Comparative Study of Decision Tree Models for Bearing Fault Detection and Classification

Armin Moghadam[1], *Fatemeh Davoudi Kakhki*[1,2]

[1] Department of Technology, San Jose State University,
San Jose, CA 95192, USA

[2] Machine Learning & Safety Analytics Lab, Department of Technology, San Jose State University,
San Jose, CA 95192, USA

ABSTRACT

Fault diagnosis of bearings is essential in reducing failures and improving functionality and reliability of rotating machines. As vibration signals are non-linear and non-stationary, extracting features for dimension reduction and efficient fault detection is challenging. This study aims at evaluating performance of decision tree-based machine learning models in detection and classification of bearing fault data. A machine learning approach combining the tree-based classifiers with derived statistical features is proposed for localized fault classification. Statistical features are extracted from normal and faulty vibration signals though time domain analysis to develop tree-based models of AdaBoost (AD), classification and regression trees (CART), LogitBoost trees (LBT), and Random Forest trees (RF). The results confirm that machine learning classifiers have satisfactory performance and strong generalization ability in fault detection, and provide practical models for classify running state of the bearing.

Keywords: Bearing Fault, Decision Trees, Time Domain Statistical Features, Machine Learning Classifiers

INTRODUCTION

Induction motors are one of the most significant parts of manufacturing systems operation and are prone to failure due to various electrical and mechanical stress and long operation hours (Toma et al., 2020). Fault diagnosis in induction motors is important for condition monitoring, enhancing reliability and availability of motors, and avoiding major loss in production and downtime (Tahir et al., 2017).

Bearing is one of the most vulnerable parts in a motor and drive systems, and typically consists of four components of the inner raceway, outer raceway, ball, and cage. Faults in any of these components creates changes in vibration signals, and therefore, monitoring and analyzing vibrational signals is helpful in understanding bearing faults based on its location (Chen et al., 2017). In other words, comparing faulty signals with normal conditions allows for fault detection (Sugumaran & Ramachandran, 2011). Bearing fault is the most commonly-occurred type of faults, responsible for 30-40% of all the machine failures (Chen et al., 2017; Jallepalli & Kakhki, 2021).Therefore, efficient modeling methods for analyzing and detecting bearing faults is critical in improving the functionality of industrial and manufacturing systems (Milo et al., 2014).

Data-driven modeling such as machine learning (ML) techniques utilize past data to generate meaningful predictive patterns in both classification and regression problems, with categorical and numerical target variables, respectively (Badarinath et al., 2021). Among various supervised ML models, decision tree techniques are powerful tools for classification and regression purposes. Adaptive boosting decision trees, known as AdaBoost (AD), is an iterative algorithm for constructing a classifier based on a linear combination for classification of a target variable. Moreover, AD is a simple enhancing process of weak classification algorithms, and this process can enhance the ability for data classification by reducing both bias and variance through ongoing training. The AD algorithm first begins with building different tree classifiers on the same training set, and then combining all those weak classifiers to construct a final strong classifier through adaptively adjustment of the weak classifier errors obtained by weak learners. The algorithm itself is achieved by changing the data distribution, which determines the weight of each sample based on the correct classification of each sample in each training set and the accuracy of the prior overall classification. The new data set with modified weights is sent to the lower classifier for training. Finally, the classifier obtained by each training is combined together to be the final decision classifier (Wu et al., 2020). Classification and Regression Tree (CART) (Breiman et al., 2017) is another powerful classifier that can create meaningful outputs by partitioning the feature space and building binary decision trees recursively. CART classifier is popular in application due to being flexible, interpretable, and straightforward to comprehend (Murari et al., 2020). LogitBoost (LBT) is another ensemble-based decision tree that uses boosting for reducing bias and variance in the classification tasks by applying logistic regression principles on the AdaBoost generalized additive model (Tehrany et al., 2019). Random Forest (RF) algorithm is another tree-based ensemble classifier that reduces the generalization error through building many trees as the basis for classification (Fang et al., 2020). The big number of trees use bagging method which adds randomness to the tree building process (Liakos et al., 2018).

Developing data-driven ML models need large datasets that include data on both faulty and normal conditions (Ebrahimifakhar et al., 2020). One of the main benchmark datasets for bearing fault analysis is available through the Case Western Reserve University (CWRU) website database (Fault & Data, n.d.), which has been used as a benchmark set for developing various statistical, machine learning, and deep learning models with the purpose of detecting and classifying faults in bearings (Zhang et al., 2020). The main challenge for modeling CWRU data and consequently various results is that the original data do not have a manual for instruction on classification experiments, and therefore leaves researchers with the challenge of interpretation and selection of feature extraction methods for achieving higher accuracy rates and useful models (Rauber et al., 2020).

This paper demonstrates a data-driven fault detection modeling approach based on statistical machine learning classifiers, specifically decision-tree based methods. This study contributes to the current literature on practicality of ML models in classification and detection of bearing faults by evaluating the performance of decision tree models including AD, CART, LBT, and RF.

The rest of this paper is organized as follows. Section 2 explains the process for extracting statistical features from original data files, and developing ML classifiers for bearing fault detection. Section 3 includes details of the results, followed by a brief discussion of the study contribution and future research direction in Conclusion, presented in Section 4 of the paper.

MODEL DEVELOPMENT AND FEATURE EXTRACTION

Comparing faulty vibration signals with normal signals is a popular method to detect faults in bearings (Sugumaran & Ramachandran, 2011). However, due to the large data dimension from vibration signals at different locations of the bearings, feature extraction methods should be applied for reducing data dimensions and improving data-driven modeling approach. Among various feature extraction methods, this paper uses statistical feature extracted from the vibration signals based on similar approaches in (Kankar et al., 2011; Sugumaran & Ramachandran, 2011; Tahir et al., 2017).

One of the main benchmark datasets available for bearing fault analysis is available through the Case Western Reserve University (CWRU) website database, which contains data files on both faulty and normal conditions of bearing, and includes the details of the data (Fault & Data, n.d.). Considering the 48kHz bearing faults in the original data, there are at least 480,000 sample points, or the length of time series data. To train the ML classifiers, the data is segmented to reduce the data dimensionality. Furthermore, to include data points representative of the entire bearing time series data, samples at certain intervals such as a fixed gap in samples, rather than choosing first few consecutive data points, are drawn. Also, the segmentation process can reduce the overlap between two data sample intervals. In this data, the sample rate is 48 kHz, approximately 230 sample points per revolution for each type of bearing fault with length of 2048, and number of class labels of 10 for normal and faulty bearings, shown in **Error! Reference source not found.**. Therefore, Considering the 48KHz bearing faults at drive end, a final data set is prepared with dimension of (230*10, 2048) for the 10 various labels.

Table 1: Dataset obtained from CWRU bearing faults

Bearing condition	**Fault size (inch)**	**Created label**
Bearing fault (Ball)	0.007	Ball_0.007
	0.014	Ball_0.014
	0.021	Ball_0.021
Inner race fault (IR)	0.007	IR_0.007
	0.014	IR_0.014
	0.021	IR_0.021
Outer race fault (OR)	0.007	OR_0.007
	0.014	OR_0.014
	0.021	OR_0.021
Normal	-	Normal

In the next step, for each vibration signal, from either type of faulty or normal bearing, statistical feature of minimum, maximum, mean, standard deviation, root mean square error, skewness, kurtosis, crest form, and form factor are calculated and extracted, for the 48 kHz data (**Error! Reference source not found.**). The features extracted are then used as input variables for developing decision tree ML models to classify and predict the fault location and status in the dataset. The data for modeling is partitioned in 70% for training and 30%

for testing. The model performance is judged based on the test data.

Table 2: Statistical features extracted from the data

Statistical feature	**Equation**
Mean	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
Standard Deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}}$
Root Mean Square	$X_{rms} = \frac{\sqrt{\sum_{i=1}^{n} {x_i}^2}}{n}$
Skewness	$X_{skewness} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3}$
Kurtosis	$X_{kurtosis} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4}$
Crest Factor	$C_f = \frac{x_{max}}{x_{min}}$
Form factor	$F_f = \frac{\mu}{X_{rms}}$

To assess the performance of the ML classifiers, both overall model accuracy and class-specific accuracy is important. The former shows the general ability of the ML classification power, and the latter shows the capability of the model in distinguishing among various class output labels, and reveals the model success in specifically detecting and classifying a label. Multi-label confusion matrix is used for determining this task.

Using the confusion matrix for any classifier, the terms of specificity, sensitivity and precision can be obtained to characterize the performance of a classifier per class. Recall represents the proportion of correctly classified labels, while precision evaluates class agreement of the data labels with the positive labels defined by the classifier. Another metric that is gained from the multi-level confusion matrix is F-score, and is a harmonic mean of the precision and sensitivity. Furthermore, in cases of working with class imbalance issue, another effective evaluation criterion for performance of ML classification models is using the Matthews correlation coefficient (MCC) (Chicco & Jurman, 2020) as a more reliable evaluation metric compare to overall accuracy rate (Chicco et al., 2021). Another criterion for assessing class-specific accuracy is identifying the area under curve (AUC) from the receiver operating characteristic curve (ROC). The ROC curve depicts the value of specificity vs sensitivity, and the AUC is between 0 and 1. The importance of AUC values is in the fact that they provide information about the usefulness and predictive power of the model in specifically classifying each label, in particular when more than two classes are present as outputs in the dataset. The higher the AUC values, the more powerful a classifier is in differentiating between class labels.

RESULTS

We applied feature extraction of vibration signals though time domain analysis that were used for training decision tree-based machine learning classifiers. The time domain statistical parameters extracted from vibration signals include minimum, maximum, mean, standard deviation, root mean square, skewness, kurtosis, crest factor and from factor. This is an effective method for significantly reduce the dimension of data before training any machine learning models.

In the next step, AD, CART, LBT, and RF decision trees are used for detection and classification of bearing fault conditions. Considering 48KHz drive end bearing data, the accuracy comparison based on average values for overall accuracy rate, overall error rate, precision, recall, F-measure, and ROC area for all the classifiers is presented in **Error! Reference source not found.**. When applied to the CWRU bearing fault data set, all models showed high accuracy in classifying the localized bearing faults at multilevel in either ball, inner race, or outer race of a 48kHz drive end bearing with different diameter measurements at 0.007, 0.014, 0.021 inches. All tree classifiers have an average overall accuracy over 95%.

Table 3: ML model performance per classifier (on average) for 48K data on test set

ML Classifier	Precision	Recall	F-Measure	MCC	ROC Area
AD	0.957	0.957	0.957	0.952	0.994
CART	0.961	0.961	0.961	0.957	0.997
LBT	0.966	0.965	0.965	0.962	0.997
RF	0.963	0.962	0.962	0.959	0.998

All models have a high F-measure between 0.957 to 0.965 which belongs to LBT and is marginally higher than the others. The same is observed looking at the average ROC area, with RF model demonstrating the highest value of 0.998. The results confirm that the supervised machine learning classifiers have satisfactory performance and strong generalization ability, and therefore, can be used to classify the running state of the bearing. The MCC values are close to 1 (between 0.952-0.962) for all classifiers that show the high quality of the models in distinguishing among the various classes of the localized faults.

CONCLUSION

In this study, a machine learning approach combining the tree-based classifiers with derived statistical features is proposed for localized bearing fault classification. The effectiveness of the proposed approach is validated using the publicly available Case Western Reserve University (CWRU) dataset. The main challenge for modeling CWRU data and consequently various results is that the original data do not have a manual for instruction on classification experiments, and therefore leaves researchers with the challenge of interpretation and selection of feature extraction methods for achieving higher accuracy rates and useful models. The results of this study are comparable with other research in which CWRU data was analyzed with similar approach of applying time domain statistical features to decision tree models with high accuracy rates (Nishat Toma & Kim, 2020). The future direction of this study is to compare the performance of decision tree models on other available bearing fault datasets, and developing deep learning models to assess how deep learning and machine learning performance are compared for bearing fault classification and detection.

REFERENCES

Badarinath, P. V., Chierichetti, M., & Kakhki, F. D. (2021). A machine learning approach as a surrogate for a finite element analysis: Status of research and application to one dimensional systems. *Sensors*, *21*(5), 1–18. https://doi.org/10.3390/s21051654

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. In *Classification and Regression Trees*. https://doi.org/10.1201/9781315139470

Chen, Z., Deng, S., Chen, X., Li, C., Sanchez, R. V., & Qin, H. (2017). Deep neural networks-based rolling bearing fault diagnosis. *Microelectronics Reliability*, *75*, 327–333. https://doi.org/10.1016/j.microrel.2017.03.006

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. https://doi.org/10.1186/s12864-019-6413-7

Chicco, D., Tötsch, N., & Jurman, G. (2021). The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*, 1–22. https://doi.org/10.1186/s13040-021-00244-z

Ebrahimifakhar, A., Kabirikopaei, A., & Yuill, D. (2020). Data-driven fault detection and diagnosis for packaged rooftop units using statistical machine learning classification methods. *Energy and Buildings*, *225*, 110318. https://doi.org/10.1016/j.enbuild.2020.110318

Fang, P., Zhang, X., Wei, P., Wang, Y., Zhang, H., Liu, F., & Zhao, J. (2020). The classification performance and mechanism of machine learning algorithms in winter wheat mapping using Sentinel-2 10 m resolution imagery. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app10155075

Fault, S., & Data, T. (n.d.). *Case Western Reserve University Bearing Data Center. Bearings Vibration Data Sets, Case Western Reserve University*. https://csegroups.case.edu/ bearingdatacenter/home

Jallepalli, D., & Kakhki, F. D. (2021). *Data-Driven Fault Classi fi cation Using Support Vector Machines* (Vol. 3). Springer International Publishing. https://doi.org/10.1007/978-3-030-68017-6

Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with Applications*, *38*(3), 1876–1886. https://doi.org/10.1016/j.eswa.2010.07.119

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. In *Sensors (Switzerland)*. https://doi.org/10.3390/s18082674

Milo, M. W., Harris, B., Bjerke, B., & Roan, M. (2014). Anomaly detection in rolling element bearings via hierarchical transition matrices. *Mechanical Systems and Signal Processing*, *48*(1–2), 120–137. https://doi.org/10.1016/j.ymssp.2014.02.004

Murari, A., Peluso, E., Lungaroni, M., Rossi, R., & Gelfusa, M. (2020). Investigating the physics of tokamak global stability with interpretable machine learning tools. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/APP10196683

Nishat Toma, R., & Kim, J.-M. (2020). Bearing Fault Classification of Induction Motors Using Discrete Wavelet Transform and Ensemble Machine Learning Algorithms. *Applied Sciences*. https://doi.org/10.3390/app10155251

Rauber, T. W., da Silva Loca, A. L., Boldt, F. de A., Rodrigues, A. L., & Varejão, F. M. (2020). An experimental methodology to evaluate machine learning methods for fault diagnosis based on vibration signals. *Expert Systems with Applications*, *April*, 114022. https://doi.org/10.1016/j.eswa.2020.114022

Sugumaran, V., & Ramachandran, K. I. (2011). Effect of number of features on classification of roller bearing faults using SVM and PSVM. *Expert Systems with Applications*, *38*(4), 4088–4096. https://doi.org/10.1016/j.eswa.2010.09.072

Tahir, M. M., Khan, A. Q., Iqbal, N., Hussain, A., & Badshah, S. (2017). Enhancing Fault Classification Accuracy of Ball Bearing Using Central Tendency Based Time Domain Features. *IEEE Access*. https://doi.org/10.1109/ACCESS.2016.2608505

Tehrany, M. S., Jones, S., Shabani, F., Martínez-Álvarez, F., & Tien Bui, D. (2019). A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theoretical and Applied Climatology*. https://doi.org/10.1007/s00704-018-2628-9

Toma, R. N., Prosvirin, A. E., & Kim, J. M. (2020). Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. *Sensors (Switzerland)*. https://doi.org/10.3390/s20071884

Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., & Hong, H. (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena*. https://doi.org/10.1016/j.catena.2019.104396

Zhang, S. S., Zhang, S. S., Wang, B., & Habetler, T. G. (2020). Deep Learning Algorithms for Bearing Fault Diagnosticsx - A Comprehensive Review. In *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2972859