Bearing Fault Diagnosis using Machine Learning Models

Rahul Pandey
Miet, Meerut
Meerut

Vishal Dham
Miet, Meerut
Meerut

Nikhil Chaudhary
Miet, Meerut
Meerut

Rakesh Sambhyal
Miet, Meerut
Meerut

ABSTRACT
The bearing serves as a crucial element of any machinery with a gearbox. It is essential to diagnose bearing faults effectively to ensure the machinery's safety and normal operation. Therefore, the identification and assessment of mechanical faults in bearings are extremely significant for ensuring reliable machinery operation. This comparative study shows the performance of fault diagnosis of bearings by utilizing various machine learning methodologies, including SVM, KNN, linear regression, ridge regression, XGB regression, AdaBoost regression, and cat boosting regression. Bearings are like the unsung heroes of the mechanical world, immensely supporting and guiding the smooth motion in everything, from your car's wheel to the propeller in a ship. However, like other mechanical components, over the course of time, the constant use of bearings can lead to wear and tear, which may ultimately result in a fault.

Bearing faults can manifest in several ways, including vibration, noise, heat, and changes in lubrication that reduce the efficiency of a machine. Therefore, it is essential to regularly monitor the bearings and inspect them to detect any issues early on. The aim of this present work is to use the various ML methodologies and their application to the bearing's data to monitor the condition of the machine's bearing. The present work is carried out in four phases. In the first phase, the data from various loads is collected. In the second phase, the data undergoes exploratory data analysis (EDA).

Keywords
Bearing, Gearbox, Fault Diagnosis, Exploratory Data Analysis (EDA).

1. INTRODUCTION
Condition-based monitoring (CBM) is a crucial practice for maintaining the reliability of rotating machinery by identifying potential issues at an early stage. This approach helps prevent system failure by allowing for preventative measures to be taken in a timely manner. Bearing fault

diagnosis is particularly important as bearings undergo distinct stages when subjected to different loads, including the healthy stage, damage stage, and extra damage stage. Early detection of faults is crucial to preventing total system failure. In this study, various machine learning models will be employed to diagnose faults in bearings. These algorithms will analyze the behavior of bearings under varying loads and identify any deviations from normal functioning using factors such as sound emissions, vibrations, and variations in temperature. The findings of this search can reveal the enhancement of predictive maintenance strategies to enhance the reliability and performance of rotating machinery.

Bearing faults are a common cause of machinery breakdowns and can lead to equipment damage and high maintenance costs. To prevent such issues, early diagnosis and detection of faults are crucial. In the past few years, several methods have been introduced to detect bearing faults by analyzing vibration signals. One such approach is the utilization of statistical characteristics obtained from the vibratory signals. In this regard, an innovative way of diagnosing issues in bearings by utilizing statistical characteristics extracted from vibratory signals was submitted in [1]. The authors demonstrated the effectiveness of their approach in identifying healthy and faulty bearings through experimental data and compared their approach with existing methods, showing promising results. Their research suggests that their approach can be employed in the condition monitoring of industrial machinery to prevent equipment failure and reduce maintenance costs.

Ball bearings serve as a fundamental component of machinery that revolves, and their faults can lead to equipment breakdown and high maintenance costs. Machine learning techniques have recently been suggested for the accurate diagnosis of ball bearing faults. [2] investigated for ball bearing fault diagnosis by employing vibration signals and statistical characteristics obtained from the bearings.

Their research showed that the SVM algorithm outperformed the other methods, demonstrating high accuracy in fault diagnosis. The study concludes that machine learning methods can be effective in diagnosing ball bearing problems and are appropriate in condition monitoring of industrial machinery.

The suggested technique was assessed on a set of vibration signals obtained from bearings having various types of defects and juxtaposed with conventional machine learning methods [3].

2. PROPOSED METHODOLOGY
2.1 SVM
A category of supervised machine learning approach is Support Vector Machine (SVM), which operates on the principles of statistical learning theory. The present research highlights that Support Vector Machine (SVM) is a potent approach to perform classification and regression tasks, particularly in scenarios where the data available is inadequate, as in the context of fault diagnosis. This research proposes the utilization of Support Vector Machines (SVMs) for several reasons, such as their capability to provide high precision, overcome overfitting, and work well with limited data for fault identification. Additionally, SVMs possess the potential to perform efficiently in less time and are suitable for handling high-dimensional data without overfitting due to their automatic regularization mechanism. Moreover, SVMs

demonstrate a promising approach for various applications as they can effectively operate with small datasets.

2.2 SVM Kernels

The kernel function is a powerful mathematical tool that enables Support Vector Machines (SVM) to perform complex, nonlinear classifications of one-dimensional data in higher-dimensional spaces.

- **Linear Kernel Function-** The SVM utilizes the linear kernel function to transform the input data to a higher-dimensional space, which is a straightforward yet effective approach, making it easier to separate and classify. Its effectiveness lies in its ability to capture linear relationships between variables.
- **Polynomial-** It is also used to Convert data into a space with a higher number of dimensions, its uniqueness lies in its capacity to capture intricate connections between data points, even when they are not linearly separable.
- **Radial-** It is used for classification and clustering. It measures the similarity between two points based on their distance between each other.
- **Quadratic-** Quadratic Kernel is a type of radial basis function used for non-linear classification. It maps data points into a higher dimensional space to make them separable.

2.3 KNN

KNN, also known as K-Nearest-Neighbor, is a machine learning algorithm that falls under the category of supervised learning techniques. This algorithm is used for making highly accurate predictions. However, KNN is used for classification and regression problems.

Classification problem: - When the targeted column or dependent variable has a fixed number of categories, then it comes under classification. We use two approaches to solve the Classification problem i.e., Euclidean Distance and Manhattan distance.

Regression problem: - Whenever the targeted column or dependent variable is continuous in nature, it becomes a regression problem. In this, we take the average distance of the data points by choosing the k value, where k is a hyperparameter and its value is not fixed.

2.4 Decision Tree

A popular machine-learning method for its ability to provide transparency, flexibility, and feature selection capabilities is the decision tree. Unlike other machine-learning methods that are often seen as “black boxes,” decision trees generate a tree-like structure that is intuitive and easy to interpret. This characteristic is especially useful in industries like healthcare, law, and finance where interpretability is essential.

One of the key strengths of the decision tree is its versatility in handling both categorical and numerical data. This makes them highly adaptable to a wide range of problems, from predicting customer churn to detecting fraudulent transactions. Moreover, decision trees are highly robust to missing data and can handle noisy data, making them ideal for real-world applications where data quality is not always perfect.

Another major advantage of the decision tree is its feature selection capabilities. By identifying the most relevant feature in the dataset, the decision tree can help researchers reduce the dimensionality of the data, leading to improved model accuracy and faster training times. This feature is particularly useful in domains where the cost of data acquisition or storage is high.

3. EXPERIMENTAL ANALYSIS

3.1 Dataset Description

This study involves the utilization of a dataset comprising three stages of bearing faults, with the inclusion of a healthy bearing in the first stage for comparative analysis. The dataset comprises data at different loads, ranging from 0k to 4k, for every stage of bearing faults. Additionally, the distribution of data within each stage is not uniform. Three phases make up the dataset, where the first stage represents a healthy bearing, the second stage includes data of bearings with certain degrees of damage, and the third stage comprises data of bearings that have undergone more severe damage (see Fig.1).

- **First Stage**
  The initial stage of the dataset corresponds to the healthy bearing stage, with the inclusion of readings at various loads of 0k, 1k, 2k, 3k, and 4k.
- **Second Stage**
  The second stage of the dataset encompasses data of bearings that have undergone some degree of damage, with recordings available at different loads of 0k, 1k, 2k, 3k, and 4k, depicting the damage to the bearings.

BEARING DATA																
S.No	UNIT	1st Stage (new bearing)					2nd Stage (some amount of damaged bearing)					3rd stage (extreme amount of damaged bearing)				
		0k	1k	2k	3k	4k	0k	1k	2k	3k	4k	0k	1k	2k	3k	4k
1	0	-0.0003865	0.000312	8.26E-06	0.001426	0.00091	-1.41722	-0.98888	-10.5582	-6.79987	-2.62613	-10.5582	-0.92308	-1.12564	-0.25735	-0.55151
2	0.024	-0.0003865	0.000312	-0.00051	0.001426	0.00091	-0.63279	-1.88169	0.369681	-7.26046	-1.48947	0.750285	-0.82245	-0.69085	-0.4612	-0.45862
3	0.048	-0.0003865	-0.0002	8.26E-06	0.001426	0.001426	-0.21993	-1.36304	1.197979	-5.62967	-1.25466	0.976067	-0.25864	-0.75665	-0.79406	-0.83535
4	0.072	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-0.68827	-1.16693	-1.06887	-3.12543	-1.85976	-10.0873	-1.19273	-0.47281	-0.9992	-0.85341
5	0.096	-0.0003865	-0.0002	8.26E-06	0.00091	0.00091	-1.24692	-1.10758	-10.5582	-2.04038	-2.20682	-10.5582	-1.62494	-0.39669	-0.40443	0.127127
6	0.12	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-1.45335	-0.97985	-10.5582	-6.58957	-2.63387	-10.5582	-1.14886	-1.03404	-0.18768	0.103903
7	0.144	-0.0003865	0.000312	-0.00051	0.001426	0.00091	-1.00953	-1.75138	0.2226	-7.32497	-1.33465	0.536115	-0.67924	-0.89212	-0.51668	-0.73472
8	0.168	-0.0003865	-0.0002	-0.00051	0.001426	0.00091	-0.71665	-1.45722	1.043157	-5.6	-1.30369	1.154113	-0.80052	-0.65085	-0.85083	-0.62247
9	0.192	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-0.85857	-1.25853	0.004559	-3.1422	-1.50238	-10.5582	-0.64569	-0.51409	-0.68311	-0.38766
10	0.216	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-1.20563	-0.51668	-10.5582	-2.89577	-2.32422	-10.5582	-0.83793	-0.64182	-0.31541	-0.19155
11	0.24	-0.0003865	-0.0002	8.26E-06	0.00091	0.001426	-1.13467	-1.0779	-10.5582	-5.81933	-2.53452	-10.5582	-0.91405	-1.02888	0.087131	0.058747
12	0.264	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-0.76568	-1.07145	0.199377	-7.29917	-1.38239	0.577401	-0.61215	-1.16693	-0.32315	-0.6973
13	0.288	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-0.65601	-1.47012	1.039286	-6.41926	-0.30638	1.240555	-1.06758	-0.5399	-0.8947	-0.42378
14	0.312	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-0.76955	-0.59538	-0.69601	-3.21187	-1.71655	-10.5582	-1.23144	0.169703	-0.50248	-0.44055
15	0.336	-0.0003865	-0.0002	8.26E-06	0.001426	0.00091	-0.87277	-0.9605	-10.5582	-2.17456	-2.70096	-10.5582	-1.68042	-0.80568	-0.23542	-0.01737
16	0.36	-0.0003865	-0.0002	8.26E-06	0.00091	0.00091	-0.75149	-1.29853	-10.5582	-6.39217	-2.83901	-9.70149	-0.77987	-1.1308	-0.06382	-0.42636
17	0.384	-0.0003865	-0.0002	8.26E-06	0.00091	0.00091	-0.55667	-1.06371	0.520633	-8.34938	-2.00039	0.248404	-0.93598	-0.87793	-0.05608	-0.59022
18	0.408	-0.0003865	-0.0002	0.000524	0.001426	0.00091	-0.80181	-1.49463	1.465047	-5.62451	-0.3954	0.99413	-0.92953	-0.36056	-0.80181	-0.57344

Fig. 1. Overall Dataset Utilized

- **Third stage**
  The third stage of the dataset includes data of bearings that have undergone an additional level of damage beyond the second stage. Recordings of these bearings are available at various loads of 0k, 1k, 2k, 3k, and 4k, providing insights into the nature and extent of the extra damage sustained by the bearings

Table 1 Statistical Features of First Stage at Different load

Statistical feature	1st stage (0k)	1st stage (1k)	1st stage (2k)	1st stage (3k)	1st stage (4k)
Mean	-0.00128	0.00039	-0.00020	0.001239	0.00081
Variance	3.155e-07	8.365e-08	6.723e-08	7.874e-08	6.109e-08
Root mean square	0.001403	0.000488	0.000329	0.001270	0.000867
Skewness	0.0502	0.0006	-0.2728	-0.2990	-0.6284
Standard deviation	0.000562	0.000289	0.000259	0.000281	0.000247
Kurtosis	2.1385	2.3659	1.39714	4.4217	4.3355

Table 2 Accuracy with distinct Approaches

Statistical feature	1st stage (0k)	1st stage (1k)	1st stage (2k)	1st stage (3k)	1st stage (4k)
Mean	-4.87233	-4.718315	-4.569394	-1.771385	-0.737018
Variance	30.364	29.991	4.315	0.648	0.318
Root mean square	7.355545	7.228699	5.019388	1.945778	0.928224
Skewness	0.228	-0.03344	0.8618	-0.78598	-6.47081
Standard deviation	5.511	5.477	2.077	0.805	0.564
Kurtosis	1.0700	1.0852	2.4529	3.77	57.9134

3.2 Result

The table 2 represents the precision of diverse machine learning techniques on the bearing dataset across all three stages of degradation. The models were trained to categorize the condition of the bearing based on the given data. The SVM model with a linear kernel function achieved the highest accuracy of 98.5%, followed closely by the models with quadratic and ridge kernels, achieving 98.3% and 98.1%, respectively. The other models like KNN, Random Forest, Linear Regression, XGB Regressor, Adaboost Regressor, and Random Forest also achieved good accuracy ranging from 97.2% to 98.2%. However, the Catboosting Regressor model achieved the lowest accuracy of 94.6%.

Overall, the high accuracy rates of these models demonstrate their effectiveness in predicting the health status of the bearing across all three stages of degradation (see Fig.2).

The models can be used for developing predictive maintenance strategies to detect and mitigate any potential damage in the early stages.

4. CONCLUSION

In conclusion, the analysis of the bearing dataset reveals significant changes in data characteristics as the bearing undergoes degradation. Statistical features indicate tighter clustering in the healthy stage and increased spread and skewness in damaged stages, with severe damage showing even greater dispersion. SVM models exhibit high accuracy for classification, particularly with complex datasets, while different kernel functions are effective for linear and nonlinear separation. KNN is computationally expensive but performs well, while decision trees have lower accuracy but Random Forest improves with ensemble learning. Linear regression is suitable for regression tasks. SVM with a linear kernel achieves the highest accuracy overall. These findings can inform monitoring and predictive maintenance strategies for bearings, saving costs by identifying potential failures in advance.

5. REFERENCES

[1] Altaf, M., Akram, T., Khan, M.A., Iqbal, M., Ch, M.M.I. and Hsu, C.H., 2022. A new statistical features based approach for bearing fault diagnosis using vibration signals. *Sensors*, *22*(5), p.2012.

[2] Kankar, P.K., Sharma, S.C. and Harsha, S.P., 2011. Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with applications*, *38*(3), pp.1876-1886.

[3] He, M. and He, D., 2017. Deep learning based approach for bearing fault diagnosis. *IEEE Transactions on Industry Applications*, *53*(3), pp.3057-3065.

[4] Han, T., Zhang, L., Yin, Z. and Tan, A.C., 2021. Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement*, *177*, p.109022.

[5] Zhang, S., Zhang, S., Wang, B. and Habetler, T.G., 2020. Deep learning algorithms for bearing fault diagnostics—A comprehensive review. *IEEE Access*, *8*, pp.29857-29881.

[6] Cristianini, N., & Shawe-Taylor, N. J. (2000). An introduction to support vector machines. Cambridge: Cambridge University Press.

[7] Widodo, A., & Yang, B.-S. (2007). Review on support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21, 2560–2574.

[8] Tyagi, C.S., 2008. A comparative study of SVM classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing. *International Journal of Mechanical and Mechatronics Engineering*, *2*(7), pp.904-912.

[9] Vakharia, V., Gupta, V.K. and Kankar, P.K., 2016. A comparison of feature ranking techniques for fault diagnosis of ball bearing. *Soft Computing*, *20*, pp.1601-1619.

[10] Han, T., Zhang, L., Yin, Z. and Tan, A.C., 2021. Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement*, *177*, p.109022.

[11] Sharma, A., Jigyasu, R., Mathew, L. and Chatterji, S., 2018, May. BearinSep. 2019.