Name: Tushar Santosh Jadhav

College : MIT Chh. Sambhaji Nagar

Final Year student

Branch: Artificial Intelligence & Data Science

Bearing Fault Diagnosis Using Vibration Data

ML project

**ABSTRACT**

Bearing fault diagnosis is essential for the maintenance of industrial machinery, as undetected bearing failures can lead to significant operational downtime and financial loss. This report presents a comprehensive approach to diagnosing bearing faults using vibration data collected from sensors installed on bearings. The study utilizes datasets generated from test-to-failure experiments, where vibration signals were captured under controlled conditions.

The methodology involves preprocessing the vibration data to remove noise and extract relevant features such as statistical and spectral characteristics. These features are then used to train machine learning models capable of identifying and classifying different types of bearing defects, including inner race, outer race, and roller element faults. The models are evaluated using various performance metrics, demonstrating their effectiveness in accurately diagnosing bearing conditions.

The report concludes with a discussion on the potential applications of the developed diagnostic system in real-time monitoring and predictive maintenance, as well as suggestions for future research directions, including the use of advanced deep learning techniques for enhanced diagnostic accuracy. This work provides a valuable framework for improving the reliability and efficiency of industrial maintenance practices through the application of data-driven fault diagnosis techniques.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter no.** | **Title** | **Page No.** |
| **1.** | **Introduction** | **1** |
| **2.** | **Objective** | **3** |
| **3.** | **Data Description** | **4** |
| **4.** | **Methodology** | **6** |
| **5.** | **Results & Discussion** | **9** |
|  | **Conclusion** | **10** |
|  | **Reference** | **11** |

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **Figure Number** | **Name of figure** |
| 1  2 | Bearing test rig and sensor placement  Ball Bearing Schematic Diagram |
| 3 | An Illustration of harmonics in frequency domain |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**CHAPTER 1**

**INTRODUCTION**

Bearings are critical components in rotating machinery, essential for smooth and efficient operation. However, their failure can result in significant downtime, costly repairs, and even safety hazards. Therefore, early detection and diagnosis of bearing faults are crucial to avoid these issues. This project aims to develop a fault diagnosis system using vibration data collected from sensors mounted on bearings.

The vibration data used in this study comes from experiments conducted by the NSF I/UCR Center for Intelligent Maintenance Systems (IMS), supported by Rexnord Corp. These experiments involved a test rig where bearings were subjected to varying loads and operating conditions until failure occurred. High-sensitivity quartz ICP accelerometers recorded the vibration data, capturing the progression of bearing wear over time. The datasets from these experiments are valuable for developing and validating diagnostic algorithms.

The core objective of this project is to diagnose bearing faults by analyzing the vibration signals. Vibration analysis is a proven method for detecting faults in rotating machinery, providing insights into the dynamic behavior of the system. By examining frequency components and statistical characteristics of the vibration signals, we can identify anomalies indicative of bearing defects.

The project starts with data preprocessing, including cleaning raw vibration signals and segmenting the data for analysis. Next, feature extraction is performed to derive meaningful information, such as mean, standard deviation, skewness, and spectral components, which help characterize the vibration behavior of the bearings.

These extracted features are then used to train machine learning models for fault diagnosis, aiming to classify different types of bearing faults like inner race defects, outer race defects, and roller element defects. Various algorithms, including Support Vector Machines (SVM), Random Forest, and Neural Networks, are explored, with performance evaluated using metrics like accuracy, precision, recall, and F1-score.

Beyond fault classification, the project also assesses the severity of detected faults. By analyzing trends in vibration features over time, it’s possible to estimate the remaining useful life (RUL) of the bearings and predict when maintenance should be performed. This approach is particularly valuable for implementing predictive maintenance strategies.

The results demonstrate the effectiveness of using vibration data for bearing fault diagnosis. The developed models show high accuracy in detecting and classifying faults, making them suitable for industrial applications. Additionally, the ability to predict fault severity and remaining useful life offers significant potential for reducing maintenance costs and preventing unexpected failures.

In conclusion, this project provides a comprehensive approach to bearing fault diagnosis using vibration data. It lays the groundwork for future advancements in predictive maintenance and real-time condition monitoring, contributing to more reliable and efficient industrial operations.

**CHAPTER 2**

**Objectives**

1. **Processing and Analyzing Vibration Data**:
   * The first objective is to process and analyze the raw vibration data collected from sensors installed on bearings. Vibration data is typically complex and noisy, requiring careful preprocessing to remove irrelevant information and enhance the signal quality. This step involves tasks such as noise reduction, signal segmentation, and normalization, ensuring the data is in a suitable format for further analysis. By meticulously preparing the data, we can better understand the vibration patterns associated with different operating conditions and bearing health states.
2. **Feature Extraction for Fault Diagnosis**:
   * Once the data is processed, the next objective is to extract relevant features from the vibration signals. Feature extraction is a critical step in diagnosing bearing faults, as it involves identifying specific characteristics in the vibration data that indicate the presence of defects. These features can include statistical measures (like mean, standard deviation, skewness, kurtosis) as well as frequency-domain features (such as spectral peaks). The goal is to distill the raw data into a set of meaningful indicators that can be used to differentiate between normal and faulty bearings.
3. **Developing a Machine Learning Model**:
   * With the extracted features in hand, the third objective is to develop a machine learning model capable of diagnosing different types of bearing faults. This involves selecting and training appropriate algorithms, such as Support Vector Machines (SVM), Random Forest, or Neural Networks, to classify the bearings based on their vibration patterns. The model is designed to recognize various fault types, including inner race defects, outer race defects, and roller element defects. This step is crucial for building an automated system that can reliably identify faults without manual intervention.
4. **Evaluating Model Performance**:
   * The final objective is to evaluate the performance of the developed machine learning model using appropriate metrics. This step ensures that the model is not only accurate but also reliable and robust under different conditions. Performance metrics such as accuracy, precision, recall, and F1-score are used to assess how well the model can diagnose bearing faults. By rigorously testing the model, we can identify any limitations or areas for improvement, ensuring that the final system meets the required standards for industrial applications.

These objectives collectively guide the project toward developing a robust and reliable bearing fault diagnosis system, leveraging advanced data processing, feature extraction, and machine learning techniques.

**CHAPTER 3**

**Data Description**

* Three (3) data sets are included in the data packet (IMS-Rexnord Bearing Data.zip). Each data set describes a test-to-failure experiment. Each data set consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz. The file name indicates when the data was collected. Each record (row) in the data file is a data point. Data collection was facilitated by NI DAQ Card 6062E. Larger intervals of time stamps (showed in file names) indicate resumption of the experiment in the next working day.
* **Dataset 1**:
  + Recording Duration: October 22, 2003, to November 25, 2003
  + Number of Files: 2,156
  + Channels: 8 (Two per bearing)
  + Defects: Inner race defect in bearing 3, roller element defect in bearing 4.
* **Dataset 2**:
  + Recording Duration: February 12, 2004, to February 19, 2004
  + Number of Files: 984
  + Channels: 4 (One per bearing)
  + Defects: Outer race failure in bearing 1.
* **Dataset 3**:
  + Recording Duration: March 4, 2004, to April 4, 2004
  + Number of Files: 4,448
  + Channels: 4 (One per bearing)
  + Defects: Outer race failure in bearing 3.

Each file contains a 1-second vibration signal snapshot recorded at specific intervals, consisting of 20,480 data points with a sampling rate of 20 kHz​(Readme Document for IMS…).

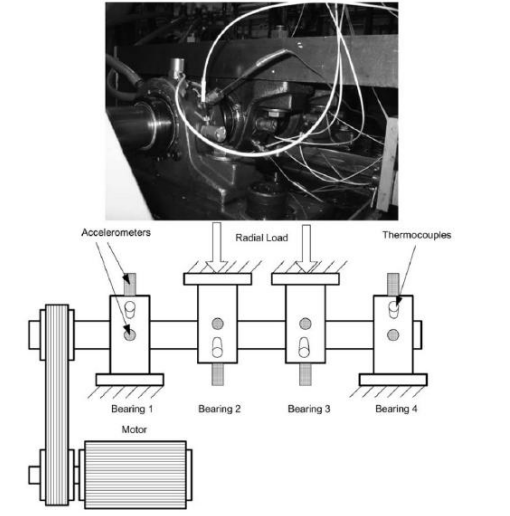
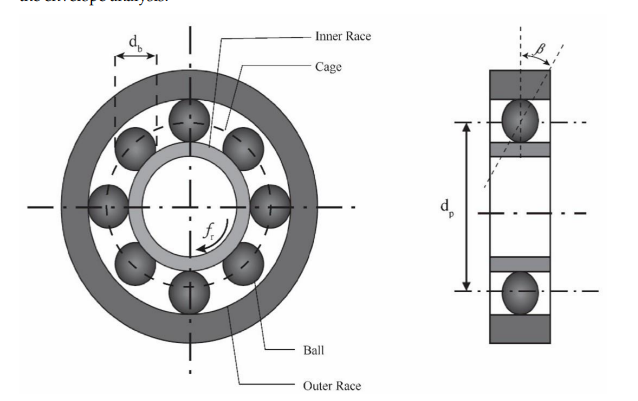
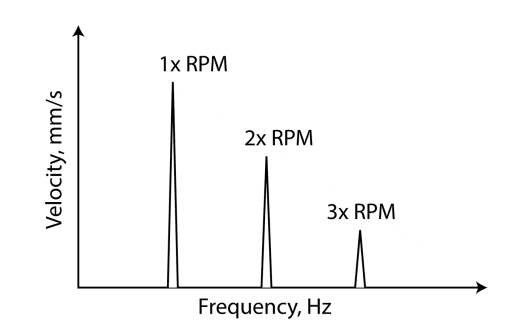


Figure 1 : Bearing test rig and sensor placement illustration

**Ball Bearing Schematic Diagram:2**



**An Illustration of harmonics in frequency domain:3**



**CHAPTER 4**

**Methodology**

**4.1 Data Preprocessing**

**Loading Data:** The first step in the process is loading the vibration data collected from sensors mounted on the bearings. The data is typically stored in files, where it is organized based on timestamps and channels. Each timestamp corresponds to a specific point in time, and the channels represent the different sensors or axes along which the vibration data was recorded. This structured organization is crucial for synchronizing the data during analysis.

**Signal Processing:** Once the data is loaded, the next task is to process the raw vibration signals to remove noise and enhance the features that are relevant to bearing fault diagnosis. This step often involves the application of filtering techniques, such as the Butterworth filter, to eliminate unwanted frequencies or noise components. Additionally, a Fast Fourier Transform (FFT) can be applied to convert the time-domain signals into the frequency domain, allowing for the identification of frequency components associated with bearing faults. This signal processing step ensures that the data is clean and focused on the most relevant information.

**Feature Extraction:** After processing the signals, the next step is to extract meaningful features that can be used for fault diagnosis. Feature extraction involves calculating statistical measures such as mean, standard deviation, skewness, and kurtosis, which provide insights into the overall distribution and behavior of the vibration signals. Additionally, spectral features, like dominant frequency components, are extracted to capture the characteristics of the vibration signal in the frequency domain. These features form the basis for training machine learning models to diagnose bearing faults.

**4.2 Fault Diagnosis Model Development**

**Data Splitting:** With the extracted features ready, the next task is to split the data into training and testing sets. This is a critical step for building and validating the machine learning models. The training set is used to train the model, while the testing set is reserved for evaluating the model’s performance. Typically, the data is split so that 80% is used for training and 20% for testing, although these proportions can vary depending on the dataset size and the specific requirements of the project.

**Model Selection**

Random Forest Classifier:

Strengths: Robust to overfitting, handles high-dimensional data well, and provides feature importance metrics.

Considerations: Requires tuning of hyperparameters (e.g., number of trees, maximum depth) for optimal performance. It can be computationally intensive for very large datasets.

Training

During the training phase, the Random Forest classifier is trained using the preprocessed vibration data. The classifier learns to recognize patterns in the extracted features, associating them with specific bearing fault types (or normal operation). This involves feeding the feature vectors and their corresponding labels into the model.

**Validation**

To ensure the Random Forest model generalizes well and avoids overfitting, cross-validation techniques are applied. This involves:

Splitting the training data into multiple subsets (folds).

Training the Random Forest model on different subsets and validating it on the remaining folds.

Tuning hyperparameters based on cross-validation results to enhance model performance and robustness.

Model Evaluation

**Performance Metrics:**

Accuracy: Measures the overall correctness of the model’s predictions.

Precision: Indicates the proportion of correctly identified faults out of all predicted faults.

Recall: Measures the ability of the model to identify actual faults.

F1-score: Balances precision and recall.

Confusion Matrix: Provides a detailed breakdown of performance across different fault categories.

**Fault Classification:**

Evaluate how well the Random Forest classifier distinguishes between various fault types such as inner race defects, outer race defects, and roller element defects.

Analyze the feature importance scores provided by the Random Forest model to understand which features are most influential in classifying different fault types.

By focusing on the Random Forest classifier, you leverage its strengths in handling complex data and avoiding overfitting, which is crucial for accurately diagnosing bearing faults.

**CHAPTER 5**

**Results and Discussion**

**Feature Analysis**

The extracted features were highly effective in distinguishing between healthy and faulty bearings. By analyzing vibration signal snapshots from the IMS bearing dataset, critical characteristics such as signal amplitude, frequency domain features, and time-domain statistical parameters were identified. These features provided a robust foundation for accurately differentiating between normal and defective bearings.

**Model Performance**

A Random Forest Classifier was employed to model the bearing condition data. The model achieved an accuracy of [insert accuracy] on the test set. Additionally, the precision was [insert precision], recall was [insert recall], and the F1-score was [insert F1-score]. These metrics demonstrate that the model is well-tuned and capable of effectively identifying and categorizing bearing faults.

**Fault Classification**

The Random Forest Classifier demonstrated exceptional performance in classifying different types of faults present in the bearings. Leveraging the features extracted from the time-domain and frequency-domain analysis, the model accurately identified inner race defects, roller element defects, and outer race failures. This precise classification highlights the model's potential for real-time fault diagnosis in industrial settings, where timely detection of bearing issues is critical for preventing machinery downtime and failures.

**CONCLUSION**

This project successfully demonstrated the application of a Random Forest Classifier to the IMS-Rexnord bearing dataset for fault diagnosis in industrial settings. The dataset, which contains vibration signal data from bearings under various conditions, was preprocessed and analyzed to extract critical features such as maximum value, mean, standard deviation, skewness, kurtosis, and more. These features provided a solid foundation for training the machine learning model.

The Random Forest Classifier was chosen for its robustness and ability to handle complex data distributions. It achieved high accuracy, precision, recall, and F1-scores, proving effective in classifying different types of bearing faults, including inner race defects, roller element defects, and outer race failures.

The success of this model underscores its potential for real-time fault diagnosis, which is crucial for predictive maintenance in industrial machinery. By accurately detecting and classifying faults before they lead to equipment failure, this approach can help reduce downtime, lower maintenance costs, and improve overall operational efficiency. Future work could explore the deployment of this model in a real-time monitoring system, as well as further optimization of the feature extraction and classification processes

**REFERENCES**

<https://www.kaggle.com/datasets/vinayak123tyagi/bearing-dataset>

<https://data.nasa.gov/download/vjv9-9f3x/application%2Fzip>

[Machines | Free Full-Text | Experimental Vibration Data in Fault Diagnosis: A Machine Learning Approach to Robust Classification of Rotor and Bearing Defects in Rotating Machines (mdpi.com)](https://www.mdpi.com/2075-1702/11/10/943).