FEATURE EXTRACTION FROM VIBRATION SIGNATURE ACQUIRED FROM RAILROAD BEARING ONBOARD CONDITION MONITORING SENSOR MODULES

A Thesis

by

KEVIN QUAYE

Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE IN ENGINEERING

Major Subject: Mechanical Engineering

The University of Texas Rio Grande Valley
August 2024

FEATURE EXTRACTION FROM VIBRATION SIGNATURE ACQUIRED FROM RAILROAD BEARING ONBOARD CONDITION MONITORING SENSOR MODULES

A Thesis
by
KEVIN QUAYE

COMMITTEE MEMBERS

Dr. Constantine Tarawneh
Chair of Committee

Dr. Ping Xu
Co-Chair of Committee

Dr. Arturo Fuentes
Committee Member

Dr. Heinrich Foltz
Committee Member

Dr. Dimah Dera
Committee Member

August 2024

ABSTRACT

Quaye, Kevin, Feature Extraction from Vibration Signature Acquired from Railroad Bearing Onboard Condition Monitoring Sensor Modules. Master of Science in Engineering (MSE), August 2024, 55 pp., 8 tables, 27 figures, 25 references.

The University Transportation Center for Railway Safety (UTCRS) has designed an onboard monitoring system that tracks vibration waveforms over time, to obtain a direct and more accurate indicator of bearing health. The data collected by these sensors is used for vibrational analyses of the bearings. The speed of the bearing is a fundamental parameter needed to carry out the analysis. GPS data can be used to determine bearing speed if available; however, due to size, cost and power constraints GPS is typically not available at the sensor location. This means that analysis must be delayed until the data is uploaded to a location where it can be matched with GPS data, unless the speed is determined immediately at the bearing. It is proposed to solve this problem with the introduction of Machine Learning (ML) algorithm that would extract features, such as the speed, from the vibration data that is already being acquired by the onboard sensing module. The PSD plots contain embedded signatures corresponding to the speed. Rapid extraction of this data would allow for real time analysis of bearing condition as the train is moving, which could be sent to a cloud accessible by train dispatchers and railcar owners for assessment of bearing health and scheduling of proactive maintenance before defects reach a critical size.

DEDICATION

This thesis is dedicated to my late father, Thomas Quaye and to my lovely Quaye family and friends in the valley whom I met in my years at UTRGV. My father, Thomas, sacrificed a lot for me to have the opportunity to pursue a career in Mechanical engineering. My family for always cheering me on to keep going and never give up. I could not have made it without the support of my friends who were always there to provide a shoulder to lean on in difficult time

ACKNOWLEDGEMENTS

I will always be grateful to Dr. Constantine Tarawneh, my research advisor and chair of my thesis committee, for all his mentoring and advice. Thank you for giving me the opportunity to join your research team. You have always spurred me on to be the best engineer I could ever be. My thanks go to my thesis committee members: Dr. Ping Xu, Dr. Dimah Dera, Dr. Arturo Fuentes and Dr. Heinrich Foltz for their tutoring and assistance throughout this research and thesis writing. Their advice, input, and comments on my thesis helped me ensure the quality of my intellectual work.

Additionally, I would like to thank the University Transportation Center for Railway Safety (UTCRS) team for their contribution to my successful journey. Thank you to everyone who welcomed both past and present ones. It has been an honor and great privilege to work with and learn from each of you. Thank you, guys.

Finally, I would like to acknowledge the financial support provided by the University Transportation Center for Railway Safety (UTCRS) funded through USDOT Grant No. 69A3552348340.

TABLE OF CONTENT

Page

ABSTRACT iii
DEDICATION iv
ACKNOWLEDGEMENTS v
LIST OF TABLES viii
LIST OF FIGURES ix
CHAPTER I: INTRODUCTION AND BACKGROUND 1
1.1 Introduction to Tapered Roller Bearings 1
1.1.1 Classification of Defects and Failures 2
1.2 Wayside Detection Technologies 2
1.2.1 Trackside Acoustic Detector Systems (TADS™) 3
1.2.2 Wheel Impact Load Detectors (WILDs) 4
1.2.3 Hot Bearing Detectors (HBDs) 5
1.3 Studies of Derailment Incidents 5
1.3.1 Norfolk Southern Train 32N 5
1.3.2 Canadian Pacific Freight Train 220-24 6
1.4 Onboard Wireless Monitoring Technologies 7
1.4.1 UTCRS Wireless Onboard Sensor Module 8
1.4.2 HUM Industrial Technology Boomerang 9
1.5 Preliminary Work 9
1.6 Motivation 12
CHAPTER II: BACKGROUND OF MACHINE LEARNING AND PRACTICAL APPLICATIONS 14
2.1 Introduction to Machine Learning 14
2.2 Performance Metrics 16
2.2.1 Mean Absolute Error (MAE) 16
2.2.2 Mean Squared Error (MSE) 16

2.2.3 Root Mean Squared Error (RMSE) 17
2.2.4 Coefficient of Determination (R^2) 17
2.2.5 Cost Functions 18
2.3 Linear Regression 19
2.4 Introduction to Deep Learning 19
2.4.1 Recurrent Neural Networks (RNN) 20
2.5 Random Forest Model 21
2.6 XGBoost Regression Model 22
CHAPTER III: EXPERIMENTAL SETUP AND PROCEDURES 24
3.1 An Overview of the Laboratory 24
3.1.1 Four-Bearing Test Rig (4BT) 26
3.2 Overview of the Data 29
3.2.1 Data Recording 30
3.2.2 Data Analysis 30
CHAPTER IV: METHODOLOGY AND MODEL SPECIFICATIONS 35
4.1 Datasets 35
4.1.1 Feature Set 37
4.2 Random Forest Specifications 37
4.3 XGBoost Regression Specifications 38
4.4 Linear Regression Specifications 39
CHAPTER V: RESULTS AND DISCUSSION 41
5.1 Descriptive Analysis 41
5.2 Random Forest Regression Results 44
5.3 XGBoost Regression Results 47
5.4 Linear Regression Results 49
CHAPTER VI: CONCLUSIONS AND FUTURE WORK 51
REFERENCES 53
VITA… 55

LIST OF TABLES

Page

Table 1: Bearing Standards by AAR: Classes, Dimension, and Load Capacities 25

Table 2: Angular speeds of axle to equivalent track speed 25

Table 3: Dataset with input features (Frequency, PSD, Load) and target value (Speed) 36

Table 4: Experiments ID used for generating dataset 36

Table 5: Parameter distribution of Random Forest model 38

Table 6: Parameter distribution for XGBoost Regression model 39

Table 7: Performance metrics of Random Forest Models 46

Table 8: Performance metrics of XGBoost Models 49

LIST OF FIGURES

Page

Figure 1: Detailed view of a tapered roller bearing(Hernandez, 2020) 2

Figure 2: Trackside Acoustic Detector Systems (TADS™)(Stewart, Flynn, Marquis, & Sharma & Associates, 2019). 4

Figure 3:Wheel Impact Load Detectors (WILDs)(LBFoster, n.d.). 4

Figure 4: Diagram of a Hot Bearing Detector (HBD) (Stewart, Flynn, Marquis, & Sharma & Associates, 2019). 5

Figure 5: Canadian Pacific freight train 220-24 Derailment Location Map (Freight, 2011). 7

Figure 6: UTCRS Wireless Onboard Sensor Module (Cuanang, 2020). 8

Figure 7: A field mounted HUM Industrial Technology Boomerang (Cantu, 2021) 9

Figure 8: Defect Detection Algorithm (Montalvo, 2019) 11

Figure 9: Frequency Spectrum Plots (0-1000Hz) (a) A Defect-free bearing (b) Outer ring defect (c) Inner ring defect (d) A roller defect (Montalvo, 2019). 12

Figure 10: An unrolled recurrent neural network (Colah, 2015). 21

Figure 11: Random Forest Regression Model (Random Forest Regression in Python, 2023) 22

Figure 12; Gradient Boosting Decision Trees (NVIDIA, 2024) 23

Figure 13: Four-Bearing Test Rig (4BT) (Montalvo, 2019). 27

Figure 14: Top and Rear views of 4BT including sensor locations (Montalvo, 2019). 28

Figure 15: Bearing adapter showing accelerometer locations (Montalvo, 2019). 28

Figure 16: Placement of thermocouples on the railroad bearing.
(red dots represent the locations of the regular K-type thermocouples and
black dots represent the locations of the bayonet K-type thermocouples) 29

Figure 17: Vibration and Temperature analysis plots (Martinez, 2023). 31

Figure 18: Correlation Analysis at Level 3 for a Bearing Cup (outer ring) (Martinez, 2023). 33

Figure 19: Correlation Analysis at Level 3 of a Bearing Cone (inner ring) (Martinez, 2023). 34

Figure 20: Correlation heatmap of all variables in the dataset 42

Figure 21: Feature importance of features 43

Figure 22: Plot of the features 44

Figure 23: Plot of Predicted vs. Actual Speed values with all features 45

Figure 24: Plot of predicted vs. actual speed values without the load feature 46

Figure 25: XGBoost model predictions with all features 47

Figure 26: Plot of predicted vs. actual values of speed without the load feature 48

Figure 27: Predicted vs. Actual speed values of Linear Regression model 50

CHAPTER I

INTRODUCTION AND BACKGROUND

From 2013 to 2022, 1671 derailments have been reported by the Federal Railroad Administration (FRA). 8.2 % of which were due to journal bearing defects (Federal Railroad Administration, n.d.). The University Transportation Center for Railway Safety (UTCRS) designed an onboard monitoring system that tracks vibration waveforms over time to assess bearing health through three analysis levels. However, the speed of the bearing, a fundamental parameter for these analyses, is often acquired from GPS data, which is typically not available at the sensor location. To solve this issue, this work proposes to employ Artificial Intelligence (AI) and Machine Learning (ML) algorithms to extract the speed and other essential features from existing vibration data, which can eliminate the need for additional speed sensors. This innovation would enable real-time analysis of bearings while the train is in motion.

1.1 Introduction to Tapered Roller Bearings

Figure 1 shows a tapered roller bearing assembly subcomponents, i.e., the cup, spacer ring, cage, 23 rollers, a cone, the grease seal, and the wear ring. A tapered-roller bearing is made up of a cup with two cone assemblies, separated by a spacer ring. A bearing is secured with two seals which prevent foreign materials from contaminating the lubricant in the bearing assembly. The cup (outer ring) stays stationary when the train is in motion since it bears the load of the freight railcar.

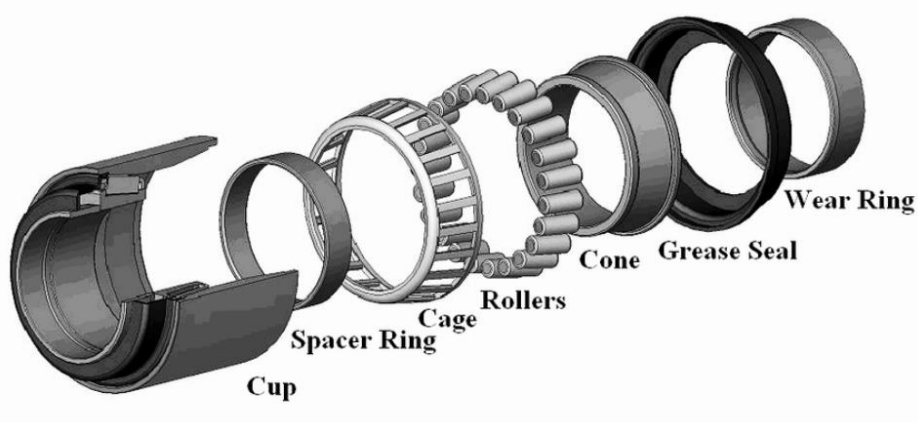

Figure 1: Detailed view of a tapered roller bearing(Hernandez, 2020)

As the freight train moves, the cup stays static while the inner rings (cones) rotate freely within the cup. Due to the rotation of the rollers in the cone, they receive the least amount of wear and stress, whereas the cup suffers the most stress and wear. (Hernandez, 2020).

1.1.1 Classification of Defects and Failures

Bearing defects fall into three general categories, namely; localized, distributed, and geometric defects. Localized defects include cracks, pits, and spalls on the rolling components of the bearing. A water-etch defect occurs when water enters the bearing, and the grease begins to degrade resulting in metal-to-metal friction which in turn results in the rapid wear of raceways. A water-etch is an example of a distributed defect. Geometric defects occur when there are inconsistencies during the manufacturing process that results in the bearing components being out of tolerance (Hernandez, 2020).

1.2 Wayside Detection Technologies

Wayside Detection technologies refer to the technologies that are used to monitor the condition of critical railcar components as the train moves on the tracks. These technologies are mounted on the sides of the track or directly on the rails. To fully assess the condition of a

railroad tapered roller bearing the temperature, load, and vibration profiles acquired from the bearing are all needed. The most widely used wayside technologies in North America that capture these parameters are the Hot Bearing Detectors (HBDs), Trackside Acoustic Detector Systems (TADS^{TM}), and the Wheel Impact Load Detectors (WILD).

Since their implementation, HBDs have reduced derailments, but they have also raised the percentage of false positives, or non-verified bearings. The primary drawback of HBDs is their highly reactive nature, which means that by the time the temperature reacts to the presence of a large flaw or defect, it might be too late to prevent a derailment. It takes approximately three miles for a train consisting 100 freight cars traveling at 50 miles per hour to come to a complete halt. This increases to about 6 miles for a train with 150 freight cars moving at the same speed (Tarawneh, Aranda, Hernandez, Crown, & Montalvo, 2020).

1.2.1 Trackside Acoustic Detector Systems (TADS^{TM})

Stationary devices known as Trackside Acoustic Detector Systems (TADS^{TM}) are used to track the acoustic emissions of tapered roller bearings. When tapered roller bearings exhibit a significant amount of spalling, an acoustic bearing detector known as TADS^{TM} will identify them as “growlers” based on their acoustic response. Train operators receive a signal for an instant train stoppage if a TADS^{TM} detects a growler.

Figure 2: Trackside Acoustic Detector Systems (TADS^{TM})(Stewart, Flynn, Marquis, & Sharma & Associates, 2019).

1.2.2 Wheel Impact Load Detectors (WILDs)

At fixed locations along railroad routes, Wheel Impact Load Detectors (WILD) measure the wheel-to-rail contact force generated by the wheels impacting the rails. Businesses like LBFoster use these detectors to stop excessive wheel-rail impacts, which can cause derailments. The WILD systems consist of a series of strain gauges that are installed on the rail tracks. An image of a live WILD site can be seen in Figure 3 (LBFoster, n.d.).

Figure 3:Wheel Impact Load Detectors (WILDs)(LBFoster, n.d.).

1.2.3 Hot Bearing Detectors (HBDs)

Using infrared sensors, HBDs monitor the temperature profiles of vital rail rolling stock components. These sensors specifically check to determine if any bearing that is in service has a temperature that is 94.4℃, or 170℉, higher than ambient. HBDs are often spaced 25 to 40 km (15 to 25 mi) on average, depending on the location of the lines and the volume of traffic, but the spacing can be as far as 64 km (40 mi) on rail tracks in rural areas. In North America, there are already more than 6,000 HBDs in use (Stewart, Flynn, Marquis, & Sharma & Associates, 2019). Figure 4 is a schematic diagram of HBD.

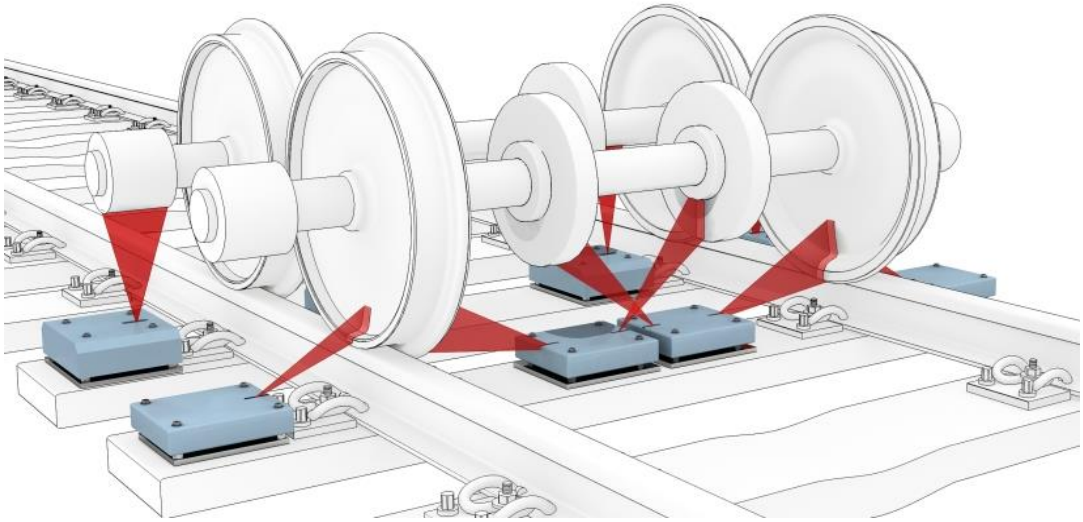

Figure 4: Diagram of a Hot Bearing Detector (HBD) (Stewart, Flynn, Marquis, & Sharma & Associates, 2019).

1.3 Studies of Derailment Incidents

1.3.1 Norfolk Southern Train 32N

On February 3, 2023, Norfolk Southern train 32N derailed, resulting in 35 cars being damaged. The incident happened in East Palestine, Ohio. The railway industry was unable to

stop this derailment because of the limitations of the existing condition monitoring systems and the types of problems they detect. The failing bearing's temperature at an HBD was 253°F above ambient before the train was signaled to stop. Most class I railroads utilize a 170°F above ambient alarm threshold for their HBDs to signal trains to stop for a visual inspection of the wheel-axle in question. The culprit bearing was photographed by multiple cameras along the route with flames for about 45 minutes. It was not flagged to halt until the HBD recorded 253°F above ambient. By that time, the train had derailed due to the faulty bearing that had caught fire for more than 20 miles. Since temperature is a reactive indicator, if the bearing temperature signals a problem, it can already be too late to mitigate its catastrophic failure. This is the main concern with the existing reliance on HBDs. This derailment served as additional evidence of the industry's need for more proactive condition monitoring technology (Board, 2023).

1.3.2 Canadian Pacific Freight Train 220-24

On January 26, 2011, Canadian Pacific Railway freight train 220-24 derailed, causing twenty cars to spill non-odorized liquified petroleum gas between Sudbury and Mactier, Canada. A roller bearing seized due to a spall on the inboard cup raceway. On the axle, the seized roller bearing overheated and finally burned off. Eight different HBDs failed to signal for the train to stop, which led to this derailment. The map indicates that the derailment occurred at a place where the distance between the HBDs was greater than 25 miles (40km).

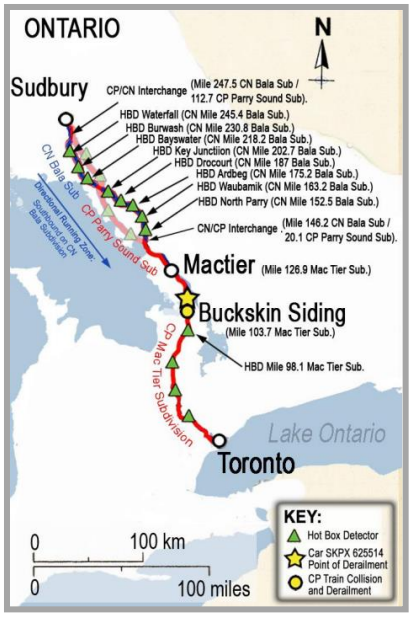

Figure 5: Canadian Pacific freight train 220-24 Derailment Location Map (Freight, 2011).

Canadian Pacific installed additional HBDs in these sites where the distances between detectors were as wide or larger in reaction to this derailment (Freight, 2011).

1.4 Onboard Wireless Monitoring Technologies

The newest development in bearing condition monitoring technology is the use of onboard sensors that can track load, vibration, temperature, and other data in real time. While the vibration profiles enable a more proactive detection of bearing defects, the temperature profiles enable detection of impending failure. Lastly, the load profile provides information about the stresses that are applied to the components. One or more of these characteristics are reflected in the onboard sensors that are now being field tested in a number of different class I railroads and short lines.

1.4.1 UTCRS Wireless Onboard Sensor Module

Over the past decade, the University Transportation Center for Railway Safety (UTCRS) has been developing a wireless sensor module, depicted in Figure 6, which can monitor the temperature and vibration signatures of tapered roller bearings in real time. This technology was licensed by HUM Industrial Technology, Inc., which has commercialized this technology and deployed it in several pilot field tests and freight rail operations.

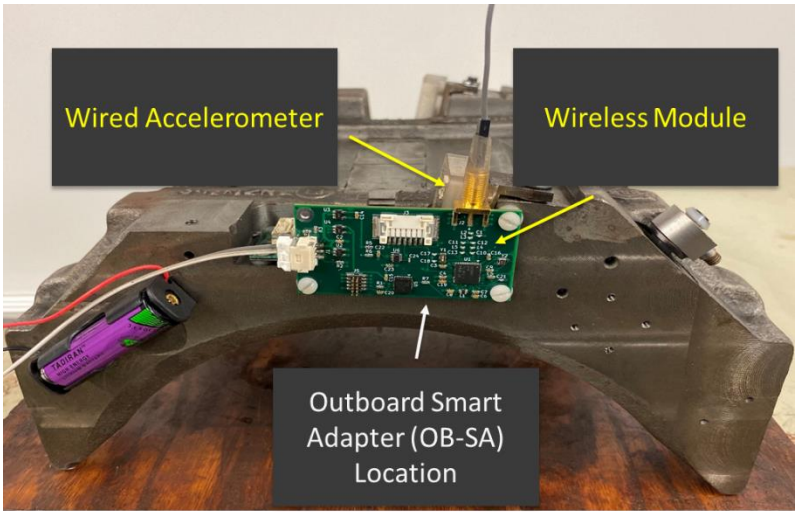

Figure 6: UTCRS Wireless Onboard Sensor Module (Cuanang, 2020).

One of the most notable advantages of this sensor, which was created with academic and research goals in mind, is that it attaches on the bearing adapter, meaning that the sensor module is reusable regardless of the state of the actual bearing. Every ten minutes, this wireless sensor module gathers data at a rate of 5,120Hz for 4 seconds. After the data is gathered, the sensor uses Bluetooth to send it to a central monitoring device (Cuanang, 2020).

1.4.2 HUM Industrial Technology Boomerang

A field application version of the UTCRS developed wireless sensor, manufactured by HUM Industrial Technology Inc., is called the Boomerang. A field-tested version of the HUM Boomerang is shown in Figure 7.

With the help of sensors installed on the bearing adapter, this condition monitoring device is able to monitor the temperature and vibration responses of the tapered roller bearing. The HUM Boomerang is capable of measuring the vibration levels within a bearing to within ±1 g and outputting bearing operating temperatures to within ±4°C.

Figure 7: A field mounted HUM Industrial Technology Boomerang (Cantu, 2021)

Every ten minutes, the sensors gather data at a rate of 5,200 Hertz for one to four seconds (as programmed) (Cantu, 2021).

1.5 Preliminary Work

The UTCRS team of researchers designed an onboard monitoring system that uses sensors to track the vibration waveforms over time on bearings, to obtain an accurate indicator of the health of bearings (Joseph, Constantine, & Arturo, 2018). The resulting Defect Detection Algorithm (DDA) is shown in Figure 8. The data collected by the onboard sensors is used for

analyses at three different levels. Level 1 analysis determines whether the bearings are defective or healthy based on the root mean square (RMS) of the vibration data being above or below the average threshold for defect-free bearings. Level 2 analyzes the power spectrum density (PSD) of the vibration signatures to determine the location of the defect whether it developed on the cup (outer ring) raceway, the cone (inner ring) raceway, or the rollers. Finally, Level 3 analysis determines the size of the defect using a library of previously measured defects.

The DDA uses frequency-domain analysis and relies on the rotational speed of the bearings to carry out the vibrational analysis. The MATLAB algorithm calculates the power spectral density (PSD) to determine whether the bearing component is faulty or not. Faulty bearings exhibit a power spike at the corresponding defect frequency and its harmonics, while a healthy bearing will exhibit PSD that is lower and not concentrated at the defect frequencies. The defects are classified by the tracking of the fundamental frequencies of the bearings based on the rotational speed of the bearings. Examples of each kind of bearing condition, along with the corresponding defect frequency and its harmonics are shown in Figure 9.

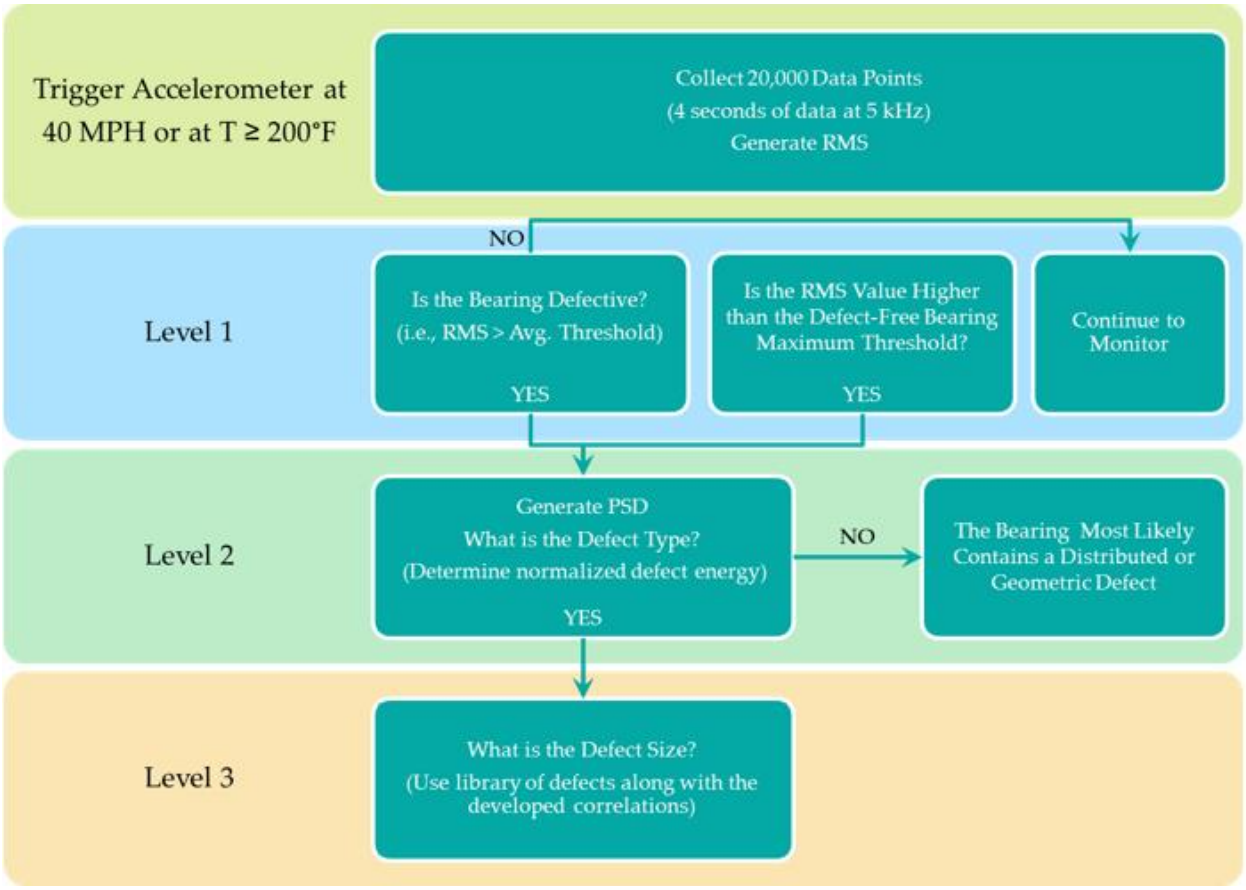

Figure 8: Defect Detection Algorithm (Montalvo, 2019)

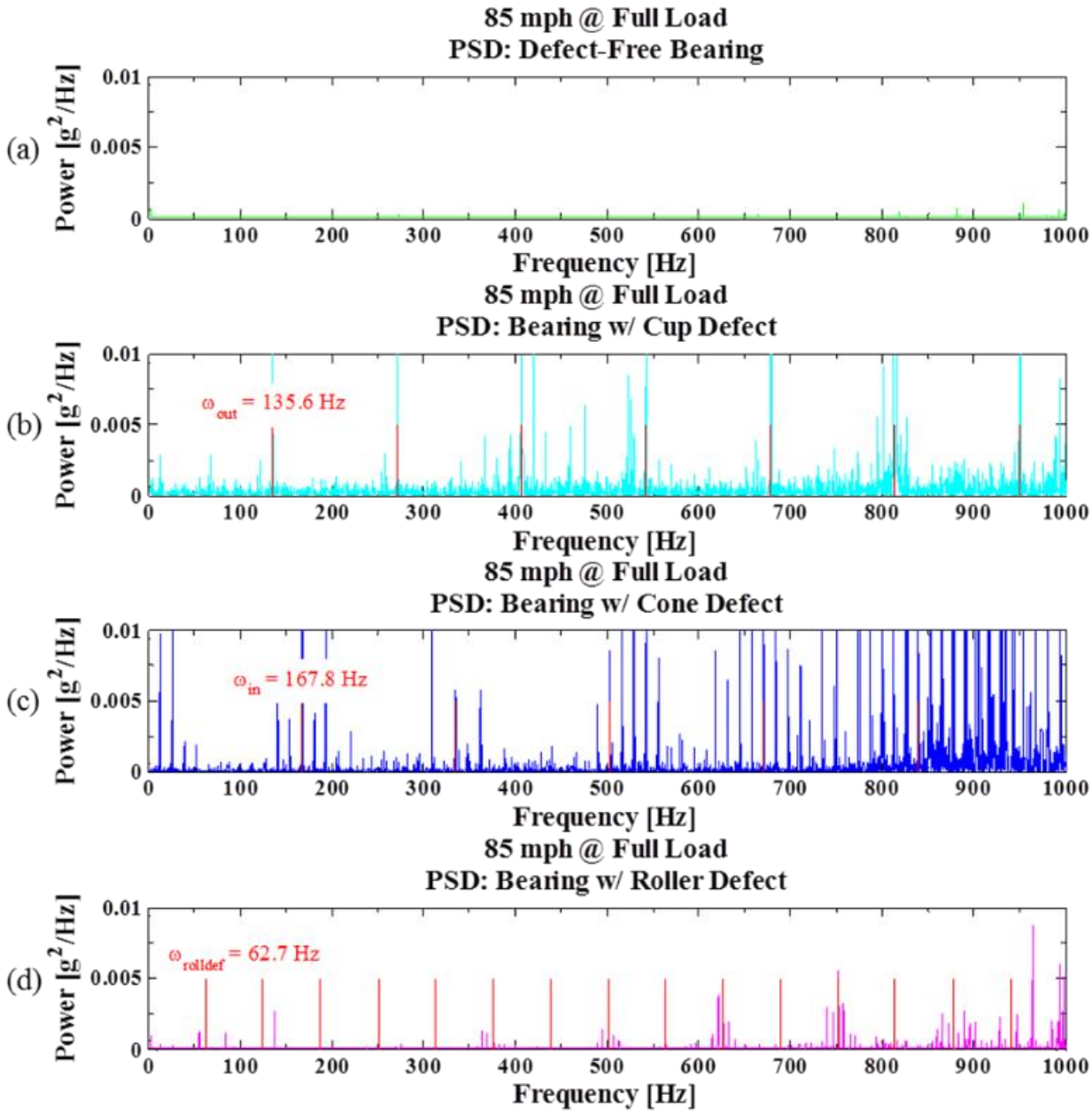

Figure 9: Frequency Spectrum Plots (0-1000Hz) (a) A Defect-free bearing (b) Outer ring defect (c) Inner ring defect (d) A roller defect (Montalvo, 2019).

1.6 Motivation

In recent years, the railway industry has experienced several catastrophic derailments predominantly due to the failure of bearings and wheel assemblies. Therefore, the safety, reliability, and performance of wheel-axle-bearings have become an issue of great concern in the

industry. Condition monitoring and fault detection of railcar bearings are complex and labor-intensive. Hot-Box (or Bearing) Detectors (HBDs) and the Trackside Acoustic Detection System (TADS^{TM}) are the currently utilized bearing condition monitoring systems. However, research has shown that these technologies are not proactive in detecting bearing health, i.e., these systems were designed to identify bearings that are near the end of their service life. To that end, the UTCRS developed onboard sensors were designed to identify bearing and wheel defects at their early stages of development, affording railcar owners and operators the opportunity to track their deterioration and schedule proactive maintenance cycles that minimize disruption to rail operations.

The developed algorithm uses the bearing's temperature and vibration signatures to assess its health. However, this algorithm requires the speed of the bearing to be known to carry out the vibrational analysis. GPS can be used to determine bearing speed if available, however due to cost, size, and power constraints, GPS data is not available at sensor location. This delays the analysis until the data is uploaded to a location where it can be matched with GPS data. The work presented in this thesis seeks to employ Machine Learning (ML) algorithms that would extract features such as the bearing rotational speed, from the vibration data that is already being acquired by the onboard sensing module. This would allow real time assessment of bearing health as the train is moving, which could be sent to a cloud accessible by train dispatchers and railcar owners for their analysis and scheduling of timely maintenance before defects approach imminent failure. Hence, a major outcome of this work is to mitigate train derailments and unplanned and costly field replacements and delays.

CHAPTER II

BACKGROUND OF MACHINE LEARNING AND PRACTICAL APPLICATIONS

2.1 Introduction to Machine Learning

Machine Learning (ML) is a subcategory of Artificial Intelligence (AI) which generally is defined as the capability of a machine to mimic intelligent human behavior. Systems built with AI are used to perform complicated tasks in a way that resembles how humans solve problems. ML uses an approach that allows computers to program themselves through experience. Machine learning begins with data (numbers, photos, or text) like photos of people, inventory records, time series data from sensors, or reports on sales. The data collected and prepared to be used is referred to as training data. The larger the dataset, the better the performance (Brown, 2021).

In addition, programmers have to select an ML model to use for a specific task. The model is supplied with training data to train itself to find patterns or make predictions. Over a period, the programmer can alter the model, this may include changing the parameters to give a more accurate result. Some of the data is excluded from training and used as evaluation data, to test how the ML model performs when it is introduced to a new database. The performance of the model on the evaluation data gives an idea of how the model will perform in the future with different sets of data (Brown, 2021).

ML systems can be descriptive, meaning that the system uses the data to explain what happened; predictive, meaning the system uses the data to predict what will happen; or prescriptive, meaning the system will use the data to make suggestions about what action to take. There are three subdivisions of machine learning: supervised, unsupervised and reinforcement learning. Labeled data sets are used to train supervised ML models, enabling them to improve in accuracy over time. For instance, an algorithm may be taught with human-labeled images of dogs and other objects to teach the computer how to recognize dog images on its own. The most popular kind of ML employed today is supervised learning. In unsupervised learning, an algorithm searches unlabeled data for patterns. Unsupervised machine learning is the ability to identify trends or patterns that humans are not consciously searching for. An unsupervised ML software, for instance, may examine online sales data and recognize various customer types making purchases. Reinforcement learning creates a reward system and uses trial and error to teach machines the optimal course of action. By informing the machine when it made the proper choice, reinforcement learning can be used to teach models to play games or teach autonomous cars to drive. This allows the machine to gradually learn what actions it should take (Brown, 2021).

This research uses supervised learning models such as Random Forest and XGBoost Regressor. These models learn the relationship between input and output. The inputs are referred to as features and are usually denoted ‘X variables’ and the output is commonly known as the target variable or ‘Y variable’. This research seeks to use a labeled data, thus, data that contains both the features and the target variables (Ali, 2022).

2.2 Performance Metrics

Performance metrics or evaluation metrics are essential in ML to assess the performance of models. Some common performance metrics used for regression analysis are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R^2 (Coefficient of Determination).

2.2.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a simple evaluation metric used to evaluate the accuracy of regression models. It is computed by finding the absolute difference between the predicted values and actual target values as shown in Eq. (1). Unlike MSE and RMSE, the MAE gives equal weight to all the errors, regardless of their direction. For the MAE, different errors are not inflated like in the MSE and RMSE where larger errors are penalized, increasing the mean error value due to the square of the larger error value (Schneider & Xhafa, 2022).

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}|\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}| \tag{1}$$

where n is the number of observations in the dataset, y_i is the actual value of the *i*-th observation and $\hat{y}_i$ is the predicted value for the *i*-th observation. $|y_i - \hat{y}_i|$ is the absolute difference between the actual value and the predicted value.

2.2.2 Mean Squared Error (MSE)

Mean Squared Error (MSE) is a metric used to calculate errors in statistical models. It measures the average squared difference between the actual target values and the predicted values. Since there is a likelihood of equal numbers of both negative and positive differences

between actual and predicted values, squaring these differences eliminates the situation of them adding up to zero. The formula for calculating the MSE is

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y}_i - \hat{y}_\mathrm{i})^2 \tag{2}$$

2.2.3 Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) is calculated by finding the square root of the MSE. This means that the units of the RMSE are the same as the original units of the predicted target value. It is common to train the regression model with MSE and evaluate its performance using the RMSE (Schneider & Xhafa, 2022). The RMSE can be calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y}_i - \hat{y}_\mathrm{i})^2} \tag{3}$$

2.2.4 Coefficient of Determination (R^2).

The Coefficient of Determination (R^2) is a metric that measures the accuracy of line of fit of a model. In the context of regression, it is a measure of how well the regression line approximates the actual data. It is useful in models when the model is used to predict future outcomes. The formula for calculating R^2 is shown in Equation (4) below:

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}}$$

$$R^2 = 1 - \frac{\sum(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2}{\sum(\mathrm{y_i} - \bar{\mathrm{y}}_\mathrm{i})^2} \tag{4}$$

The R^2 values range from zero (0) to one (1). A Coefficient of Determination, $R^2 = 1$, denotes that the variation in the dependent variable (y) is fully accounted for by the independent variable (x) values. When $R^2 = 0.83$, 83% of the variation in the y values is accounted for by the x values

When $R^2 = 0$, none of the variation in the y values is accounted for by the x values.

2.2.5 Cost Functions

The cost function is an important component of ML that measures the difference between predicted values and the actual target values. It is also known as the loss function. An algorithm such as the backpropagation algorithm uses the gradient of the loss function to adjust the model's parameters and minimize the error margin, resulting in a better performance on the dataset. There are different types of loss functions which are classified based on the task for which they are applied. They mostly apply to regression and classification machine learning problems. For regression models, we usually use the Mean Square Error (MSE) / L2 Loss, Mean Absolute Error (MAE) / L1 Loss and Huber Loss / Smooth Mean Absolute Error (Alake, 2023). Equations (1) and (2) describe the MAE and MSE loss functions respectively. Equation (5) below describes the Huber loss function.

$$L_{\delta}\big(y, f(x)\big) = \begin{cases} \frac{1}{2}\big(y - f(x)\big)^2, \text{for } |y - f(x)| \leq \\ \delta \cdot \left(|y - f(x)| - \frac{1}{2}\delta\right), \text{otherwise} \end{cases} \tag{5}$$

where f(x) represents the predicted values generated by the model. The Huber loss function uses a piecewise function that implements the squared error function and alternates to the absolute error function where they are best applicable. The delta (δ) in Eq. (5) represents the difference between the predicted and actual values (Martinez, 2023).

2.3 Linear Regression

Linear regression is a type of ML algorithm that determines the linear relation between the dependent variable and one or more independent variables or features by computing a linear equation to observed data. Linear regression is a type of supervised learning, where the algorithm learns from labeled data to make predictions based on the input data (geeksforgeeks, 2024). A linear regression model with only one independent feature is known as Simple linear regression, and when there is more than one independent feature, it is known as Multiple linear regression.

The algorithm implemented in this research is a Multiple linear regression model with speed being our target variable (y) and the PSD, load and frequency being our input features (X_n). The equation below represents the Linear regression model used:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n \quad (6)$$

where: y is the dependent variable, X_1, X_2,, X_n are the independent variables. β_0 is the intercept and $\beta_1, \beta_2, \ldots, \beta_n$ are the slopes.

The main goal of the linear regression algorithm is to find the line of best fit equation that can predict the dependent variable, i.e. the speed, based on the independent variables (PSD, load and frequency).

2.4 Introduction to Deep Learning

Deep learning is a subset of ML methods. It works based on neural networks with the ability to discover representations needed for feature detection or classification from raw data. “Deep” is an adjective that depicts the use of multiple layers in the network. It uses hidden layers together with weights and biases to make predictions. Deep learning models such as deep neural networks, recurrent neural networks, convolutional neural networks and transformers have been

used in various fields including speech recognition, natural language processing, medical image analysis, fault and anomaly detection and many more. These models have been proven to produce results comparable to human expert performance. In some cases, their performances have exceeded that of humans (Wikipedia, 2024).

2.4.1 Recurrent Neural Networks (RNN)

One of the most common deep learning algorithms is the Recurrent Neural Networks (RNNs). RNN is a type of artificial neural network which works best with sequential data or time series data. RNNs are usually used for language translation, natural language processing (NLP), speech recognition, feature engineering, and image captioning. They are utilized by popular applications such as Siri, voice search and Google translator. The unique feature of RNNs is their "memory" as they take information from previous inputs to influence the current input and output (IBM, 2024).

RNNs differ from the feedforward networks in that they share the same weight parameters across each layer, whereas feedforward networks have distinct weights per node. This parameter sharing allows RNNs to process sequential data efficiently but also introduces some constraints such as exploding gradients, where gradients become excessively large resulting in weights represented as NaN (Not a Number), and vanishing gradients, where the weights become overly small and hinder learning. Figure 10 shows the architecture of RNNs, which can be thought of as multiple copies of the same network, each passing some information to the next cell.

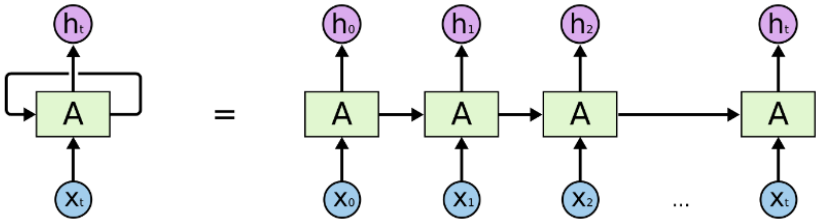

Figure 10: An unrolled recurrent neural network (Colah, 2015).

2.5 Random Forest Model

Random Forest is an ensemble learning technique that generates a more stable and accurate forecast by combining the predictions from several decision trees. It is a kind of technique for supervised learning that can be applied to tasks involving regression and classification. By combining the predictions from several models, ensemble learning is an ML technique that produces predictions that are more reliable and accurate. It is a method that increases the learning systems' overall performance by utilizing the combined intelligence of several methods as shown in Figure 11 (Random Forest Regression in Python, 2023).

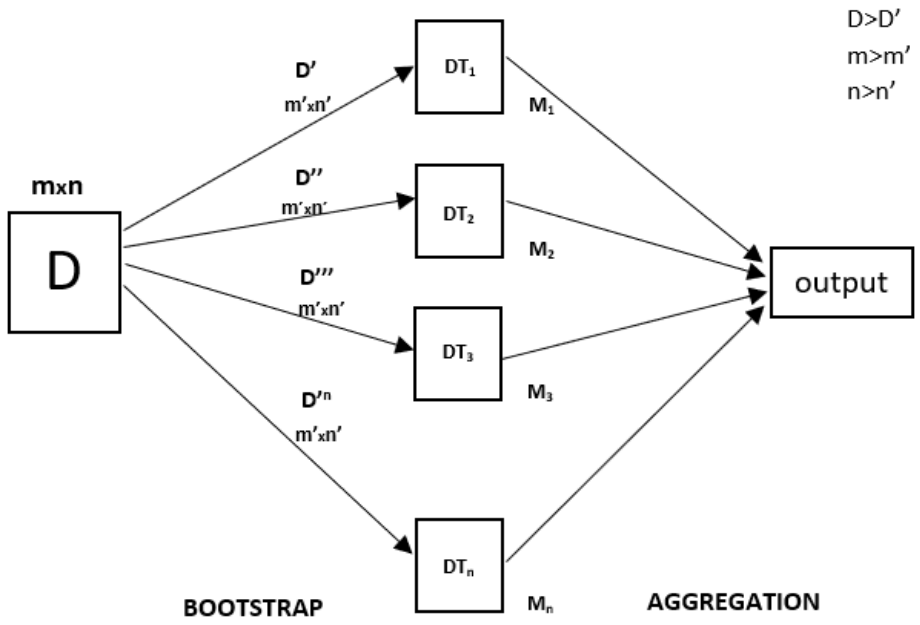

Figure 11: Random Forest Regression Model (Random Forest Regression in Python, 2023)

Each decision tree has a significant variance, but when we aggregate them all at once, the variance that results is low since each decision tree is perfectly trained on that specific sample data. As a result, the output depends on numerous decision trees rather than just one (Random Forest Regression in Python, 2023). For a classification problem the final output is determined by using the majority voting classifier, whereas in a regression problem, the final output is the mean of all the outputs. This is known as Aggregation.

2.6 XGBoost Regression Model

XGBoost is a type of ensemble learning algorithm which is used for regression, classification, and ranking problems. XGBoost stands for Extreme Gradient Boosting, which depicts the way this algorithm works. Specifically, it combines several weak models to generate a collectively strong model, hence the term ‘gradient boosting’.

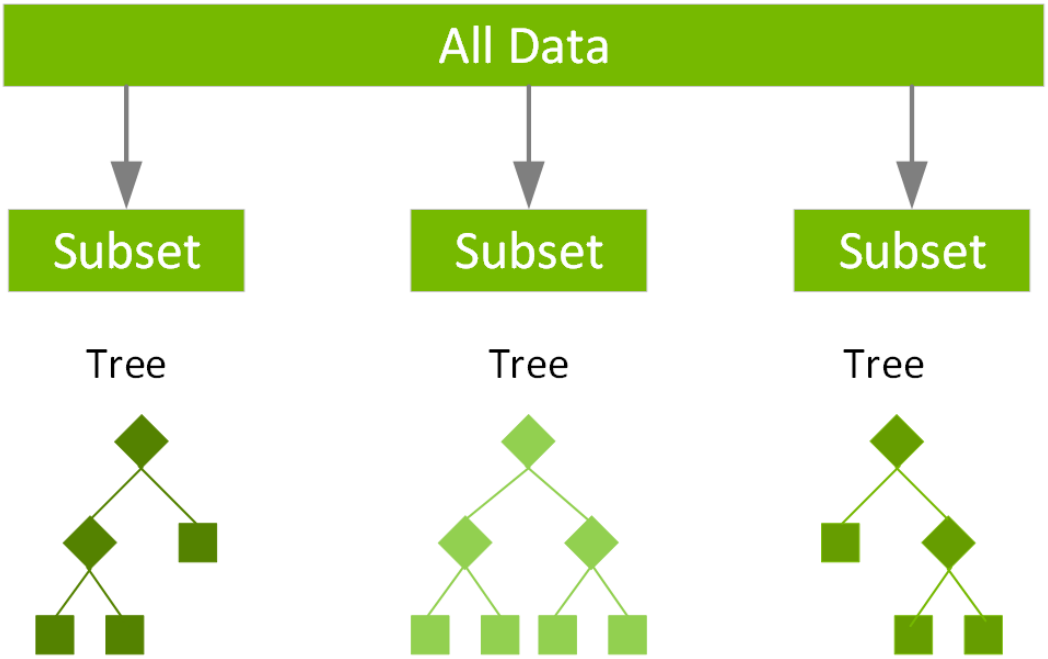

Figure 12; Gradient Boosting Decision Trees (NVIDIA, 2024)

Both Random Forest and Gradient Boosting Decision Trees (GBDT) work in similar ways by building a model consisting of multiple decision trees. The difference is with the construction of the trees. Random forest builds complete decision trees in parallel using random bootstrap samples of the data set using a method called bagging. Averaging all the decision tree forecasts results in the final prediction, whereas the XGBoost uses the "gradient boosting" technique as explained earlier (NVIDIA, 2024). Figure 12 is a schematic diagram of the Gradient Boosting Decision Trees.

CHAPTER III

EXPERIMENTAL SETUP AND PROCEDURES

3.1 An Overview of the Laboratory

The UTCRS has dynamic bearing testers used to simulate the various operating conditions experienced by tapered roller bearings in freight rail service. The four classes of railroad bearings that are tested by the dynamic bearing test rigs are the Association of American Railroads (AAR) class K, F, E and G bearings. However, the two bearings used in this thesis are listed in Table 1. The dynamic bearing testers are fitted with sensors that record the temperature, vibration, and load of the bearings. The data provided in this thesis was collected using two important loading conditions: 17% of full load (5,848 lbs. or 26kN), which represents an empty (unloaded) railcar, and 100% (34,400 lbs. or 153kN) of the total Association of American Railroads (AAR) load capacity for class F and K bearings, which corresponds to a fully loaded railcar.

The dynamic bearing test rigs are powered by a 22 kW (30 hp) variable speed motor which allows the testers to simulate the various train speeds listed in Table 2. The motors are regulated by variable frequency drives (VFDs) that maintain the desired angular speeds. The operating speeds range from 25 mph (234 rpm) to 85 mph (796 rpm) (Martinez, 2023). The test rigs can apply up to 150% of the AAR standard operating load of bearings of all classes tested. A load controller regulates this operation by controlling the pressure entering a hydraulic cylinder

and monitoring the resulting load using a load cell. The load controller is able to maintain the desired load within ±1% of full load range (Martinez, 2023).

Table 1: Bearing Standards by AAR: Classes, Dimension, and Load Capacities.

Class	Size [mm]	Size [in}	Load [kN]	Load [lbf]
Class F	165 × 305	6.5 × 12	153	34,400
Class K	165 × 229	6.5 × 9	153	34,400

Table 2: Angular speeds of axle to equivalent track speed.

Axle Rotational Speed [rpm]	**Equivalent Track Speed [mph]**	**Equivalent Track Speed [km/h]**
234	25	40
280	30	48
327	35	56
374	40	64
420	45	72
467	50	80
498	53	85
514	55	89
560	60	97
618	66	106
700	75	121
796	85	137

3.1.1 Four-Bearing Test Rig (4BT)

The Four-Bearing Test Rig (4BT) can accommodate all four classes of railroad bearings (Class F, K, G or E). The bearings are pressed onto a text axle before mounted on to the test rig for loading and testing. Figure 13 shows a picture of the 4BT, and Figure 14 shows a schematic diagram of the instrumentation setup for the 4BT. The two center bearings are top loaded while the two outer bearings at both ends of the axle are subject to bottom loading. The datasets used for this study were collected from the two center bearings (B2 and B3) since they are both top loaded just like in freight rail service. The two center bearings have adapters machined to accept two customized 70g analogue accelerometers placed in the outboard Smart Adapter (SA) and Mote (M) locations, one 500g PCB accelerometer placed in the outboard Radial (R) location, two K-type bayonet thermocouples, and one regular K-type thermocouple aligned with the two bayonet thermocouples and placed midway through the bearing cup width and held tightly by a hose clamp. Figure 15 shows the SA, M and R locations of the accelerometers on the bearing adapter (Montalvo, 2019).

Figure 13: Four-Bearing Test Rig (4BT) (Montalvo, 2019).

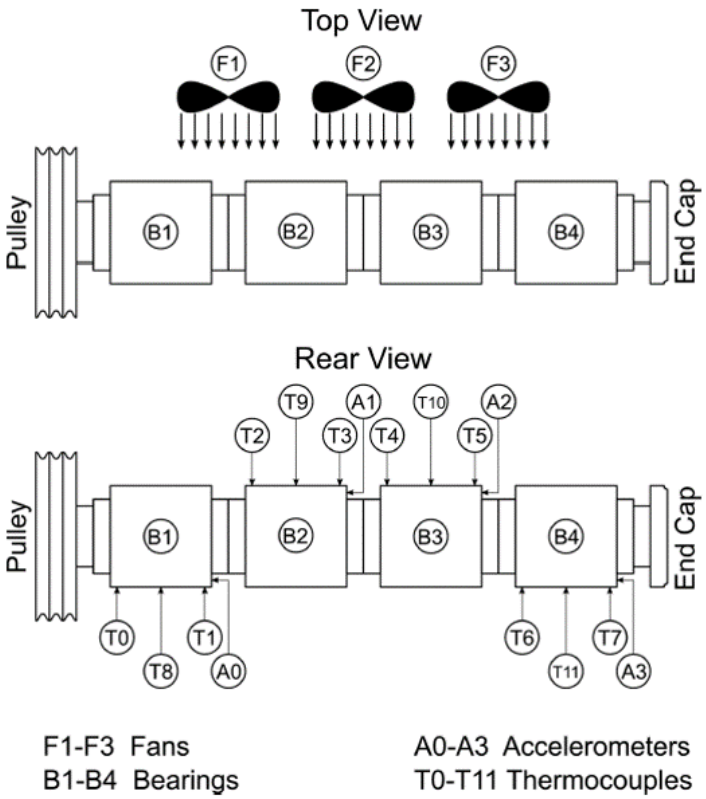

Figure 14: Top and Rear views of 4BT including sensor locations (Montalvo, 2019).

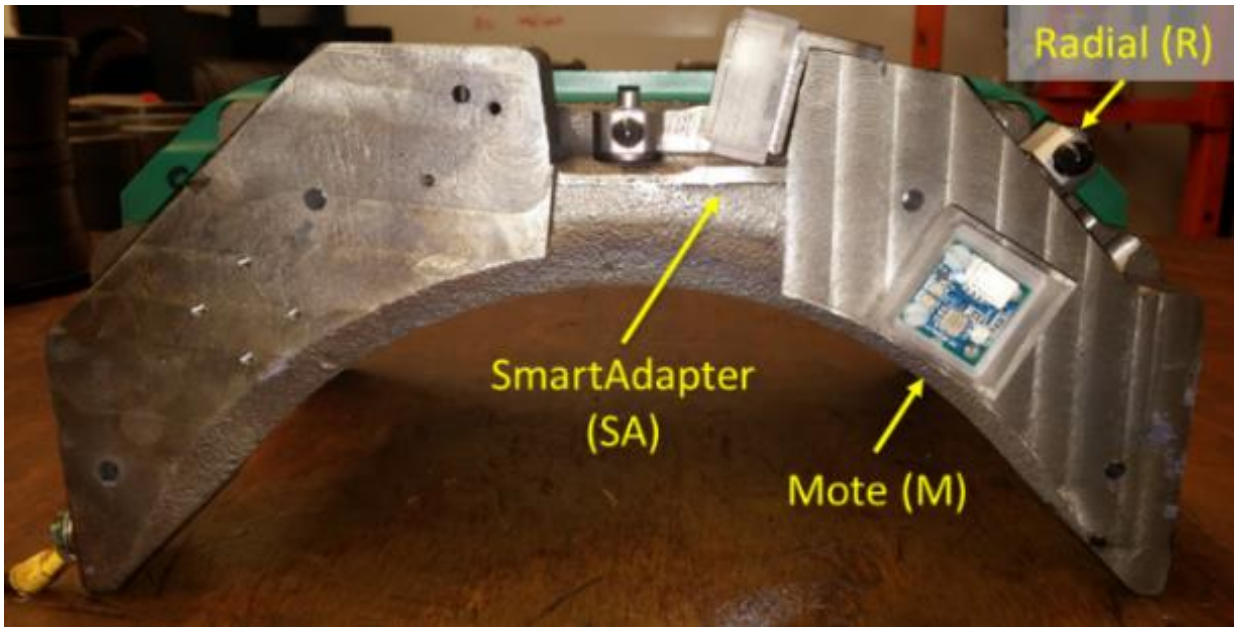

Figure 15: Bearing adapter showing accelerometer locations (Montalvo, 2019).

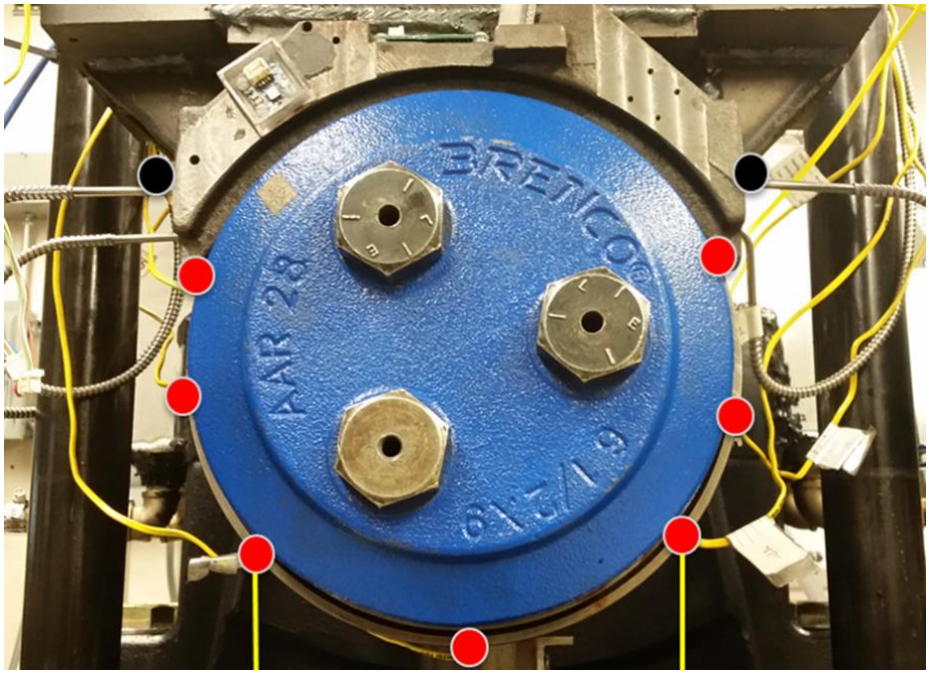

Figure 16: Placement of thermocouples on the railroad bearing. (red dots represent the locations of the regular K-type thermocouples and black dots represent the locations of the bayonet K-type thermocouples)

3.2 Overview of the Data

The data collected by the onboard sensors is recorded via a data acquisition (DAQ) system and then goes through a central processing unit for analysis. The system shows the profiles of the vibration, temperature and load of the bearings based on the data acquired. These three metrics best describe the health of the railroad tapered roller bearings as mentioned earlier. Thus, the UTCRS research team records and processes: accelerometer data measured in g, harvesting voltage, temperature data in Celsius, VFD voltage, and load cell data in lbs. In addition, the mileage run by each tapered roller bearing is also tracked and recorded.

3.2.1 Data Recording

The National Instruments (NI) PXIe-1062Q data acquisition system is used to record the data. The data is transcribed to text-based files using a the LabVIEW™. The NI TB-2627 records the temperature and an 8-channel NI PXI-4472B card records the vibration data. The temperature data was recorded at a sampling rate of 128 Hz every 20 seconds, whereas the vibration data was captured at a rate of 5.12 kHz for 16 seconds every 10 minutes respectively. The load cell data was recorded at the same sampling rate as the temperature profiles. The data is then transferred to a central processing unit for analysis.

3.2.2 Data Analysis

The UTCRS laboratory uses the MATLAB® software to run the analysis of the data collected via LabVIEW™. There are three (3) levels of analysis conducted for condition monitoring of bearings at the UTCRS. Level 1 of the algorithm determines the condition of the bearings from the temperature and vibration profiles by comparing them to two speed-dependent (preliminary and maximum) thresholds developed using a library of defect-free bearing vibration signatures acquired by laboratory testing. Figure 17 presents example plots of the temperature and vibration profiles for two test bearings.

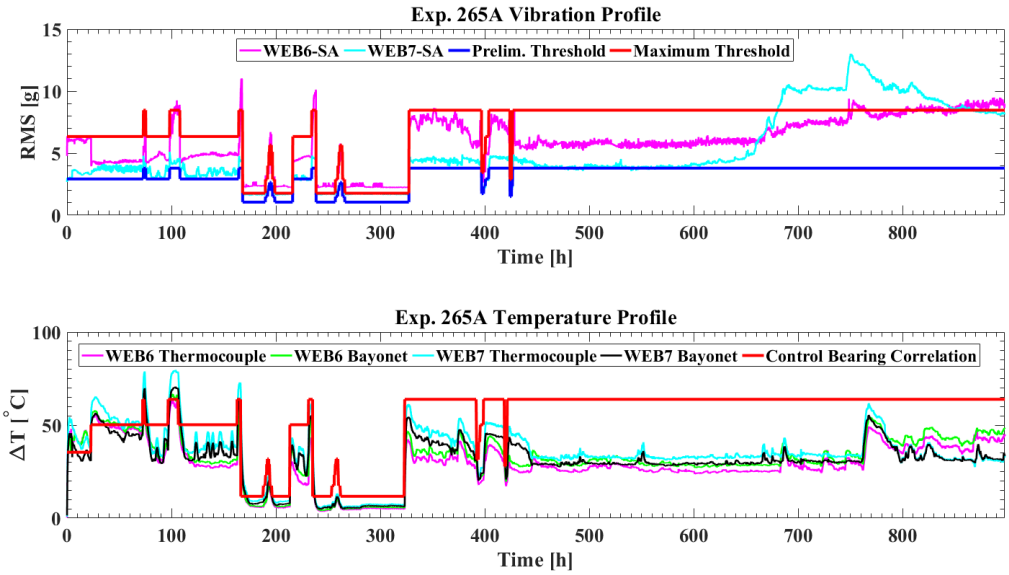

Figure 17: Vibration and Temperature analysis plots (Martinez, 2023).

The vibration plot shows the root mean square (RMS) of the vibration profile plotted against the preliminary and maximum thresholds. A first-order line of fit with a 95% confidence preliminary interval and a 45% confidence maximum interval dictates the equations that establish the thresholds. The bearing is statistically defect-free (healthy) when the vibration levels are below the preliminary threshold. The bearing is still within operable conditions and has the potential to develop flaws when the vibration level lies between the two thresholds; it is no longer regarded as optimum. A bearing exhibiting vibration levels over the maximum threshold for an extended duration (at least five hours) strongly suggests the presence of a problem. The latter prompts further analysis using the subsequent levels of the algorithm. Comparing the bearings' present temperature to a control bearing correlation allows for a comparable interpretation of the temperature plot. The two speed-dependent thresholds emphasize the fundamental reason for this research to focus on vibration data.

After the detection of defects using level 1 analysis, level 2 analysis is performed to determine the defect type (local or distributed/geometric) present within a bearing that has been found to be defective. Level 2 analysis utilizes a frequency-domain analysis and creates power spectral density (PSD) plots, where a PSD ($|X(f)|^2$) is the square of the magnitudes in the frequency domain, after which six rotational frequencies are tracked, shown in Eq. (8) – Eq. (13) (Montalvo, 2019).

$$\text{Total Power} = \int_{-\infty}^{\infty} |X(f)|^2 df \tag{7}$$

$$\omega_{cone} = \omega_o \tag{8}$$

$$\omega_{cage} = \left(\frac{R_{cone}}{R_{cone} + R_{cup}}\right)\omega_{cone} \tag{9}$$

$$\omega_{roller} = \left(\frac{R_{cone}}{D_{roller}}\right)\omega_{cone} \tag{10}$$

$$\omega_{out} = 23\omega_{cage} \tag{11}$$

$$\omega_{in} = 23\left(\omega_{cone} - \omega_{cage}\right) \tag{12}$$

$$\omega_{rolldef} = \left(\frac{R_{cup}}{R_{roller}}\right)\omega_{cage} \tag{13}$$

Using Eq. (7), the PSD plots can be created, in which a faulty component will exhibit a power spike at the corresponding defect frequency, while a healthy bearing will exhibit no significant power changes. To correctly classify the bearing defect type, the fundamental frequencies are essential, which can be tracked based on the rotational speed of the bearings, as

shown in Equations (8) – Eq. (13), among which, Equations (11) - (13) represent three different tapered roller bearings defect frequencies related to faulty outer rings, inner rings, and rollers, respectively. Examples of each kind of bearing condition, along with the matching defect frequency and its harmonics, are displayed in Figure 9.

Level 3 analysis is the last stage of condition monitoring for a bearing. It allows for the estimation of the defect's size without requiring the bearing to be stopped, dismounted, and disassembled, and visually inspected. Experimentally devised correlations that relate the bearing vibration RMS to the defect size are used for this analysis. These correlations are graphically represented in Figures 18 and 19, respectively, for the bearing cup (outer ring) and cone (inner ring).

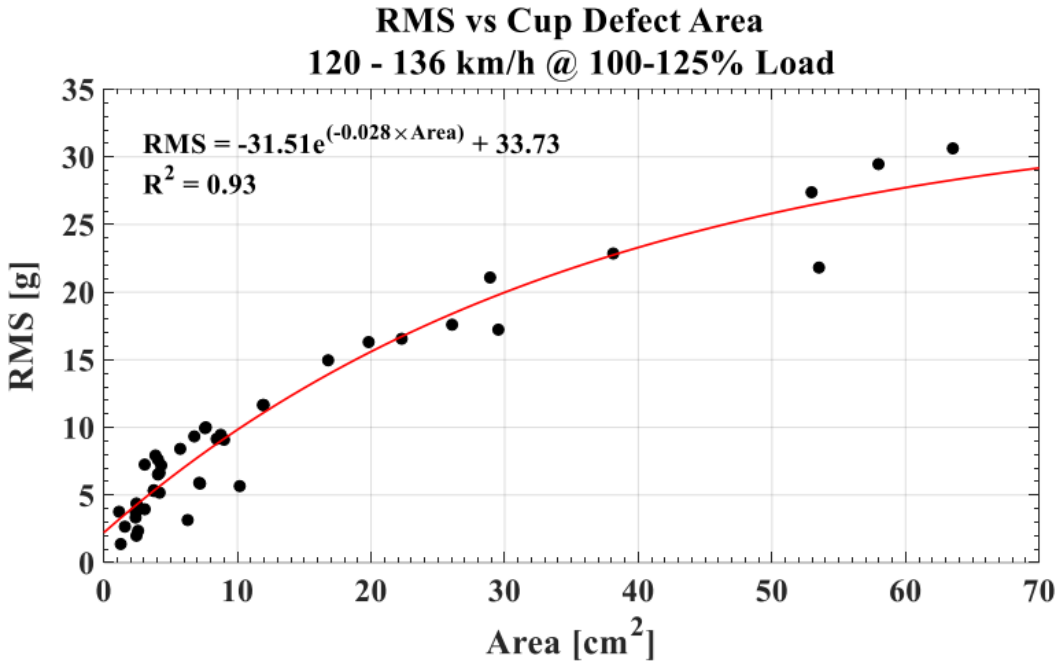

Figure 18: Correlation Analysis at Level 3 for a Bearing Cup (outer ring) (Martinez, 2023).

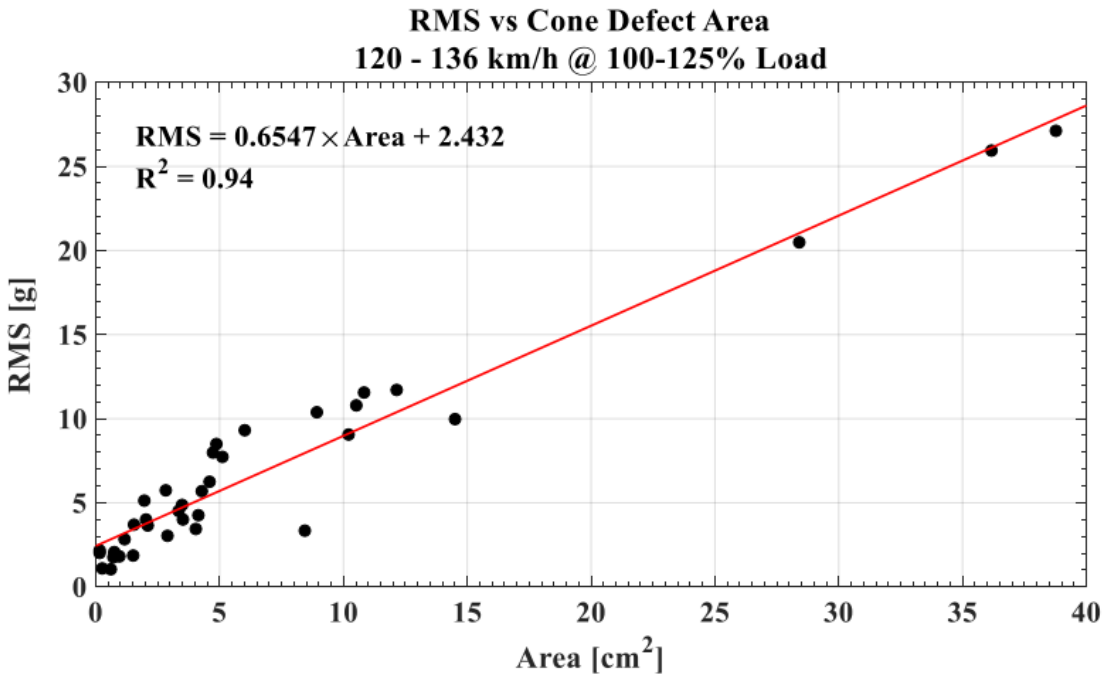

Figure 19: Correlation Analysis at Level 3 of a Bearing Cone (inner ring) (Martinez, 2023).

CHAPTER IV

METHODOLOGY AND MODEL SPECIFICATIONS

4.1 Datasets

The datasets used for training the models were obtained from data acquired from the dynamic testers that were used to assess the performance of railroad bearings at the UTCRS. The data is stored in folders with unique identifiers that show the chronological order and the iteration of the experiment. These folders were run through MATLAB to extract the frequency and PSD values from the PSD plots and then saved as a csv file in Excel. The load and speed values were acquired from the folder name, since the amount of load applied, and the speed of the testers are components of the unique folder names. For example, the first iteration of experiment 264 will have a folder named 264A. Experiment 264A may have different iterations that are based on either passing certain mileage threshold or varying testing conditions. The data recorded in these folders are the following: temperature, acceleration, motor power, harvesting voltage, VFD voltage, and load cell data. This study focuses on the accelerometer data since we are focusing on the vibration profiles of the bearings. Over eight million data samples were used to train the models, each containing three features (i.e., frequency (Hz), PSD (g^2/Hz), and Load %) and one label (i.e., the Speed in rpm). Table 3 provides example data samples. The data samples were obtained from experiments where the bearings were healthy, contained cone defects, or contained cup-defects

There were minimal data points for bearings with roller defects since the roller is the hardest component of the bearing assembly, and it does not develop spalls easily.

Table 3: Dataset with input features (Frequency, PSD, Load %) and target value (Speed)

Frequency [Hz]	PSD [g^2/Hz]	Load [%]	Speed [rpm]
0	2.72E-34	100	234
0.678	3.06E-06	100	234
1.356	4.60E-06	100	514
822.618	5.09E-04	100	790
823.296	1.65E-04	17	560
824.652	3.68E-04	17	327

The objective is to train the ML models to be able to predict the speed (dependent variable) from the input features (independent variables), i.e., frequency, PSD, and Load. Table 4 shows the experiments that were used to generate the dataset.

Table 4: Experiments ID used for generating dataset

Defective Cup Experiments				
178A	178B	178C	178D	200A
200B	200C	226A	226B	225
235A	268			
Defective Cone Experiments				
185	187	190A	190B	190C
190D	192	193	194	195A

Table 4 cont.

195B	204A	256A	260A	
Healthy (Defect-Free) Bearings				
211A	211B	211C	212A	212B
213A	213B	214A	214B	214C
267	263A			

4.1.1 Feature Set

The dataset used to train the ML models contained three features, i.e., the frequency (Hz), PSD (g^2/Hz), and Load (%) and one label (i.e., the Speed (rpm)). In the preliminary stages, frequency and PSD were the only features used, but the load was later added as we would observe its contribution to the models' predictions, as will become evident in the later chapters. The frequency values in dataset represent the sampling rate which ranges from 0 to 2,777 Hz. The sequence of data collected over 4 seconds at a sampling rate of 5.12 kHz (20,480 data points) is recorded as a single data point (average of the sequence) in the models. The load values used were 100% and 17%, which represent fully loaded and unloaded (empty) railcar conditions, respectively. The speed values used were the angular speeds of the VFD, which corresponded to different track speeds as listed in Table 2 in Chapter 3.

4.2 Random Forest Specifications

The Random Forest model employed in this analysis is a regression model which was constructed using the 'RandomForestRegressor' from the 'sklearn' library (Learn, 2024). The model is configured with 25 trees, each with a maximum depth of 25, which controls the

complexity and prevents overfitting. In addition, the model uses all the available CPU cores (n_jobs = -1) to parallelize the training process, enhancing the computational efficiency. This model is trained on the dataset to predict continuous target speed values, and its performance was evaluated using the R-squared (R^2), Mean Absolute Error (MAE) and Mean Squared Error (MSE), which provide insights into the accuracy and reliability of the model's predictions. Table 5 shows the parameter distribution of the Random Forest model.

Table 5: Parameter distribution of Random Forest model

Parameter	Values
random_state	42
n_estimators	25
max_depth	25
n_jobs	-1

The dataset is split into training and testing sets, with 20% of the data allocated for testing. The data is shuffled before splitting to ensure that the model consistently achieves the same results using the same dataset.

4.3 XGBoost Regression Specifications

The XGBoost regression model uses the 'XGBRegressor' class from the 'xgboost' library. The model has specific hyperparameters that particularly help improve the performance of the model on the given dataset. The model uses the 'gbtree' booster, which builds an ensemble of decision trees to predict continuous target values (XGBoost, 2022). Like the Random Forest model, the XGBoost model's complexity and capacity are controlled by number of estimators set

at 25, which specifies the number of trees, and a maximum depth of 25, which limits the depth of each tree, preventing overfitting. Also, all the available CPU cores were used to speed up the training. The model is trained on standardized training data and its performance is evaluated using the R-squared (R^2), Mean Absolute Error (MAE) and Mean Squared Error (MSE). The model shares the same parameter distribution as in the Randon Forest model shown in Table 6. The dataset is split into training and testing sets, with 20% of the data allocated for testing. The data is shuffled before splitting to ensure that the model consistently achieves the same results using the same dataset.

Table 6: Parameter distribution for XGBoost Regression model

Parameter	Values
random_state	42
n_estimators	25
max_depth	25
n_jobs	-1

4.4 Linear Regression Specifications

In this analysis, we implemented and trained a linear regression model using the training dataset. The model was created with an intercept term (`fit_intercept=True`) to account for the baseline level of the target variable. After fitting the model to the training data, predictions were made for both the training and testing datasets. The Mean Squared Error (MSE) was calculated for the training data to assess how well the model captured the relationship between the features and the target variable. Additionally, predictions on the testing data were generated to evaluate

the model's performance on unseen data. This process allows us to identify potential overfitting or underfitting issues by comparing the performance metrics (such as R^2, MAE, and MSE) between the training and testing datasets, ensuring that the model generalizes well to new data.

CHAPTER V

RESULTS AND DISCUSSION

5.1 Descriptive Analysis

Figure 20 is a correlation heatmap that shows the Pearson correlation coefficients among the features (frequency, PSD, and load) and the target variable (speed). The Pearson correlation is a measure of the linear relationship between two variables. It describes how the changes in one variable affect the other variable. It ranges from -1 to 1. A correlation of +1 represents a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. Equation 14 calculates it by dividing the covariance of the two variables by the product of their standard deviations.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (14)$$

where x_i and y_i are individual sample points, and $\bar{x}$ and $\bar{y}$ are the means of the x and y variables. A value ranging from 0.5 to 1 will represent a strong positive linear relationship, 0 to 0.5 a weak positive relationship. – 0.5 to 0 shows a weak negative relationship and -1 to -0.5 a strong negative relationship. The Pearson correlation is unique to the dataset used. This implies that once the dataset is changed, we would have different correlation values. Figure 20 shows that the frequency has a negligible linear correlation with the speed with a value of -0.047, the PSD

shows a moderate correlation with the speed with a value of 0.35 and the load shows the highest correlation with a value of 0.48.

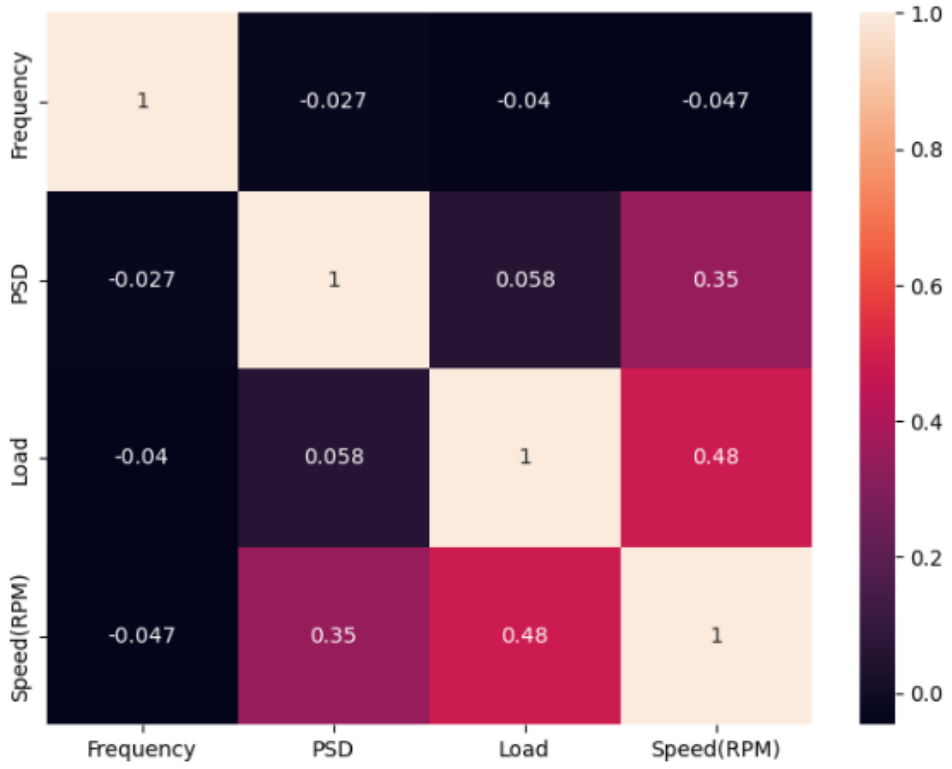

Figure 20: Correlation heatmap of all variables in the dataset

However, these correlation values do not necessarily depict the importance of each feature in the prediction of speed since they focus on showing the linear relationship amongst the variables. Other factors, such as the different datasets, data distribution (outliers, range of values), data quality (measurement errors, consistency) and many more may influence these numbers. Figure 21 shows the features and their respective importance in predicting the speed. The bar chart shows the feature importance scores for frequency, PSD, and load in predicting the speed. PSD has the highest bar with a score of 0.86 indicating that it is the feature with the most influence in predicting the speed

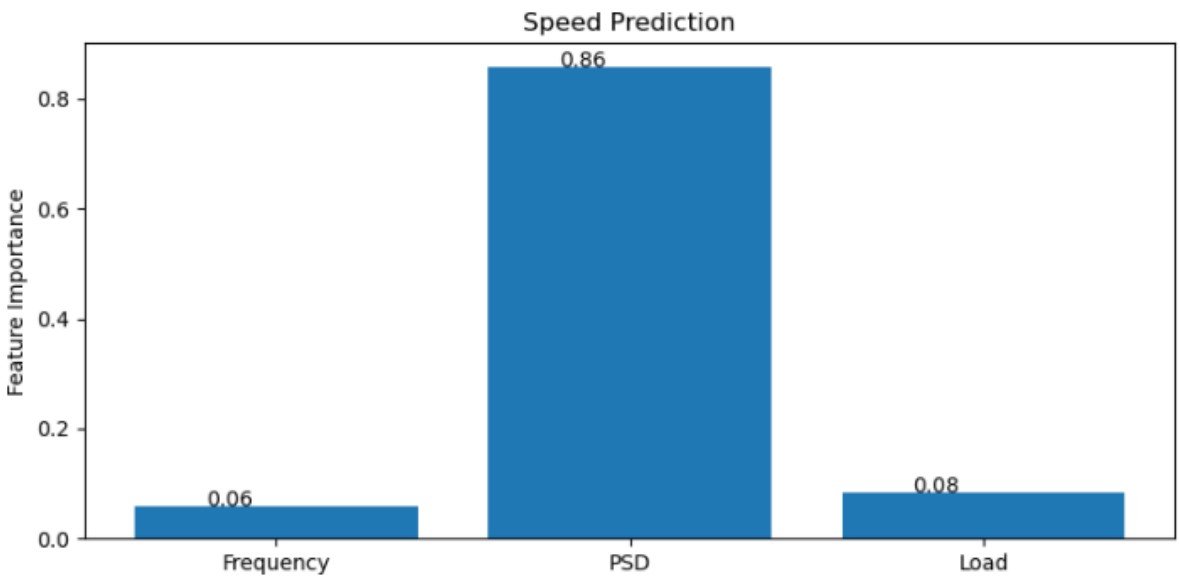

Figure 21: Feature importance of features

Figure 22 is a multi-plot chart of an overview of the frequency, PSD, load and speed in the dataset. The frequency shows some peaks over time; however, this does not suggest a strong relationship with the speed. The load shows a step-like plot that fluctuates between the 17 and 100 settings. There appears to be a corresponding relationship with speed and load, with higher load values coinciding with higher speed values. However, this does not define the relationship between the load and the speed since the load shows a binary characteristic, which might affect the linearity of the relationship. Note that the latter is an artifact of the laboratory testing where assessing the performance of the bearings requires running them at full load and high speed, which is the most common setting. It can be observed that there is a noticeable correlation between the PSD peaks and the speed. When the speed increases, the PSD tends to rise, indicating a positive relationship. This supports the assertion made from the bar chart, showing the PSD as the most important feature in predicting speed.

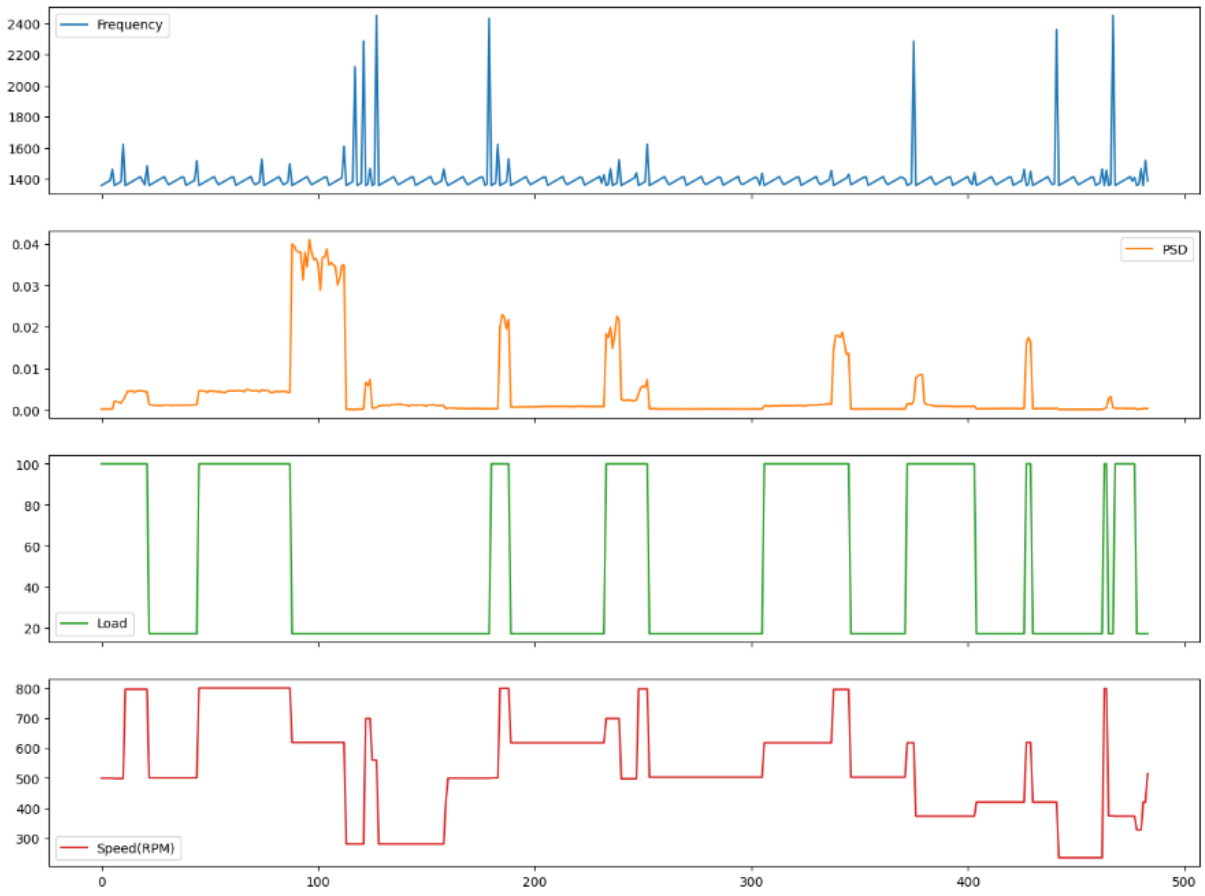

Figure 22: Plot of the features

5.2 Random Forest Regression Results

The Random Forest model was implemented to predict speed based on the features frequency, PSD and load. The model demonstrated robust performance, as depicted by its evaluation metrics. The R-squared (R^2) value was 0.84, indicating that the model explains approximately 84% of the variance in the target variable, suggesting a good fit. Furthermore, the mean absolute error (MAE) was 36.1 and the mean squared error (MSE) was 4030.2, indicating that the model's predictions were close to the actual values. Figure 24 presents the plot of the predicted values versus the actual values for a sample of 100 data points.

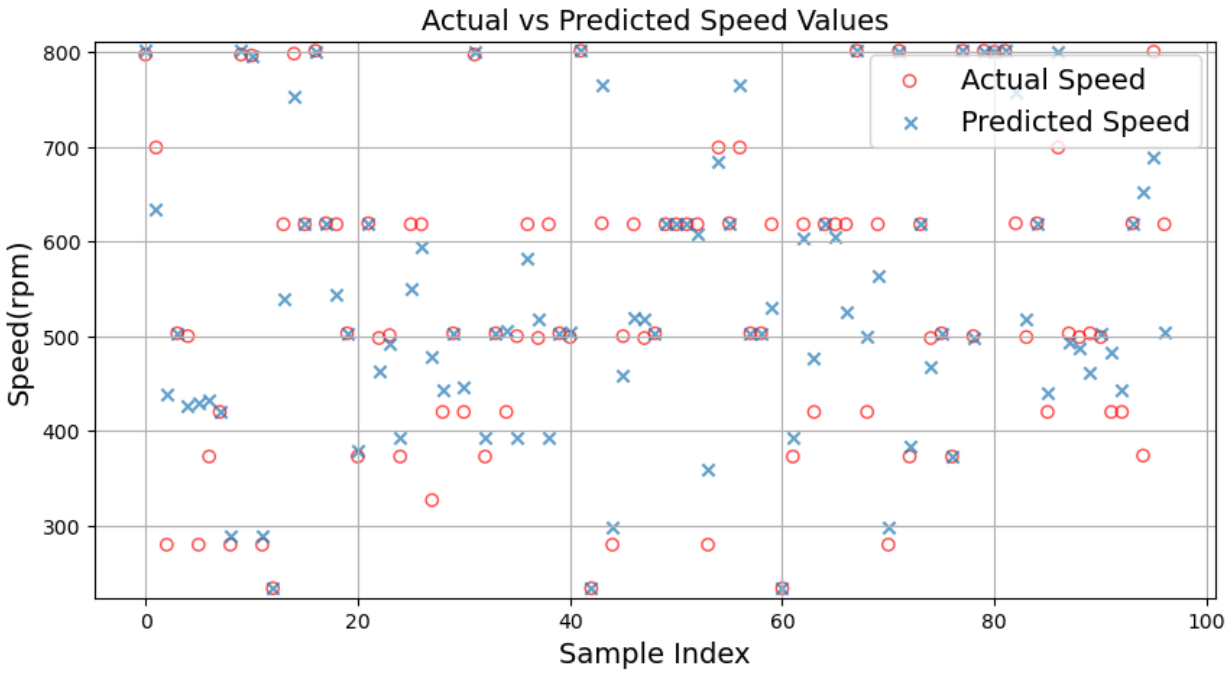

Figure 23: Plot of predicted vs. actual speed values with all features

The horizontal axis (x-axis) represents the index of the sample data points, ranging from 0 to 100. The vertical axis (y-axis) represents the speed values in revolutions per minute (rpm). The precision of the predicted values indicates how well the model's predictions align with the actual speed values. The plot shows circles and crosses overlapping or very close to each other, suggesting that the model is accurately predicting the speed. There are few instances where the model's predictions seem to be off. This can be attributed to the inability of the model to exhaustively account for all variations in the speed as evident in the R^2 value, 0.84.

In contrast to Figure 23, Figure 24 shows the predicted values versus the actual speed values when the load is excluded from the features. This means that the model was trained with the PSD and frequency as the features and the speed as the target variable. The R-squared (R^2) value was 0.68, indicating that the model explains approximately 68% of the variance in the

target variable. The mean absolute error (MAE) was 57.2 and the mean squared error (MSE) was 7854.6.

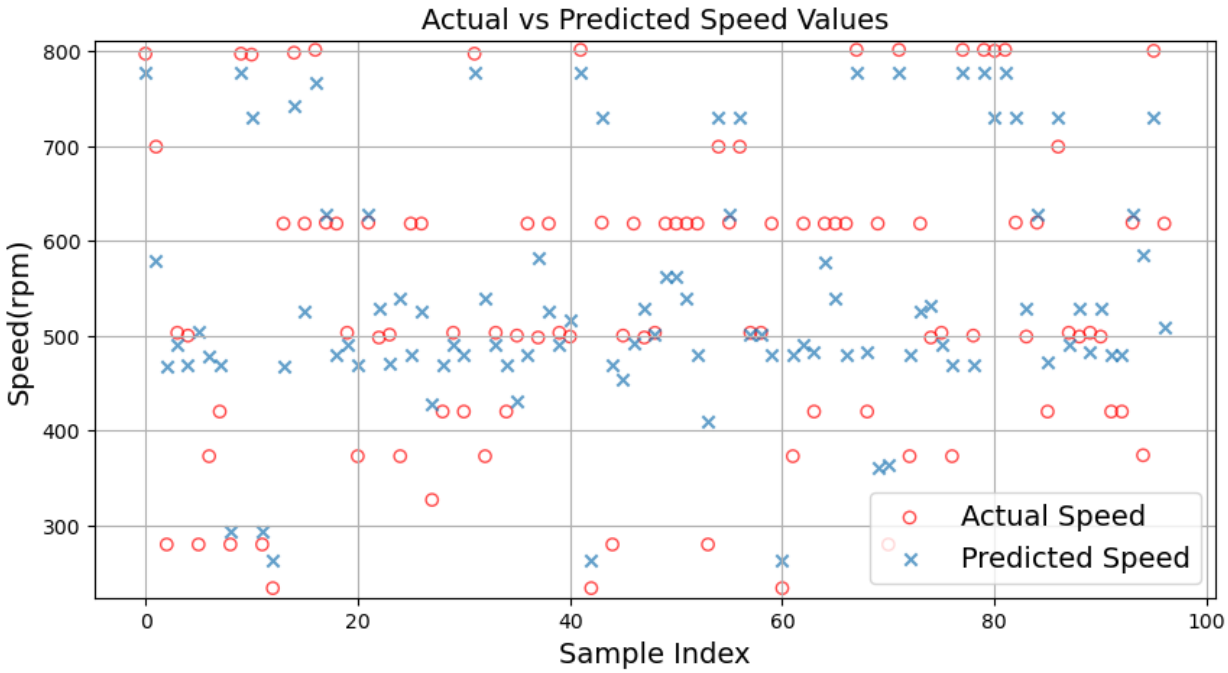

Figure 24: Plot of predicted vs. actual speed values without the load feature

It is evident from both plots, i.e., Figure 23 and Figure 24 that including the load feature improves the performance of the model's prediction, which implies that the speed is a dependent variable influenced by the load. Table 7 summarizes the comparison between the model trained with all features (Load, PSD and Frequency) and the model trained without the load feature.

Table 7: Performance metrics of Random Forest Models

Metrics	Model trained with all features	Model trained without the load feature
R-squared (R^2)	0.84	0.68
MAE	36.1	57.2
MSE	4030.3	7854.6

5.3 XGBoost Regression Results

The XGBoost Regressor is another model that was employed to predict speed based on the features frequency, PSD, and load. This also demonstrated a robust performance, as seen by its evaluation metrics. The R^2 value was 0.76, indicating that the model explains approximately 76% of the variance in the target variable, suggesting a good fit. Furthermore, the MAE was 36.0and the MSE was 6024. 3, indicating that the model's predictions were close to the actual values. Figure 25 shows the plot of the predicted values versus the actual values for a sample of 100 data points.

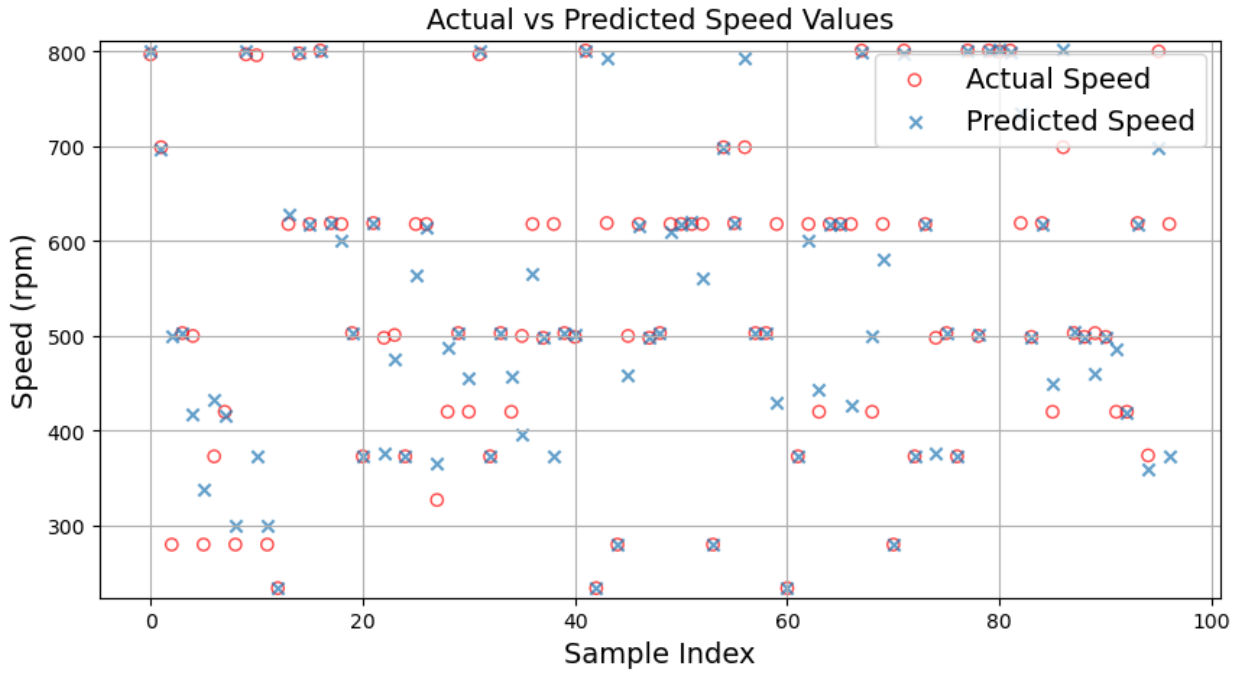

Figure 25: XGBoost model predictions with all features

The axes of the XGBoost regression plots are the same as those of the Random Forest plots. Just like the Random Forest, the XGBoost regressor model was trained without the load feature to observe the impact of the load feature on the model's prediction. Figure 26 shows a plot of the XGBoost model trained without the load feature.

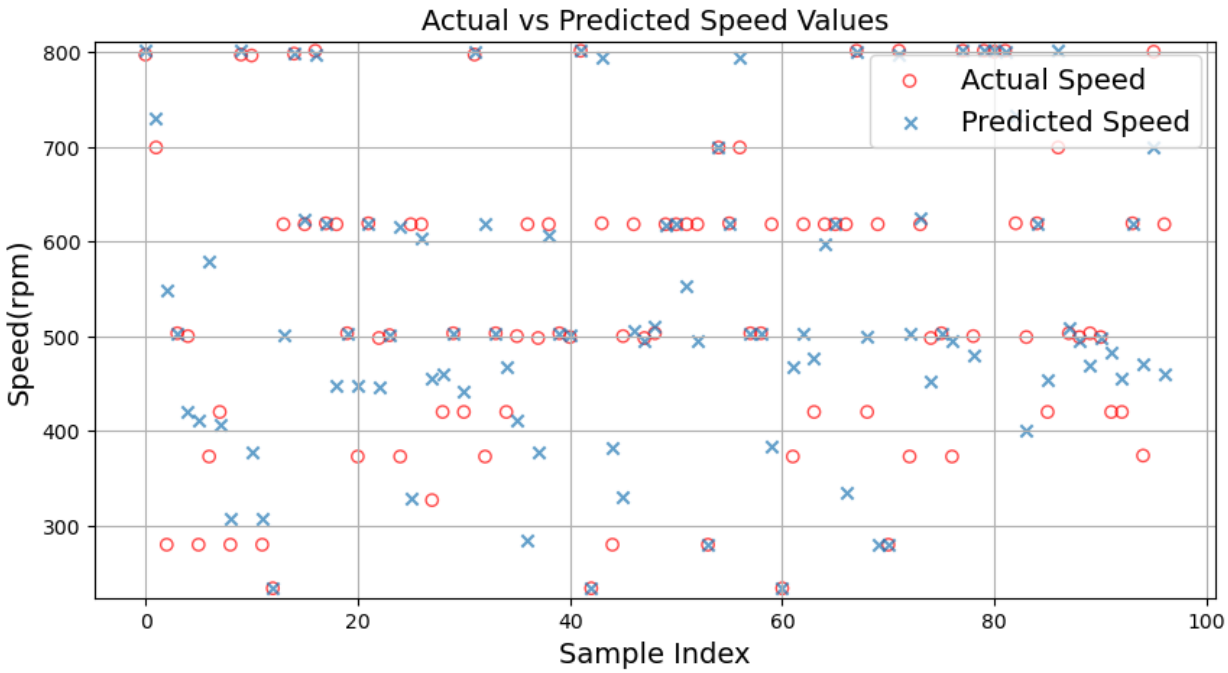

Figure 26: Plot of predicted vs. actual values of speed without the load feature

Comparing Figure 25 and 26, there is evidence of the better performance of the model when the load feature is included. In contrast to Figure 25, Figure 26 shows the predicted values versus the actual speed values when the load is excluded from the features. This means that the model was trained with the PSD and frequency as the features and the speed as the target variable. The R^2 value was 0.48, indicating that the model explains approximately 48% of the variance in the target variable. The MAE was 66.8 and the MSE was 12734.2.

Table 8 summarizes the comparison between the model trained with all features (Load, PSD, and Frequency) and the model trained without the load feature.

Table 8: Performance metrics of XGBoost Models

Metrics	Model trained with all features	Model trained without the load feature
R-squared (R^2)	0.76	0.48
MAE	36.0	66.8
MSE	6024.3	12734.2

5.4 Linear Regression Results

The linear regression model was also tested on our dataset and the performance metrics indicate that the model was overfitting, as demonstrated in Figure 27. The R^2 score of 1.0 suggests a perfect prediction accuracy, meaning that the model explains all the variance in the target variable. The MAE was extremely low with a value of 1.09e – 14, and the MSE was 7.97e – 28. The model gave a similar result when the load feature was excluded. Though this plot looks like a perfect result, we realized that the model was overfitting since it was unable to perform well when it was introduced to a different dataset. Using this linear regression model can lead to poor generalization of new or unseen data.

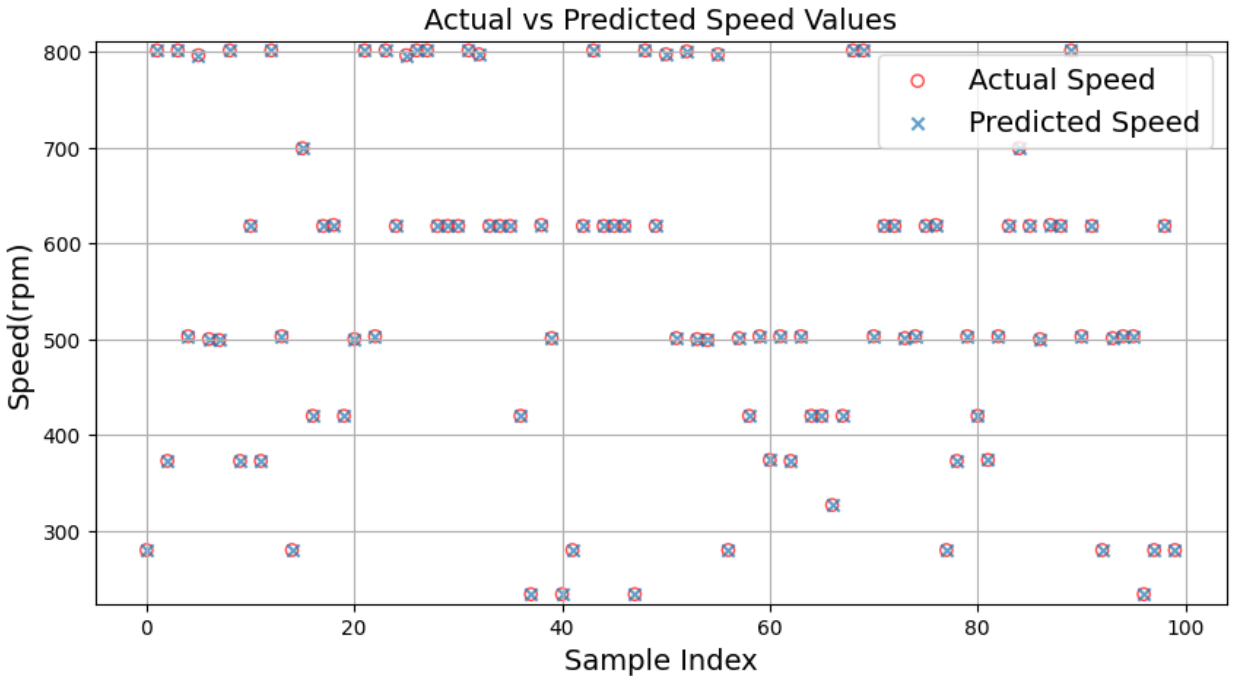

Figure 27: predicted vs. actual speed values of Linear Regression model

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

This research work was geared towards extracting features such as the speed of the bearings from the vibration signatures obtained from the onboard sensor module developed by the UTCRS. The data was initially trained with a linear regression model and the Long Short-Term Memory (LSTM) network since it was assumed that the data was sequential. Both models did not produce good results, with the linear regression model overfitting and the LSTM model providing very low accuracy. Further research suggested that the data was in tabular form, and it would be best to train it with other regression models. The Random Forest and XGBoost models were the best choices due to their performances on the data.

The Random Forest model generally predicted the speed values quite well and with few errors. However, there are some points where the predicted values deviate from the actual values, indicating that the model could be improved. The XGBoost regression model also predicted the speed values quite well, though it had a lower R^2 value and a higher MSE value compared to the Random Forest model. Though the Random Forest model shows slightly better performance compared to the XGBoost model, both models were the best performers on the dataset used. These results proved that it was possible to predict the speed values from the vibration signature obtained from our sensor modules.

Also, it was observed that the performance of both models improved with the addition of the load feature in the input feature set. Despite PSD being the most dominant feature in predicting the speed as proven by the feature importance analysis, the load feature also plays a crucial role in improving the model's accuracy. By accounting for the variations in load, the model was able to better capture the complex relationship between the operating conditions and the resulting speed, thus reducing the MAE and MSE. This improvement underscores the importance of considering multiple relevant features, even those with moderate individual impact, to achieve a more accurate predictive model.

The PSD showing the highest bar in terms of feature importance in our predictive models, emphasizes the fact that it is possible to predict the speed of the bearings with PSD plots. With further improvement of these models, there is a possibility of manufacturing a sensor that can carry out an autonomous and real-time vibrational analysis as the train is in motion without the help of additional sensors or GPS.

Some future work will include improving our dataset with more features such as the skewness, standard deviation and kurtosis of the feature set if necessary to improve the performance of the models. It was also observed during the training process that larger datasets improve the models' performance; hence, it would be appropriate to keep adding to the existing dataset. This would enable the model to learn better from the data and model the pattern of actual speed values. The models used in this research work can again be improved by hyperparameter tuning to optimize the parameters that influence the learning process of the models. This would ensure that these models perform at their best, balance computational efficiency, and adapt to specific problem requirements.

REFERENCES

Alake, R. (2023, November). Loss Functions in Mahcine Learning. Retrieved from datacamp: https://www.datacamp.com/tutorial/loss-function-in-machine-learning

Ali, M. (2022, August). datacamp. Retrieved from Supervised Machine Learning: https://www.datacamp.com/blog/supervised-machine-learning

Board, N. T. (2023, February 3). Norfolk Southern Railway Train Derailment with Subsequent Hazardous Material Release and Fires. Retrieved from NTSB: https://www.ntsb.gov/investigations/Pages/RRD23MR005.aspx

Brown, S. (2021, April 21). Machine learning, explained. Retrieved from MIT Management Sloan School: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

Colah. (2015, August 27). Understanding LSTM Networks. Retrieved from colah's blog: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Cuanang, J. R. (2020). Optimizing a Railroad Bearing Condition-Monitoring Algorithm for Use with an Onboard Wireless Low-Power Sensor Module. Edinburg, TX, United States: Master's Thesis, The University of Texas Rio Grande Valley.

Economics, L. S. (2024). What Is the Difference Between Bias and Variance? Retrieved from Master's in Data Science: https://www.mastersindatascience.org/learning/difference-between-bias-and-variance/

Federal Railroad Administration. (n.d.). Federal Railroad Administration, Office of Safety Analysis. Retrieved from 3.10 - Accident Causes: https://safetydata.fra.dot.gov/OfficeofSafety/default.aspx

Freight, C. (2011). Railway Investigation Report R11T0016 - Transportation Safety Board of Canada. Tsb.gc.ca.

geeksforgeeks. (2024, March 20). Linear Regression in Machine learning. Retrieved from geeksforgeeks: https://www.geeksforgeeks.org/ml-linear-regression/

Hernandez, V. V. (2020). Assessing the Performance of Reconditioned Railroad Tapered Roller Bearings Used in Freight Rail Service. 106.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9 (8), 1735-1780.

IBM. (2024). recurrent neural network. Retrieved from IBM: https://www.ibm.com/topics/recurrent-neural-networks

Joseph, M., Constantine, T., & Arturo, A. (2018, April). Vibration-based Defect Detection for Freight Railcar Tapered-Roller Bearing. Joint Rail Conference, 9.

LBFoster, U. (n.d.). WILD: Wheel Impact Load Detector. Retrieved from LBFoster US: https://lbfoster.com/en/market-segments/rail-technologies/solutions/rail-monitoring/track-monitoring-systems

Lee, C. (2021). Assessing the Effectiveness and Efficacy of Wireless Onboard Condition Monitoring Modules in Identifying Defects in Railroad Rolling Stock. Edinburg, TX, United States: Master's Thesis, The University of Texas Rio Grande Valley.

Montalvo, J. M. (2019). Defect Detection Algorith Optimization for Use in Freight Railcar Service. University of Texas Rio Grande Valley, Mechanical Engineering. Edinburg, TX, United States: University of Texa Rio Grande Valley.

Random Forest Regression in Python. (2023, December 06). Retrieved from geeksforgeeks: https://www.geeksforgeeks.org/random-forest-regression-in-python/

Schneider, P., & Xhafa, F. (2022). Anomaly Detection. In Anomaly Detection and Complex Event Processing over IoT Data Streams (pp. 369-381).

Sergio, M. M. (2023). Vibration-Based Machine Learning models for Condition Monitoring of Railroad Rolling Stock. University of Texas Rio Grande Valley, Mechanical Engineering. Edinburg: University of Texas Rio Grande Valley.

Stewart, M. F., Flynn, E., Marquis, B., & Sharma & Associates, S. (2019, May). An Implementation Guide for Wayside Detector Systems. 99.

Tarawneh, C., Aranda, J., Hernandez, V., Crown, S., & Montalvo, J. (2020). An investigation into Wayside Hotbox Detector Efficacy and Optimization. International Journal of Rail Transportation, 8:3, 264-284, DOI:10.1080/23248378.2019.1636721.

Wikipedia. (2024, June 26). Deep learning. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Deep_learning

VITA

Kevin Quaye was born at Awoshie in the Greater- Accra region of Ghana. He attended St. Peter's Senior High School in the Eastern region of Ghana and graduated in 2016. He received his Bachelor of Science in Aerospace Engineering with First Class Honors at the Kwame Nkrumah University of Science and Technology (KNUST) in August 2020. Kevin served as a Research and Teaching Assistant at the School of Trade Training (S. of.TT) at the Ghana Airforce base from September 2020 to July 2021. After his service he taught for a while at the St. Francis of Assisi Catholic School, Awoshie-Anyaa before leaving for the States. He gained admission to a master's degree in mechanical engineering at the University of Texas Rio Grande Valley (UTRGV), Edinburg-Texas. He moved to the United States in September 2022 to begin his graduate studies under the Presidential Research Fellowship (PRF) offered by the school. Kevin worked as a graduate research assistant at the University Transportation Center for Railway Safety (UTCRS), where he learned from faculty and fellow researchers about the Railway Industry. He graduated in August of 2024 with a Master of Science degree in Mechanical Engineering at the University of Texas Rio Grande Valley. Kevin can be contacted by email at quayekevin7@gmail.com.