearbox without hampering its working.

This work can be extended to evaluate deep learning algorithms and increase the fault prediction rate through this.

REFERENCES

[1] L. Breiman, Random forests, Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

[2] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[3] Shao, Siyu, et al. "Highly-accurate machine fault diagnosis using deep transfer learning." IEEE Transactions on Industrial Informatics (2018).

[4] A.-L. Jousselme and P. Maupin, Distances in evidence theory: Comprehensive survey and generalizations, International Journal of Approximate Reasoning, vol. 53, no. 2, pp. 118-145, 2012.

[5] S. K. Murthy, Automatic construction of decision trees from data: A multi-disciplinary survey, Data mining and knowledge discovery, vol. 2, no. 4, pp. 345-389, 1998.

[6] Wu, Xiujun. "Supp

[7] ort Vector Machine SMO Algorithms and Their Optimization." In 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019). Atlantis Press, 2019.

[8] Marsland, Stephen. Machine learning: an algorithmic perspective. Chapman and Hall/CRC, 2011.

[9] Li, Pengfei, Yongying Jiang, and Jiawei Xiang. "Experimental investigation for fault diagnosis based on a hybrid approach using wavelet packet and support vector classification." The Scientific World Journal 2014 (2014).

[10] Samanta, Biswanath, Khamis R. Al-Balushi, and Saeed A. Al-Araimi. "Artificial neural networks and genetic algorithm for bearing fault detection." Soft Computing 10, no. 3 (2006): 264-271.