STATISTICAL QUALITY CONTROL VIA MACHINE LEARNING

Lappeenranta-Lahti University of Technology LUT

Master's Program in Computational Engineering, Master's Thesis

2023

Emma Hannula

Examiners: Professor Lassi Roininen
M.Sc. Victoria Jorry

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Computational Engineering

Emma Hannula

Statistical Quality Control via Machine Learning

Master's thesis
2023
44 pages, 17 figures, 7 tables
Examiners: Professor Lassi Roininen and M.Sc. Victoria Jorry

Keywords: Bearing fault detection, Machine learning, Classification, Quality control

Faults in the rolling element bearings are a common problem in rotating equipment. Therefore predicting and identifying the faults from vibration signals has gained increased attention. The goal of this thesis is to explore and test methods used in bearing fault detection. Fourier transform and envelope analysis are introduced from classical methods and decision tree, k nearest neighbor, and artificial neural network from the field of machine learning. A literature review is conducted for each method separately and other commonly used methods are introduced. As a result, each method was able to detect the faults, and good classification accuracies were achieved. Machine learning methods are dependent on feature extraction and the fault sensitivity of the feature will affect the classification ability.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Engineering Science
Laskennallinen tekniikka

Emma Hannula

Tilastollinen laadunhallinta koneoppimisella

Diplomityö
2023
44 sivua, 17 kuvaa, 7 taulukkoa
Tarkastajat: Professor Lassi Roininen ja M.Sc. Victoria Jorry

Hakusanat: Koneoppiminen, Laakerin vian tunnistaminen, Luokittelu, Laadunhallinta
Keywords: Bearing fault detection, Machine learning, Classification, Quality control

Vierintälaakereiden viat ovat yleinen ongelma pyörivissä koneissa. Siksi vikojen ennustaminen ja tunnistaminen värähtelysignaaleista on herättänyt kasvavaa kiinnostusta. Tämän työn tavoitteena on käydä läpi ja testata menetelmiä joita on käytetty laakereiden vikojen tunnistamisessa. Klassisia menetelmistä valitaan Fourier muunnos sekä verhokäyräanalyysi. Päätöspuu, K-lähimmän naapurin menetelmä sekä neuroverkko valitaan koneoppimismenetelmistä. Kirjallisuuskatsaus toteutetaan jokaisesta menetelmästä erikseen ja sen lisäksi käydään läpi muita yleisesti käytettyjä menetelmiä. Tutkimuksen tuloksena todetaan, että jokainen menetelmä onnistuu tunnistamaan viat ja saadut luokittelu tarkkuudet olivat hyviä. Koneoppimismenetelmät ovat riippuvaisia erotelluista ominaisuuksista ja ominaisuuden herkkyys vialle vaikuttaa menetelmien luokittelukykyyn.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Professor Lassi Roininen and Victoria Jorry. Big thank you also to my family for supporting me on the journey towards the Master's degree. I am very grateful to the guild as the daily coffee breaks and various conversations kept me on track. Without all the support this would have not been possible.

Lappeenranta, October 29, 2023

Emma Hannula

LIST OF ABBREVIATIONS

ANN	Artificial neural network
MLP	Multilayer perceptron
FFT	Fast Fourier transform
DFT	Discrete Fourier transform
CWRU	Case Western Reserve University
2DPCA	Two-dimensional principal component analysis
CUSUM	Cumulative sum
CNN	Convolutional neural network
KNN	K-nearest neighbors
ReLU	Rectified linear activation function
RMS	Root mean square
SNR	Signal-to-noise ratio
STFT	Short-time Fourier transform
GoFT	Goodness of fit tests

CONTENTS

1 INTRODUCTION

Rolling element bearings are one of the most broadly used elements in machines. It is built from two races which are separated by a rolling element for example balls. The purpose of the rolling element is to help reduce friction between stationary and rotating races. Due to the long operating hours, bearings are exposed to developing faults in different areas.

Faults can be divided into three main categories:
1. Fault in the inner race,
2. Fault in the outer race, and
3. Fault in the rolling element.

The location of the fault correlates with the signal it will produce. Large faults will affect the acceleration vibration by increasing the amplitude and causing irregularities in the signal. Therefore it is possible to separate and detect different faults based only on the measured signal. Commonly this is performed with time-frequency analysis which can be divided into categories of signal processing and machine learning.

Rolling bearing fault diagnosis and detection has gained increasing attention among researchers. Diagnosing faults in the early stages can help to avoid unnecessary motor failures and therefore decrease the maintenance cost of the motors. With that motivation, a wide range of methods have been proposed to classify and detect different fault types. The largest differences between these methods are mostly in signal preprocessing and frequency extraction.

Machine learning is a subset of artificial intelligence. It implies that a system has the ability to adapt and learn in a changing environment without constant supervision and intervention by a human [2]. The field of machine learning has developed significantly due to the increased interest in machine learning methods among researchers and companies. New methods have been developed and proposed for classification, pattern recognition, and feature extraction purposes. One of the main requirements for applying machine learning is the availability of large and diverse datasets. Therefore, increased interest in rolling bearing fault detection has been a motivation to develop and simulate such data.

The process of bearing fault detection begins with data acquisition implemented by sensors or simulation equipment. Signals obtained from the accelerometers are usually long time series where acceleration vibration is measured at certain time intervals. In a long time series, each time state represented is used as a predictor causing a large space of

features. A large set of features will raise the computational cost of the process, which is why data processing and feature extraction are usually used to reduce the dimensionality of the dataset and therefore make the calculations less complex.

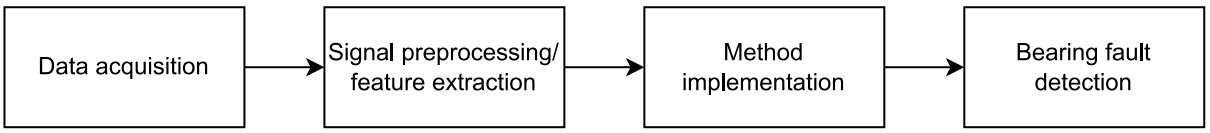


Figure 1. The process of bearing fault detection

When applying the machine learning method to any dataset, one of the major problems is how features are extracted efficiently and what features are suitable for detecting specific fault types. Therefore, the feature extraction method must be chosen carefully. Complex methods may not improve the results but make the model perform worse than simple, commonly used methods which is why their results require a closer look.

1.1 Objectives and delimitations

The purpose of this thesis is to go through commonly used methods in the field of bearing fault detection. Three popular machine learning methods and two classical methods are tested with two different datasets. The main focus is to extract commonly used features from the datasets and test them with chosen methods.

The main objectives of the thesis are:

- Compare chosen datasets and choose feature extraction methods
- Study classical and machine learning methods commonly used in bearing fault detection
- Evaluate chosen classical and machine learning methods using two datasets

The datasets used in this thesis are long time series and both datasets have one signal representing one fault. In order to use machine learning methods, datasets need to be large and diverse. Currently, only a few publicly available datasets are available which limits the possibilities of testing different methods. The quality of the dataset is not visible without a thorough data exploration.

Dividing large signals into smaller samples can be conducted in multiple different ways, especially regarding the chosen length of the samples. Choosing the sample size can have an impact on the properties of the original signal which is why the sample size must be chosen carefully. Another property affecting the properties is implemented feature extraction methods. The features can be extracted with a wide range of various signal-processing methods in the time-, frequency- and time-frequency domains. The feature extraction in this thesis is limited to time and frequency domains with commonly used extraction methods.

1.2 Structure of the thesis

This thesis is divided into seven chapters. In Chapter 2, datasets used in experiments are introduced and preprocessed. Classical bearing fault detection methods are introduced in Chapter 3. Chapter 4 focuses on common machine learning methods used in the field in question. In Chapter 5, different feature extraction methods are introduced and features are extracted from the datasets. In Chapter 6, previously introduced machine learning methods are tested with extracted features. Chapter 7 contains conclusions from the conducted experiments and their value for future research.

2 Dataset exploration and pre-processing

In this thesis, two datasets are used to test and compare chosen methods for bearing fault detection. Both datasets consist of bearing acceleration time series with multiple faulty and one normal bearing signal. Depending on the dataset, some faults are more easily separated than others.

The LUT Department of Mechanical Engineering has kindly provided the first dataset. The data was produced by simulating bearings with MATLAB Simulink and the bearing used in the simulation was a deep groove ball bearing 6010 from SFK. The dataset has not been previously used in any publicly available publications and is used in this thesis to demonstrate the performance of chosen methods with simple and easily distinguishable data.

The second dataset is provided by Case Western Reserve University (CWRU) bearing data center [1]. The dataset is acquired by inducing different-size faults into three locations of the bearing. Signals are collected by placing them into a test motor and using an accelerometer to measure the vibration. The bearing used in the experiment is a deep groove ball bearing 6205 from SFK. The CWRU dataset is a publicly available dataset that has a lot of variation. It is also one of the most commonly used datasets in bearing fault detection. Multiple classical and machine learning methods have been tested with the dataset with varying outcomes [18, 29]. Therefore the achieved results can be compared to the results in previous publications.

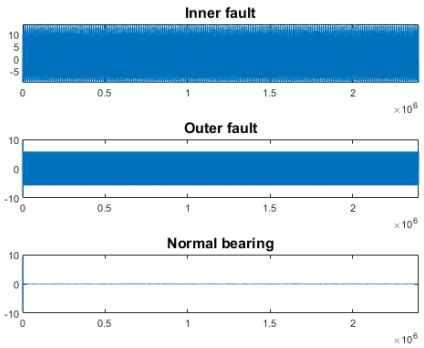
2.1 LUT rolling bearing dataset

The dataset consists of bearing acceleration data in 2.4 million time points with different types of faults; fault on the inner side and fault on the outer side. Also, data without fault was recorded. Data was collected with 12000 samples per second and with a rotating speed of 900 rpm (revolutions per minute). Later this Dataset will be referred to as "Dataset 1".

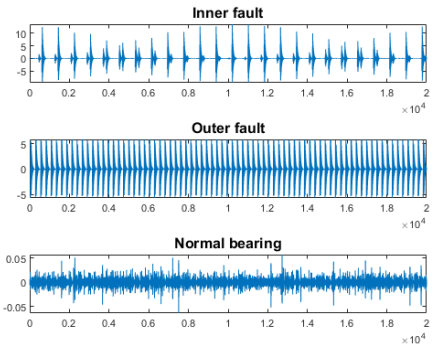
The dataset is visualized in Figure 2. The vibration of the normal bearing is very small without any large changes or irregularities in amplitude. Inner and outer race faults have large variations in amplitude. Amplitude spikes in the inner fault are roughly double the size of outer race fault spikes. Overall both inner and outer fault signals have a steady

repeating pattern throughout the signal.

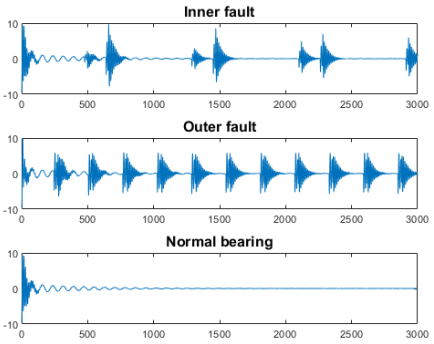
Figure 2c shows that in the beginning, acceleration differs in every case from the rest of the signal. The effect is the most visible in normal bearing where the amplitude in the beginning is 10 instead of 0.05 which remains in the rest of the signal. After roughly 2000 timesteps, the signal will become steady in all cases without any major changes.



(a) Complete signals



(b) Form of the signals

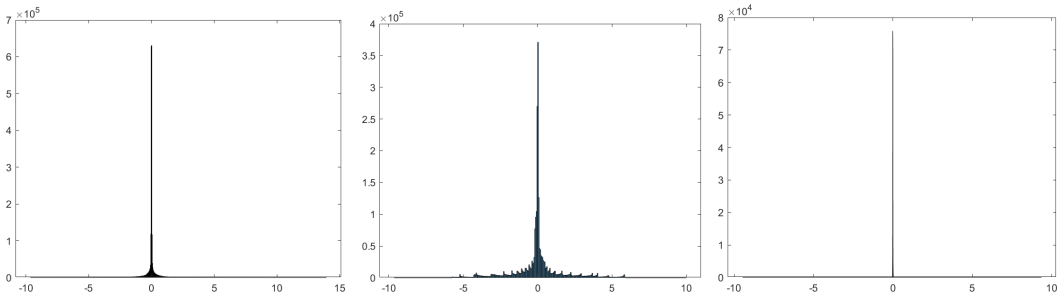


(c) Beginning of the signals

Figure 2. Signals of the Dataset 1

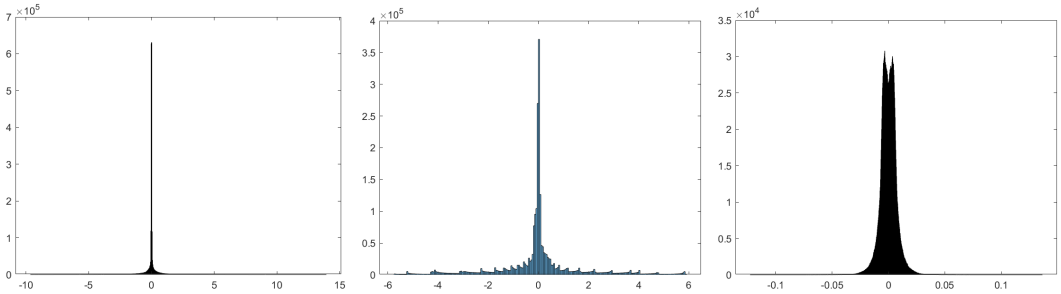
Histograms of the raw signals are in Figure 3. and histograms excluding the irregular beginning of the signals are in Figure 4. The difference between raw and partial histograms is most visible in normal bearing in which the vibration in the beginning is affecting the tails of the deviation. Removing the beginning brings out the shape of the signal. Inner and outer faults are not as clearly affected by the irregularities in the beginning.

In partial histograms, it can be noticed, that the normal bearing has the smallest standard deviation even though there are two clear spikes on both sides of zero. The tails of the distributions are larger in inner and outer race faults but in both, most values are placed



(a) Inner fault (b) Outer fault (c) Normal bearing

Figure 3. Histograms of the Dataset 1



(a) Inner fault (b) Outer fault (c) Normal bearing

Figure 4. Partial histograms of the Dataset 1

around the zero. Even though both faulty distributions are closely zero-focused, there is a small number of data points on a wide area on both sides of the big spike.

2.2 CWRU seeded fault test data

The test motor is supported by two bearings, one in the drive end and one in the fan end. Measurements with faulty bearings were made on both ends. Collected signals have an average length of 120000 with few exceptions. The signal obtained from normal bearing differs from the rest of the data because its length is 240000.

Faults induced to the bearings were sizes of 0.007 inch, 0.014 inch, 0.021 inch, 0.028 inch, and 0.040 inch. Because the outer race is stationary, measurements were made with different fault placements relative to the load zone (centered 6 o'clock, orthogonal 3 o'clock, and opposite 12 o'clock). 3 o'clock and 12 o'clock outer race data from the 0.014-inch fault were not available.

A total of 13 faulty signals were recorded with motor loads from 0 to 3 horsepower. Each

motor load has a separate speed varying from 1797 to 1720 rpm. Measurements were made in both the fan and drive end of the motor with 12000 samples per second. Drive end bearing faults were also collected at the speed of 48000 samples per second.

In this thesis, we will be using only 12000 samples per second drive end data with 0 hp motor load. Also, only faults from the size of 0.007 inches to 0.021 inches are used because the bearings used in the two largest faults were from different manufacturers. All signals used in the Dataset are listed in Table 1. Later this dataset will be referred to as "Dataset 2".

Table 1. Fault types of Dataset 2

Index	Fault location	Fault diameter (10^{-3} inch)
1	ball	7
2	ball	14
3	ball	21
4	inner race	7
5	inner race	14
6	inner race	21
7	outer race @12	7
8	outer race @3	7
9	outer race @6	7
10	outer race @6	14
11	outer race @12	21
12	outer race @3	21
13	outer race @6	21
14	-	-

The raw dataset is visualized in Figure 5. The amplitudes of signals differ between samples. In normal bearing, the amplitude is a maximum of 0.2. Most of the faulty signals have amplitude between 2 and 5. The closest signal to the normal bearing is with a 0.007-inch fault in the ball and inner race. Unlike Dataset 1, Dataset 2 does not have a significant change during the first seconds of the measurements.

2.3 Data pre-processing

In Dataset 1, the beginning of the signals are different from the rest of the dataset due to the building of acceleration in the simulation. As noticed in data exploration, the first 2000 samples of each signal differ from the rest of the signals and are therefore removed

to exclude their influence on experiences. Exploration of Dataset 2 did not reveal any irregularities at the beginning of the signal. Therefore data can be used as a whole and there is no need to remove any data points.

Both introduced datasets consisted of a single time series for each fault type. To get large enough datasets for machine learning methods, each signal needs to be divided into subsamples. Both Datasets are divided into samples with a length of 500 which grants enough training samples for both datasets. Dataset 1 has larger raw signals which results in a larger sampled Dataset than for Dataset 2. As a result, we get Dataset 1 with a size of 14325×500 with 3 classes, and Dataset 2 with a size of 3662×500 with 14 classes.

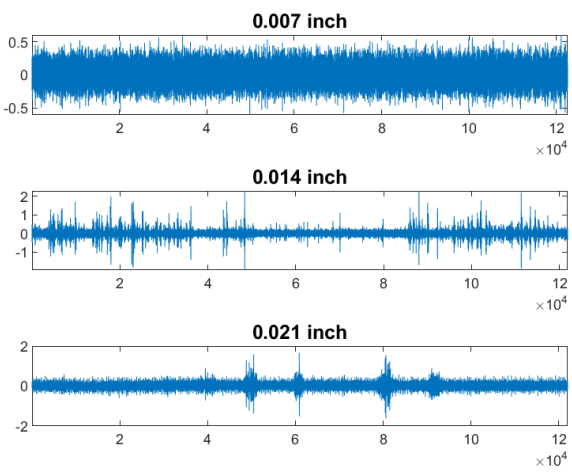
The large number of samples in a class will create a better classification model for the class. Because Dataset 1 has more samples and fewer classes, it is simpler and easier to classify. The smaller amount of data in each class in Dataset 2 might affect the creation of the model.

2.4 Dataset quality

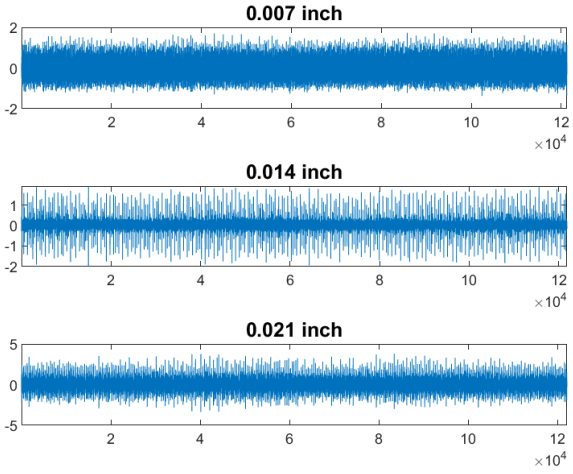
The CWRU dataset was produced by artificially inducing varying sizes of faults in different parts of the bearing and placing them into a test setup. The LUT dataset was simulated with MATLAB, which makes it an example of easily separatable bearing faults. In the simulation, only one fault type was tested in both the inner and outer races.

Because of the ideal laboratory environment, the dataset collected will differ from data collected in the actual engineering environment [7]. Laboratory-produced data is almost noiseless and faults are focused on specific areas making the faults easy to separate. In real-life situations, a bearing can have multiple faults with different severity levels making the specific fault signal harder to isolate.

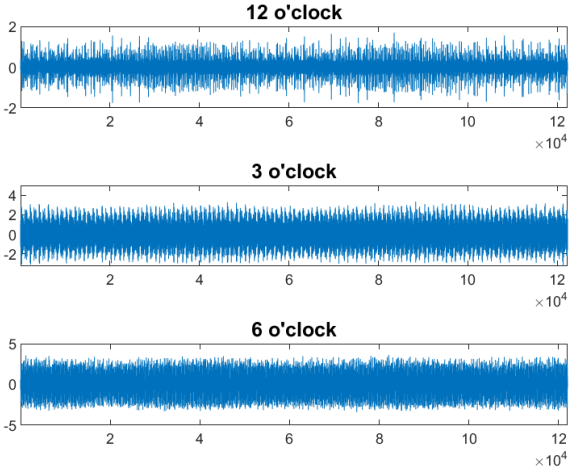
In machine learning data has to be balanced to train models correctly. A too-large difference in class sizes can result in sufficient representation of the smaller class [11]. The lengths of signals in Dataset 1 are approximately the same. The problem is with Dataset 2 where the normal signal is almost double the length of any fault signal. Therefore the normal bearing will have a much larger sampled dataset than any of the faulty bearings.



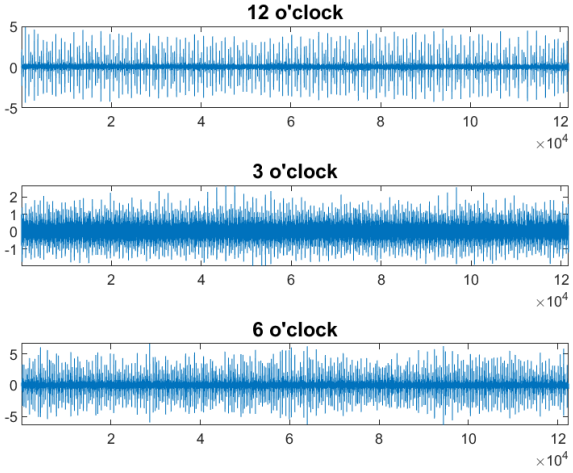
(a) Fault in ball



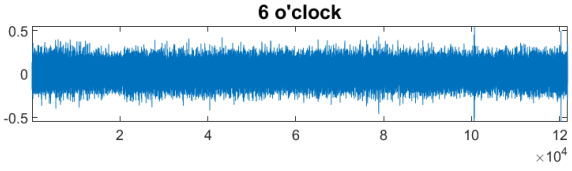
(b) Fault in inner race



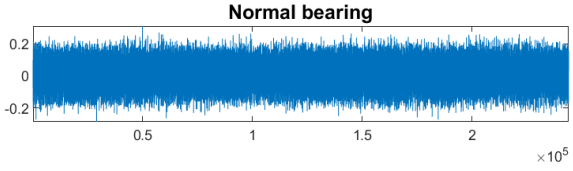
(c) 7 inch fault in outer race



(d) 21 inch fault in outer race

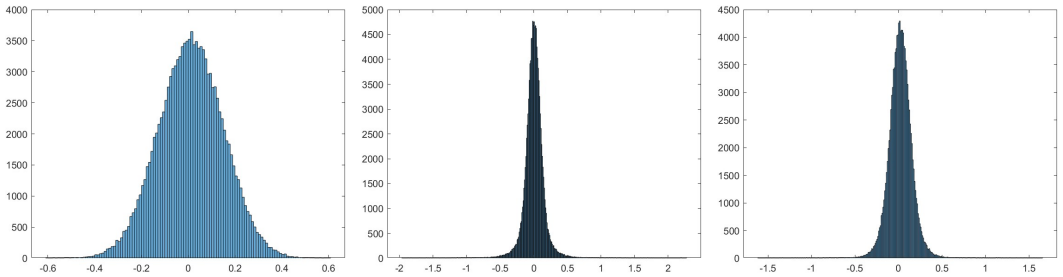


(e) 14 inch fault in outer race

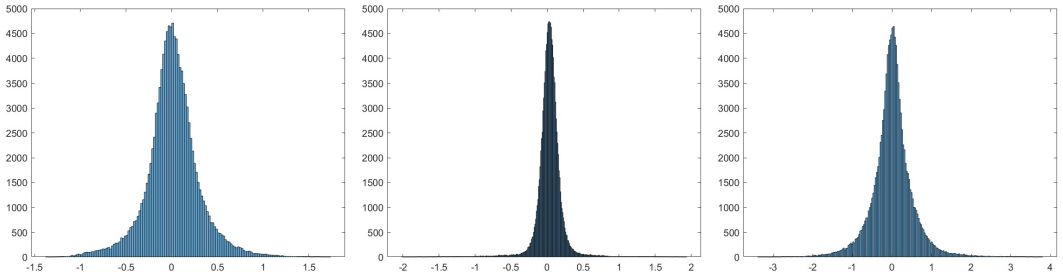


(f) Normal bearing

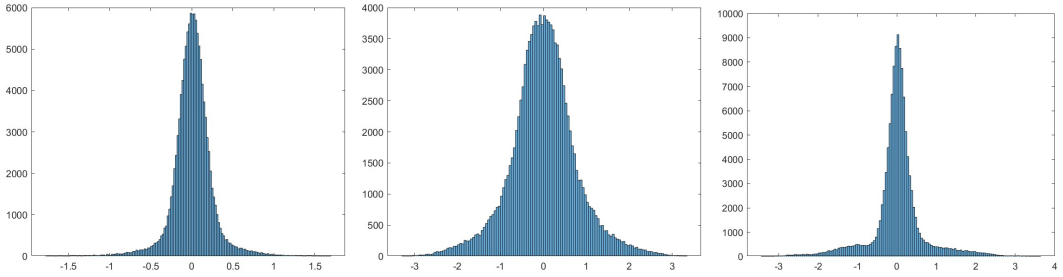
Figure 5. Signals of Dataset 2



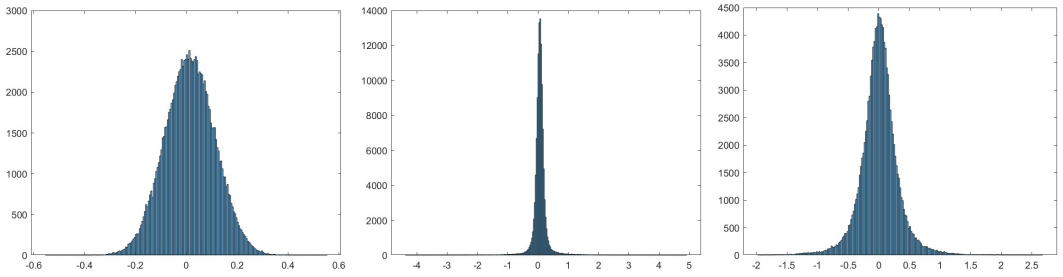
(a) Fault number 1 (b) Fault number 2 (c) Fault number 3



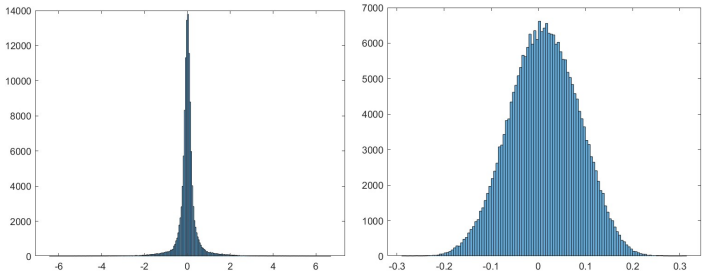
(d) Fault number 4 (e) Fault number 5 (f) Fault number 6



(g) Fault number 7 (h) Fault number 8 (i) Fault number 9



(j) Fault number 10 (k) Fault number 11 (l) Fault number 12



(m) Fault number 13 (n) Normal bearing

Figure 6. Histograms of Dataset 2

3 Classical methods

In previous studies, many analysis techniques have been proposed for bearing signals, see for example [18, 25]. Around the 1970s, one of the first papers was published on bearing diagnostics by Balderston [4]. The goal of the research was to determine if a change in frequency spectrum would indicate the onset of a bearing failure. Spectral density plots were obtained with a frequency analyzer and the plots were analyzed by hand comparing results with calculated rotational frequencies.

After recognizing the connection between bearing faults and resonance in high frequency, synchronous averaging was used to expose the faults [25]. Envelope analysis was developed to reduce the range of high-frequency analysis.

3.1 Fourier transform

Fourier transform is one of the most common methods used in signal processing. It transforms the signal into a frequency domain by composing the existing frequencies by intensity which enables frequency analysis tools to be used. When the continuous signal is replaced with a set of samples, discrete Fourier Transform (DFT) is applied [23]. DFT is defined as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}. \tag{1}$$

Inverse discrete Fourier transform is defined as

$$x_n = \frac{1}{N} \sum_{k=0}^{N} X_k e^{i2\pi kn/N}. \tag{2}$$

Due to the computational complexity of DFT, Fast-Fourier-Transform (FFT) was developed to lower the complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ [12]. FFT is mostly used variation of Fourier transform when transforming time signals into frequency domains.

Multiple methods implementing the Fourier Transform and its variations have been introduced in the field of bearing fault detection [19]. Fourier Transform is not usable as

a diagnostic method but as a step to achieve frequency-domain signal in other analysis methods.

One of the methods utilizing the FFT spectrum of the CWRU dataset was represented by Li et al. [18]. Experiments were made only with two fault sizes with different loads. They extracted FFT spectrum images from the signals and processed them with two-dimensional principal component analysis (2DPCA). Principal components acquired by 2DPCA were used to classify faults with the minimum distance method. Depending on the location of the fault and motor load achieved classification accuracy varied from 75% to 100%.

3.2 Envelope analysis

Faults in bearings affect the vibration signal by increasing the amplitude of specific frequencies. With enough information about the bearings, fault-specific vibration frequencies can be calculated separately. Therefore extracting and processing the frequencies where the fault could occur, can bring out the fault repetition frequencies in the envelope spectrum [25]. The fault frequencies might be affected by the noise, which can make them harder to identify.

The first step in the envelope analysis is to transform the time-domain signal to the frequency-domain with FFT. The envelope signal is then created by filtering the wanted frequency range and enveloping it for example with the Hilbert transform [10] which is defined as

$$\mathcal{H}[g(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(t-\tau)}{\tau} d\tau = g(t) * \frac{1}{\pi t}, \tag{3}$$

or with other demodulation techniques like bandpass filtering. The envelope spectrum is formed by taking the FFT from the enveloped signal. The peaks of the spectrum can then be compared against fault-specific frequencies. The frequency for the inner bearing fault is defined as

$$f_I = \frac{N}{2} f_0 \left(1 + \frac{D_p}{D_b} \cos(\alpha)\right), \tag{4}$$

where f_0 is the rotational speed, D_p is the diameter of the pitch and D_b is the diameter of the ball. The contact angle to the bearing is α which is assumed as zero. For outer fault, the fault frequency is defined as

$$f_O = \frac{N}{2} f_0 \left(1 - \frac{D_p}{D_b} \cos(\alpha)\right). \tag{5}$$

The performance of the envelope spectrum is affected by the demodulated band. Choosing the correct frequency band has been a big challenge but it has been solved with spectral kurtosis and kurtogram [25]. Spectral kurtosis is a statistical parameter that indicates how the impulsiveness of a signal changes with frequency. Therefore it will receive high values in frequencies where bearing fault is present. Spectral kurtosis is calculated from the short-time Fourier transform (STFT) where the local Fourier transform is calculated by moving a window along the signal. STFT is defined as

$$S(t, f) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} d\tau, \tag{6}$$

from which spectral kurtosis is calculated with

$$K(f) = \frac{\langle |S(t, f)^4| \rangle_t}{\langle |S(t, f)^2| \rangle_t^2} - 2, f \neq 0, \tag{7}$$

where $\langle \cdot \rangle$ is a time-average operator. Kurtogram is used to visualize spectral kurtosis as two-dimensional with different window lengths for the STFT.

Zhao et al. [31] implemented envelope analysis to vibration and residual signal. The purpose of the residual envelope was to remove the influence of characteristic frequencies, which is a common problem in outer race fault detection. The filtering will increase the signal-to-noise ratio (SNR) of the fault frequency and make fault detection easier.

3.3 Other methods

While Fourier transform and envelope analysis are a couple of the most commonly used and well-performing methods in bearing fault detection, multiple different methods also

have been introduced [19, 25].

Lopez-Gutierrez et al. [21] implemented empirical wavelet transform to a CWRU outer race fault. Empirical wavelet transform decomposes the signal into multiple components which can be used to extract amplitude and frequency-modulated components with Fourier transform. Three Goodness of Fit tests were applied to the achieved modes, Kolmogorov–Smirnov Test, Kuiper test, and Pearson Chi-Square Test. The method was able to detect the damage in the bearing.

Another proposal applying the same Goodness of Fit test to bearing fault detection is presented by Avina-Corral et al. [3]. The tests are applied to motor current and squared-current signal analysis. The Kuiper test is able to detect faults from the motor square-current signal. Also, other tests produced decent results for both signals. The conclusion stated that the Goodness of Fit tests are capable of working along other bearing fault detection methods.

Cumulative control charts have been used to detect small changes in the signal. Upper and lower cumulative sums are calculated from the signals and plotted with control limits. Samples outside the limits are flagged as faulty. Jawad and Jaber [14] proposed the method for bearing fault detection. They extracted time-domain features from the signal and marked standard deviation as the most sensitive to the fault and it was used as a base for the control chart. The method was able to detect small-size bearing faults but it is dependent on the extracted time-domain feature.

4 Machine learning methods

There have been multiple studies about machine learning and how it could be applied in bearing fault detection and classification, see for example [11, 17, 22]. Applied methods varied from basic decision trees and K-nearest neighbors to convolutional neural networks (CNN).

Different algorithms and methods have been tested with CWRU Dataset with varying results [22, 29]. The biggest difference between articles is how the Dataset is preprocessed. There are different styles of dividing the large signal into smaller samples and what features are used. Because the Dataset is large, many researchers choose to use only part of the data. Due to the nature of the data, this could cause differences between applied methods. The latest research focuses on whether it is possible to make classifications only using the raw signal [26].

Every machine learning method involves a learning process. This process can be divided into two main categories: supervised and unsupervised. Supervised learning uses labeled datasets and is mostly used in classification and regression tasks. Unsupervised learning tries to find patterns and similarities in the data and fits best to clustering problems.

4.1 Decision tree

A decision tree is one of the basic supervised decision-making methods. It is formed by nodes from which the node that has no incoming edges is called root and nodes that have no outgoing edges are called leaves [20]. Each leaf will represent a class and starting from the root, each node has assigned a feature or a variable. Based on the variable of the node, branches are assigned a value/range of the values from the variable. The complexity of the tree can be controlled for example by assigning a maximum number of leaves or nodes. An example of a simple decision tree that is built from three nodes and four leaves is in Figure 7.

A classification tree is a simple classification method that uses features to construct an accurate model. The problem is that decision tree models are easily overfitted and thus badly generalized [17]. Therefore careful data preprocessing and feature extraction are needed to get appropriate results from the algorithm.

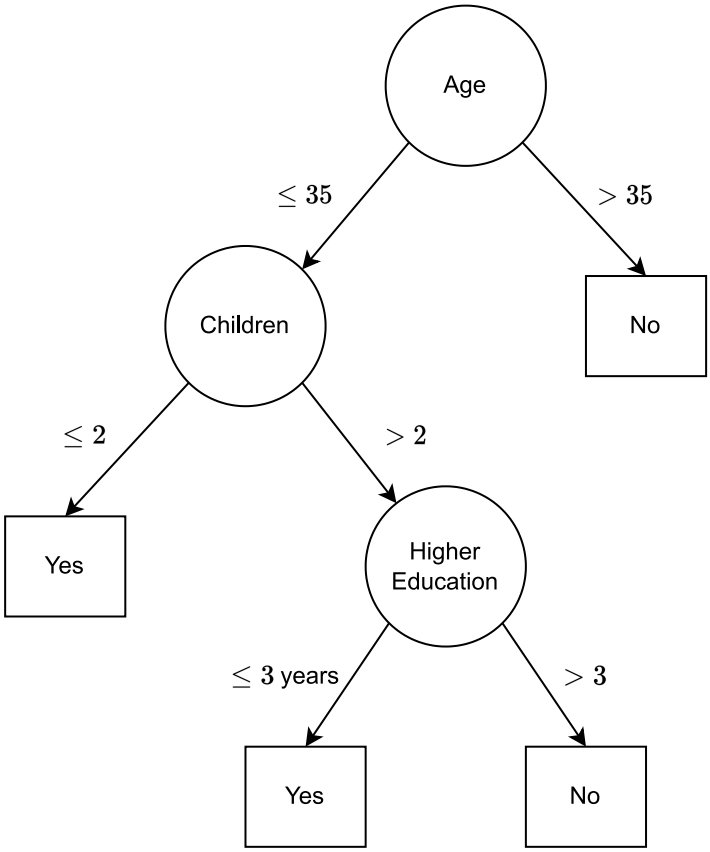


Figure 7. A simple example of a decision tree.

Euldji et al. [9] implemented a decision tree C4.5 algorithm for bearing fault detection. 15 time-domain features were extracted and used in classification. An accuracy of 94% was achieved when using all features. Reducing the amount of features by careful selection accuracy increased to 100%.

4.2 K-nearest neighbors

K-nearest neighbors (KNN) is a non-parametric supervised learning model. The model evaluates samples based on their calculated distance from other samples. The sample is classified into the class that has the most samples within the number of neighbors. The number of closest neighbors is limited with the parameter k which is defined by the user.

The distance can be determined for example with Euclidean or Mahalanobis distance. Euclidean distance calculates the distance between two samples and is defined as

$$D(X,Y) = \sqrt{\sum_{i=1}^{N}(X_i - Y_i)^2}. \tag{8}$$

Mahalanobis distance calculates the distance between a sample and a distribution and is given as

$$D = \sqrt{(y-\mu)^T \Sigma^{-1} (y-\mu)}, \tag{9}$$

where μ is the mean and Σ is the covariance matrix of the distribution.

KNN is easy to implement but for large datasets computational cost, the balance of data, and determination of parameter k could affect the performance [17]. High dimensionality becomes a problem when the distance vector is equal in length for each sample. The algorithm is also prone to overfitting which can be solved by feature extraction and dimensionality reduction example with principal component analysis.

In the field of bearing fault detection, Kumar and Upadhyaya [16] introduced a method combining wavelet transform and KNN. Wavelet transform was used to compose signals from which overall 17 time-domain features were calculated. These features were used in the KNN algorithm resulting in 83% accuracy.

Sharma and Jigyasu [27] presented weighted KNN in bearing fault detection. Experiments were done with multiple k-values and different distance metrics. Extracted time-domain features were normalized and then classified with weighted KNN. Classification accuracy varying from 97.3% to 100% was achieved with k values from 9 to 33 and Mahalanobis distance. Implementing PCA did not improve the results.

4.3 Artificial neural network

Artificial neural network (ANN) constructs from nodes and layers. The basic neural network is built from three different layers. The input layer is for receiving the data, the hidden layer is for data processing, and the output layer is for results. Several hidden layers are related to the complexity of the neural network. Simple ANN has only one hidden layer making the complete neural network 3 layers deep. An example structure of

a simple ANN is in Figure 8.

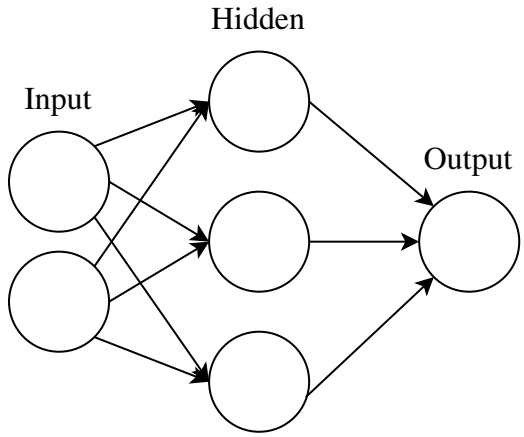


Figure 8. Simple neural network with one hidden layer

Each layer has several nodes which are connected. These connections are assigned different weights during the training process. Depending on the type of input layer and hidden layers, neural networks have different types, for example, convolutional neural network, wavelet neural network, and graph neural network [17, 28].

In a standard neural network structure, the network consists of fully connected layers which are followed by activation layers. The size of fully connected layers and the type of activation layer can be specified during the training.

In hidden layers, often used activation layers are rectified linear activation function (ReLU), Sigmoid, and Tahn. The selection of the used layer is the choice of the user, but the purpose and the size of the network can help in the decision.

The ReLU layer has become a default activation function in neural networks because networks where it was used usually perform better and faster compared to other activation functions [5]. The principle of ReLU is that it will output the input directly if the input is positive. In other cases, the output will be zero. Therefore the computation is efficient even in complex problems.

For multiclass classification, softmax is the most preferred as an output activation function. It normalizes the input vector into a probability vector of the same size. Softmax is defined as

$$S(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}, \tag{10}$$

where n is the number of classes.

4.4 Deep learning models

When a neural network has more than 3 hidden layers, it is called a deep neural network which belongs to the field of deep learning. Deep learning models are neural network-based models that can solve more complex and larger problems than traditional neural networks. One of the most common applications for deep learning is image classification as the models extract important features during the training. Automated feature extraction reduces the amount of human interference and makes the models more accurate as the features can be optimized.

Deep learning was originally invented in the 1980s, but implementing it was not possible [24]. Deep learning models require large datasets and a lot of calculation power which have been previously limited. The latest progress in computer science and the amount of new data have increased significantly and allowed the use of more complex methods.

Shao et al. [26] implemented an ensemble deep auto-encoder for bearing fault detection. A deep auto-encoder is an unsupervised feature learning model that is constructed from multiple auto-encoders. Auto-encoder is a three-layer neural network that uses raw data as input and whose goal is to minimize the reconstruction error between input and output data. Ensemble deep auto-encoder is constructed from multiple deep auto-encoders for which Shao used different activation functions. The proposed method was compared to individual deep-learning methods by using the CWRU dataset. The average classification accuracy achieved was 99.15% which was better than any other compared method.

Chen et al. [8] employed three deep neural networks to identify bearing faults; Deep Boltzmann Machines, Deep Belief Networks, and Stacked Auto-Encoders. All models were tested with raw input signals and with various preprocessed data. Deep Boltzmann Machines are constructed from multiple Boltzmann Machines which are built from symmetrically coupled stochastic units. Deep Belief Network is a combination of Boltzmann Machines and belief networks. Preprocessing was conducted by extracting time-, frequency- and time-frequency-domain features. All three models were able to predict

faults with roughly 99% accuracy when using frequency- or time-frequency-domain features.

Xiao et al. [28] proposed a graph neural network for bearing fault detection using the CWRU dataset. Extracted time- and frequency-domain features were transformed into graphs by calculating similarity and assigning weights for samples with the highest similarity. The graph is formed by connecting samples with weights larger than 0 and is finally used as an input for the network. The best average classification accuracy achieved was 99.1%.

Neupane and Seok [22] reviewed multiple publications where CWRU and a few other datasets were used with deep learning methods. They compared research made with example auto-encoders, convolutional neural networks, and deep belief networks. The review concludes that the CWRU dataset requires more benchwork study and the data quality is a big challenge.

5 Feature extraction

In most of the methods used in machine learning, data cannot be used in the form it is measured, because of its size. Information needs to be extracted as numerical features while preserving the nature of the original data. Extracting features reduces the dimensionality of the data which has a straight correlation to the complexity of the calculations.

Previous studies have shown that extracted features contribute to classification accuracy [9]. Features can be extracted in time-domain, frequency-domain, or time-frequency-domain [17]. The number of possible features can lead to increased computational cost which feature extraction is trying to solve. Some features are more sensitive to the fault than others therefore only sensitive features must be chosen.

5.1 Time-domain features

Time-domain features are basic statistics calculated from the signals themselves. They can include example mean, standard deviation, skewness, and kurtosis among many others [13]. Many time-domain features have been previously used with multiple machine learning methods [15, 17].

The mean is one of the basic features of the signals. Another varying method is a root mean square (RMS) which can be used to represent the vibration level. RMS calculates the mean magnitude of the signal disregarding their direction. In bearing fault detection, RMS might be a more presentive feature, but in our experiments, a standard mean is used.

Standard deviation tells how far on average values are from the mean. It is defined as

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})^2}, \tag{11}$$

where $\tilde{x}$ is mean and n number of samples. Asymmetry in probability density is expressed with skewness [6]. When a distribution is normally distributed, the skewness value is approximately zero. Skewness is calculated with

$$S = \frac{1}{n-1}\sum_{i=1}^{1}(x_i - \tilde{x})^3, . \tag{12}$$

Kurtosis is a measurement that indicates the size of distribution tails. It is an important indicator in bearing fault diagnostics. Kurtosis is defined as

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})^4}{\sigma^4}. \tag{13}$$

Average time-domain features were calculated from both Datasets to show the trend of each feature in different faults. For classification purposes, features were also calculated separately for each sample introduced in 2.3, to form large enough training datasets.

Results for Dataset 1 are shown in Table 2. Data from the normal bearing is very closely mean-centered and has a small standard deviation. The standard deviation in the inner and outer faults have only a small difference meaning the samples are similarly distributed around the mean. Kurtosis values indicate that the tails of the distributions have large differences and the inner fault has more values far from the mean.

Table 2. Time-domain features of the Dataset 1 for each fault

	Standard deviation	Skewness	Kurtosis	Mean
Inner fault	1.333	0.892	24.507	-1.247$\times 10^{-7}$
Outer fault	1.667	0.147	5.858	-1.902$\times 10^{-7}$
Normal bearing	0.007	6.887$\times 10^{-3}$	9.736	-3.820$\times 10^{-9}$

For Dataset 2, average time-domain features are listed in Table 3. As seen, standard deviation, mean and skewness have no significant difference between faults. Kurtosis has large differences between samples indicating that the tails of the distributions have large variation.

5.2 Frequency-domain features

Frequency-domain features can be extracted with multiple methods for example empirical mode decomposition, wavelet transform and Fourier transform [18, 30]. The Fourier transform is the most popular and widely used but offers limited information for example

Table 3. Time-domain features of the Dataset 2 for each fault

Index	Standard deviation	Skewness	Kurtosis	Mean
1	0.139	-0.009	2.985	0.013
2	0.153	0.225	17.769	0.005
3	0.134	0.033	8.549	0.021
4	0.291	0.164	5.396	0.013
5	0.195	-0.059	21.957	0.034
6	0.525	0.302	7.445	0.014
7	0.237	-0.038	8.053	0.011
8	0.769	0.066	4.239	0.007
9	0.669	0.057	7.649	0.023
10	0.100	0.001	3.056	0.014
11	0.383	-0.324	34.950	0.065
12	0.315	0.086	6.947	0.008
13	0.583	0.105	21.006	0.005
14	0.073	-0.035	2.764	0.013

of the location of the frequencies present.

Frequency-domain features used in this thesis were extracted with MATLAB Diagnostic Feature Designer. First spectral estimation of the original time-domain signal is calculated with the auto-regressive model. From the received spectrum data, peak amplitude, peak frequency, and band power are calculated.

Features extracted from the Dataset 1 are listed in Table 4. Both faulty signals have large peak amplitude while the normal bearing has a very small peak. The location of the peak is similar in all of the signals. The difference in bandpower indicates that the size of the peak is largest in the outer fault and smallest in the normal bearing.

Table 4. Frequency-domain features of the Dataset 1

	Peak Amplitude	Peak Frequency	Bandpower
Inner fault	59.970	0.091	0.896
Outer fault	51.570	0.087	1.396
Normal bearing	0.009	0.092	2.739×10^{-5}

In Table 5 are listed features calculated from Dataset 2. The values of peak amplitude have large differences, but the highest peaks are with faults 6 and 8. Peak frequency values are similar in each fault excluding the normal bearing. The variation in bandpower is also small. but different faults can be separated. The normal bearing has the lowest values in each feature.

Table 5. Frequency-domain features of the Dataset 1

Index	Peak Amplitude	Peak Frequency	Bandpower
1	0.667	0.272	0.010
2	0.633	0.265	0.012
3	0.313	0.272	0.009
4	0.648	0.279	0.043
5	0.363	0.275	0.020
6	10.570	0.252	0.140
7	1.948	0.282	0.028
8	21.889	0.230	0.299
9	6.211	0.279	0.226
10	0.085	0.293	0.005
11	3.444	0.275	0.076
12	0.581	0.257	0.049
13	5.718	0.274	0.171
14	0.033	0.084	0.003

6 Experiments and results

Classical and machine learning methods introduced in Chapter 3 and 4 are tested with datasets from Chapter 2. Classical methods are tested with raw signals, but extracted features are needed for machine learning methods.

Features for machine learning calculated in the previous Chapter, are divided into two models. Model 1 is constructed from a total of 4 features in the time domain. Calculated averages of used features are listed in Tables 2 and 3. In Model 2, the first model is expanded by adding 3 frequency-domain features making the Model 2 a size of 7 features. Averages of features are listed in Tables 4 and 5.

Table 6. Features of models used with machine learning methods

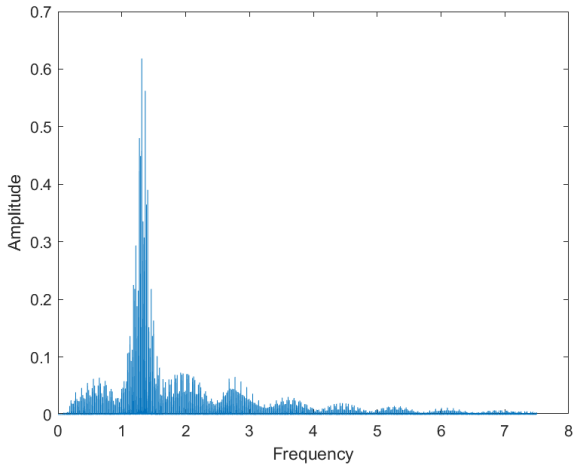
	Features	
Model 1	Standard deviation	Kurtosis
	Skewness	Mean
Model 2	Standard deviation	Mean
	Skewness	Peak Amplitude
	Kurtosis	Peak Frequency
	Bandpower	

Classification results of machine learning methods are validated with K-fold cross-validation with 5 folds. K-fold cross-validation will divide the dataset randomly into training and testing sets K times. Both datasets are standardized during the training process which is defined as

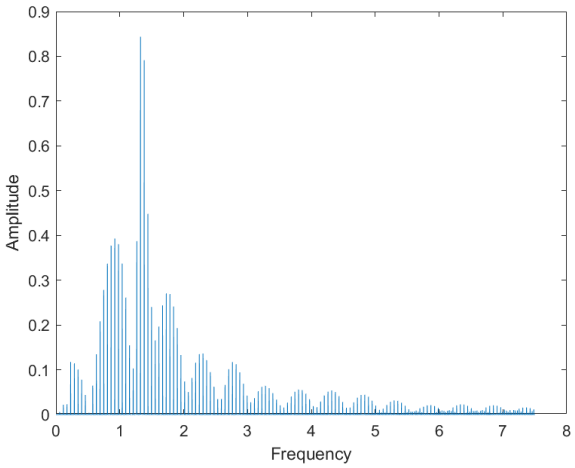
$$z = \frac{x - \mu}{\sigma}. \tag{14}$$

6.1 Fourier transform

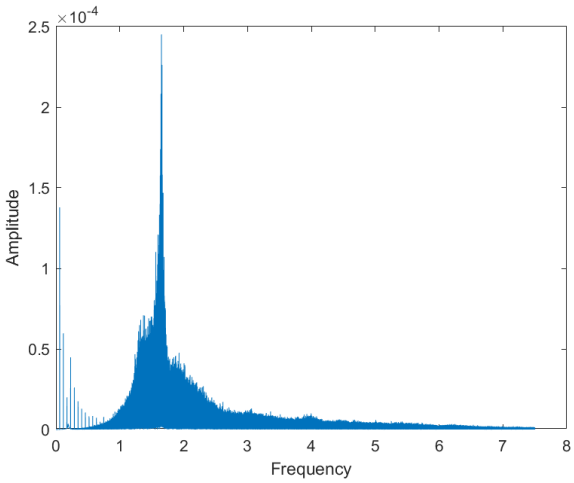
Fourier transforms are calculated from both datasets using MATLAB function `fft`. The FFT spectrum of Dataset 1 is presented in Figure 9. The shapes of the spectrum are similar in all three cases. Frequencies in normal bearing are more focused on the same area and the maximum amplitude is significantly larger than in faulty spectrums. The inner and the outer fault spectrums are mostly similar in size and shape. The difference is in the width of the large spike which can be visually divided into three separate spikes.



(a) FFT spectrum of inner fault



(b) FFT spectrum of outer fault

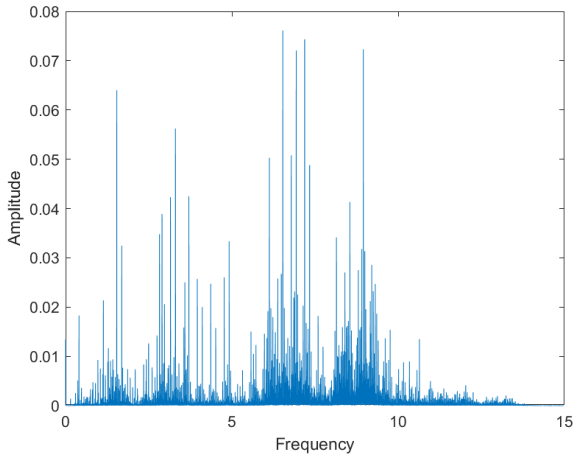


(c) FFT spectrum of normal bearing

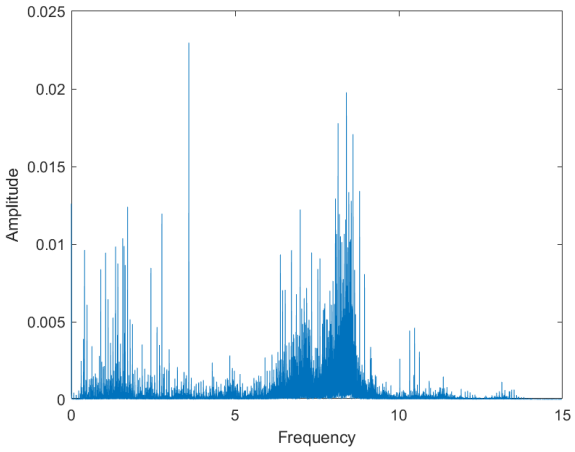
Figure 9. FFT spectrum of Dataset 1

In Figure 10 are FFT spectrums from Dataset 2. Only one fault is chosen from the ball, inner race, and outer race fault type. The spectrum of normal bearing is focused on lower frequencies and has only two thin spikes. Spectrums obtained from faulty signals are distributed to a larger area. The inner and outer fault spectrums have one wide spike in the frequencies between 5 and 10. The ball fault does not have any large and wide spikes but has a lot of thin thin spikes in almost every frequency.

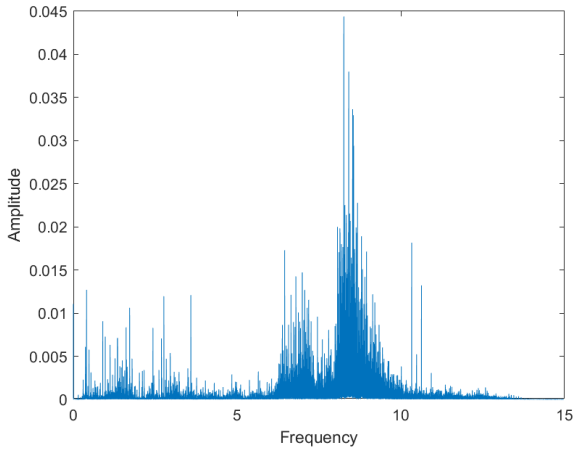
The spectrums obtained from the LUT dataset are not as clearly separatable as in the CWRU dataset. The CWRU dataset has more variation in existing frequencies and amplitudes. The largest difference between the LUT spectrums is in the amplitude of the most common frequency.



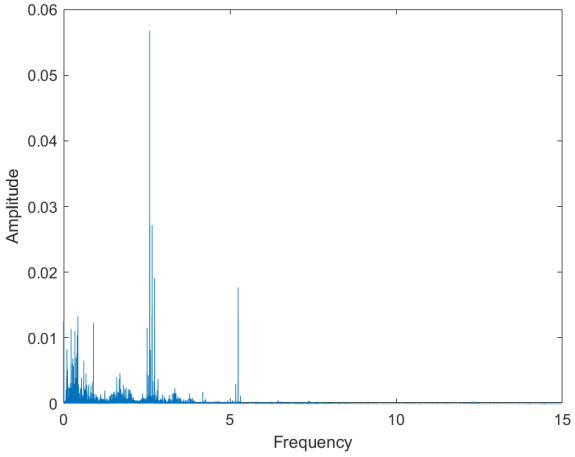
(a) FFT spectrum of ball fault number 1



(b) FFT spectrum of inner fault number 4



(c) FFT spectrum of outer fault number 7



(d) FFT spectrum of normal bearing

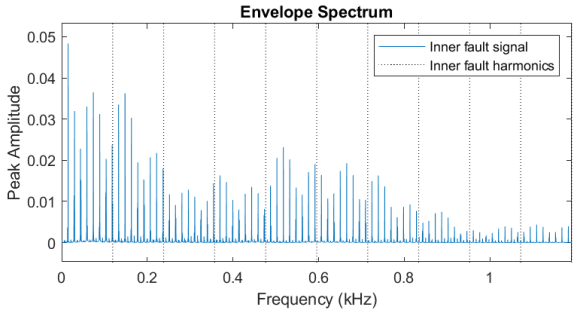
Figure 10. FFT specrum of Dataset 2

6.2 Envelope analysis

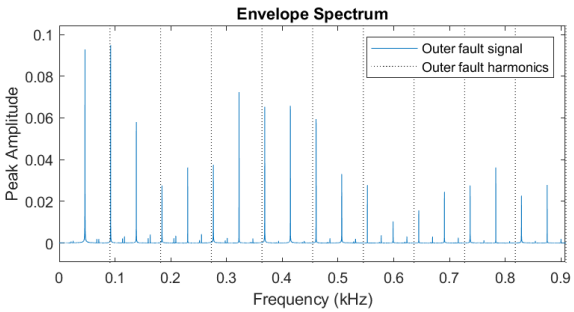
The envelope spectrum is created with MATLAB function `envspectrum` which calculates the envelope signal and spectrum from the time-domain signal. Inner and outer fault-specific frequencies are also calculated for both datasets according to Equations 4 and 5.

The spectrum and fault frequencies of inner and outer faults for Dataset 1 are visualized in Figure 11. The outer fault frequencies are clearly matching to the harmonics. In the inner fault, frequencies do not match the harmonics which can be due to the nature of the data. Also in the outer fault, harmonics match in the beginning but are slightly off at the end.

In Figure 12 are envelope spectrums of faults 4 and 7 from Dataset 2. Both are 0.007-inch



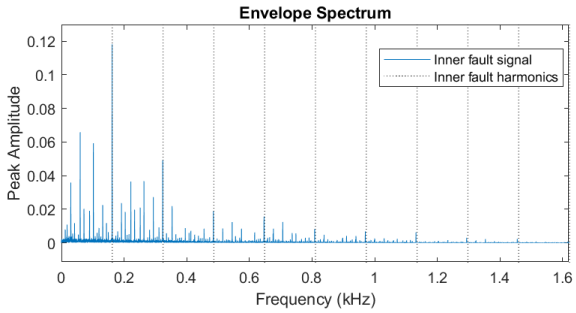
(a) Envelope spectrum of LUT inner fault



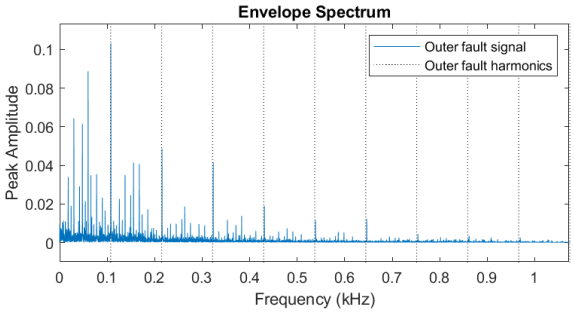
(b) Envelope spectrum of LUT outer fault

Figure 11. Envelope spectrums of Dataset 1

faults from the inner and outer races. The spectrums have clear spikes in the harmonic frequencies indicating that the fault is present. Spectrums obtained from the Dataset 2 have more variation with all frequencies.



(a) Envelope spectrum of CWRU inner fault



(b) Envelope spectrum of CWRU outer fault

Figure 12. Envelope spectrums of Dataset 2

6.3 Decision tree

The decision tree was trained by using the MATLAB function `fitctree`. Function fits binary classification decision tree to input variables. The maximum number of splits is limited by the number of variables in the input data.

Model 1 for Dataset 1 resulted in 100% accuracy. Because optimal classification accuracy was achieved with fewer features, adding more features cannot improve the accuracy.

With Dataset 2, 94% accuracy was achieved with the first model. For the second model, a classification accuracy of 97.8% was achieved. Classification results are visualized as a

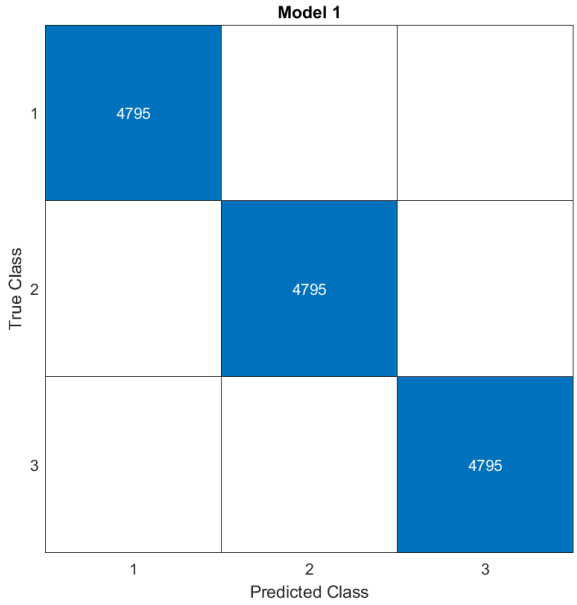
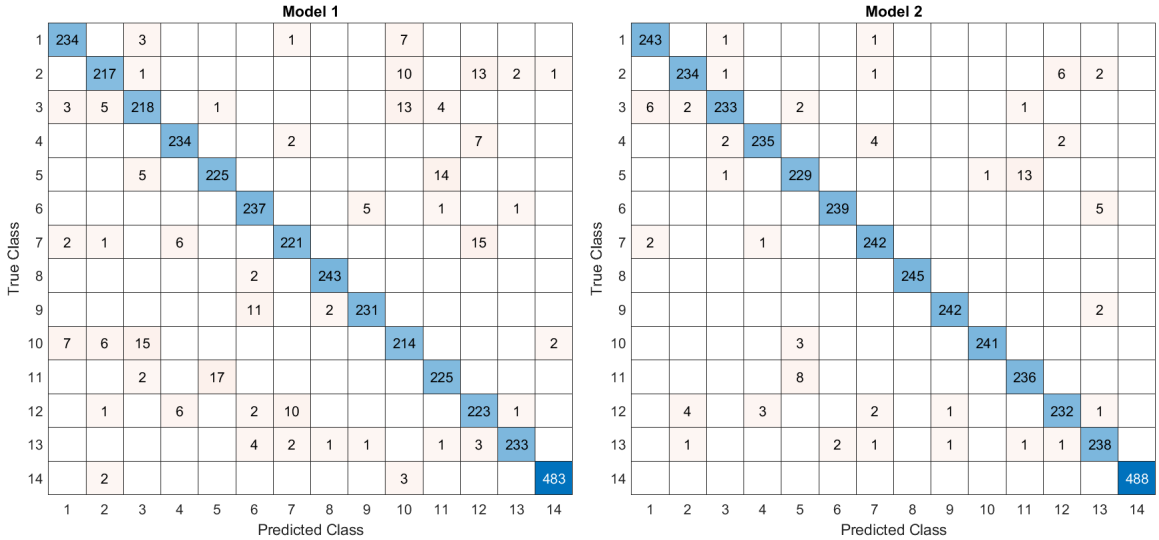


Figure 13. Confusion matrix of classification results of decision tree for Dataset 1

Confusion matrix in Figure 14. The best results in both models are with normal bearing (class 14), where the first model misclassified only a few samples and the second model had no misclassifications.



(a) (b)

Figure 14. Confusion matrix of classification results of decision tree for Dataset 2

Even though decision tree is a simple classification method, it performs well in both cases. Because Dataset 1 is easily separatable, it was expected to have good classification accuracy. Performance with Dataset 2 was originally low but more features brought the

accuracy up significantly. Signals in Dataset 2 are not as easily separated and therefore the number of classes and small training set might be problematic for a decision tree classifier.

6.4 K-nearest neighbor

Parameter k has a large influence on the accuracy of the method. Therefore both datasets are trained with multiple k values to find the optimal value. The algorithm was run 100 times with MATLAB function `fitcknn` for Model 1 with 10 different k values. The means of the accuracies were calculated and listed in Table 7.

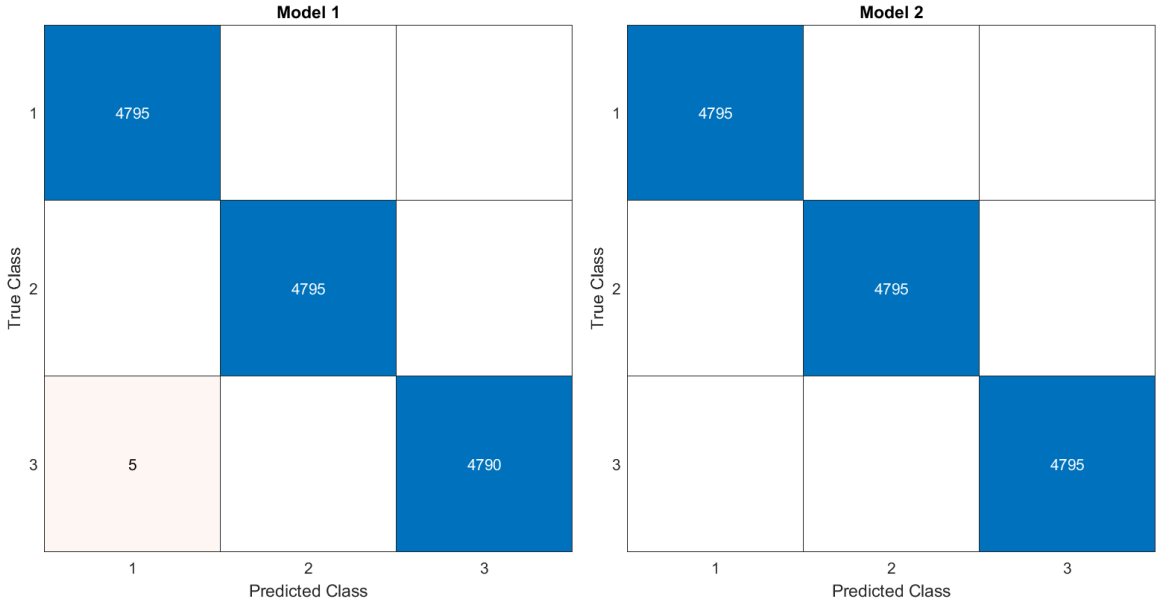
Table 7. Average accuracy (%) of Model 1 KNN with different k values

k	Dataset 1		Dataset 2	
1	99.96	100.00	92.96	97.85
2	99.91	100.00	92.67	97.53
3	99.93	100.00	93.41	97.51
4	99.88	100.00	92.90	97.31
5	99.90	100.00	92.84	97.21
6	99.83	100.00	92.50	97.01
7	99.85	100.00	92.34	96.87
8	99.79	100.00	92.03	96.69
9	99.81	100.00	91.82	96.55
10	99.76	100.00	91.59	96.38

Table 7 shows that for both datasets and models, k value 1 gives the best classification accuracy even though differences in results are slight. In further experiments, k value 1 is used meaning the algorithm will classify the samples based on the closest neighbor.

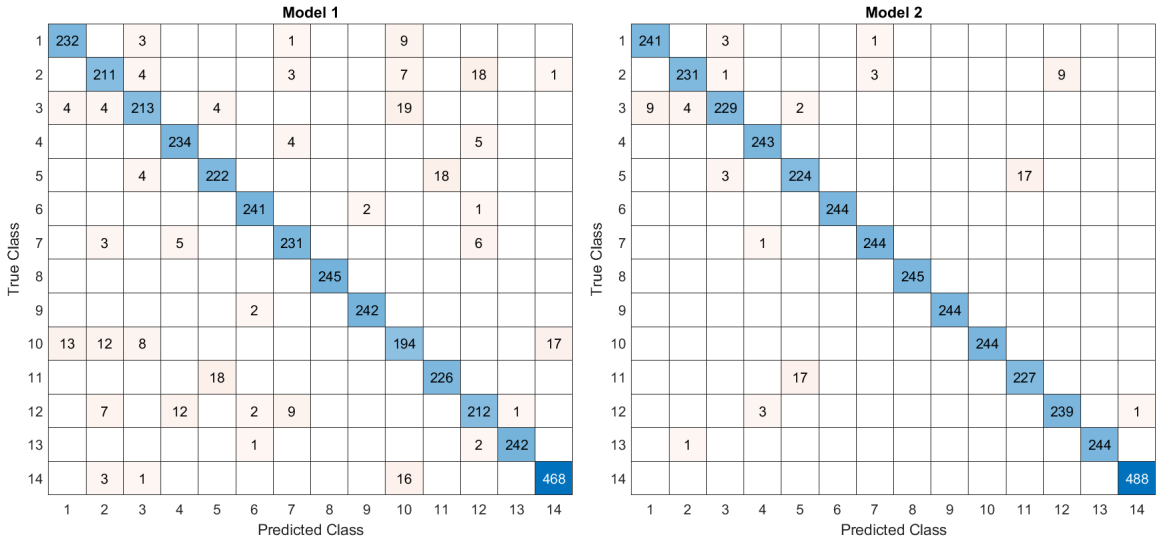
Model 1 gives a classification accuracy of 99.96% with Dataset 1 with only a few misclassifications. Adding more features increases the accuracy to 100%. The confusion matrix of KNN classification results for Dataset 1 is in Figure 15. The small error with Model 1 is due to the misclassification of a few normal bearing signals to the inner race fault. The error was fixed by adding more features.

Dataset 2 has a classification accuracy of 93% with Model 1. Model 2 increases the classification accuracy to 98%. Classification results are visualized in Figure 16. In Model 1, classification errors appear in every class. By adding more features in Model 2, most of these errors disappear.



(a) (b)

Figure 15. Confusion matrix of classification results of KNN for Dataset 1 with $k = 1$



(a) (b)

Figure 16. Confusion matrix of classification results of KNN for Dataset 2 with $k = 1$

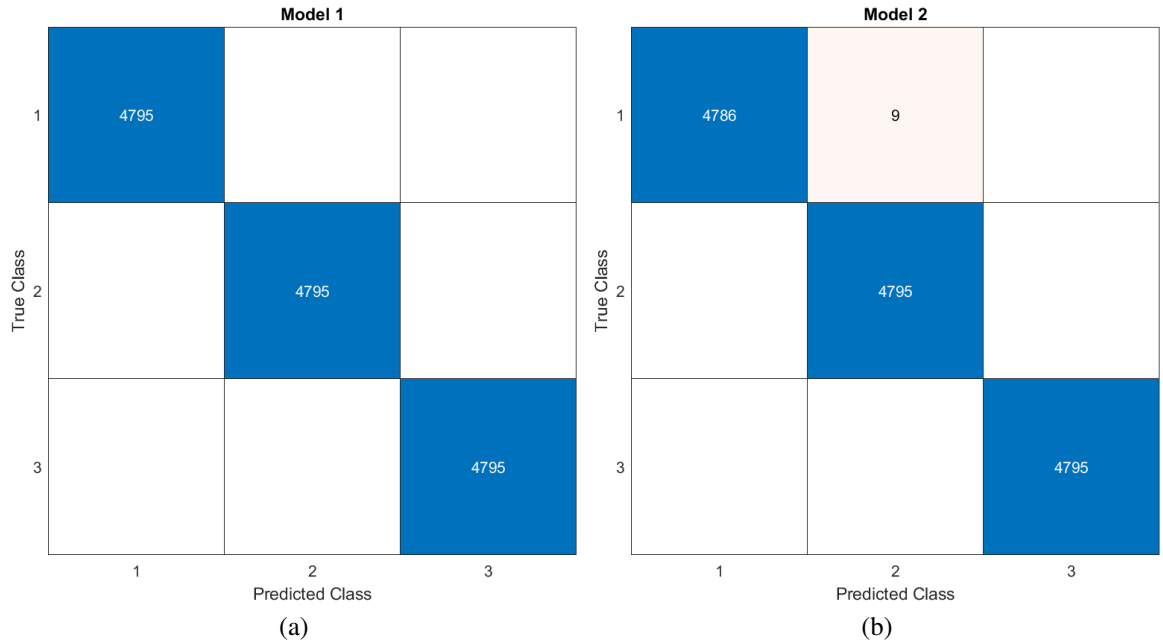
With KNN, adding more features improves the classification accuracy in both cases. Dataset 1 had almost perfect accuracy with Model 1 and Model 2 fixed a few misclassifications. The Model 1 classification with Dataset 2 had a lot of misclassifications. Model 2 improved the classification accuracy significantly. The chosen value for variable k is relatively small, but as tested, a larger value does not improve the accuracy. More features might improve the results further.

6.5 Artificial neural network

The neural network was run with MATLAB function `fitcnet`. The function creates a simple neural network that has 1 fully connected layer of size 10 nodes, a ReLU activation layer, and a softmax layer. There is a possibility to change the layer size and activation layers, but it was not necessary as the network is meant to be kept as simple as possible.

Model 1 had a classification accuracy of 100% with Dataset 1. Adding more features didn't improve the results as Model 2 got an accuracy result of 99.92%. Confusion matrixes for both models are in Figure 17.

Classification accuracy for Dataset 2 with Model 1 was 95.5%. Adding more features improved the accuracy to 98.5%. Confusion matrixes are in Figure 18. With Model 1, the classification errors are distributed between all classes. The larger set of features improves the classification, and some classes are perfectly classified.

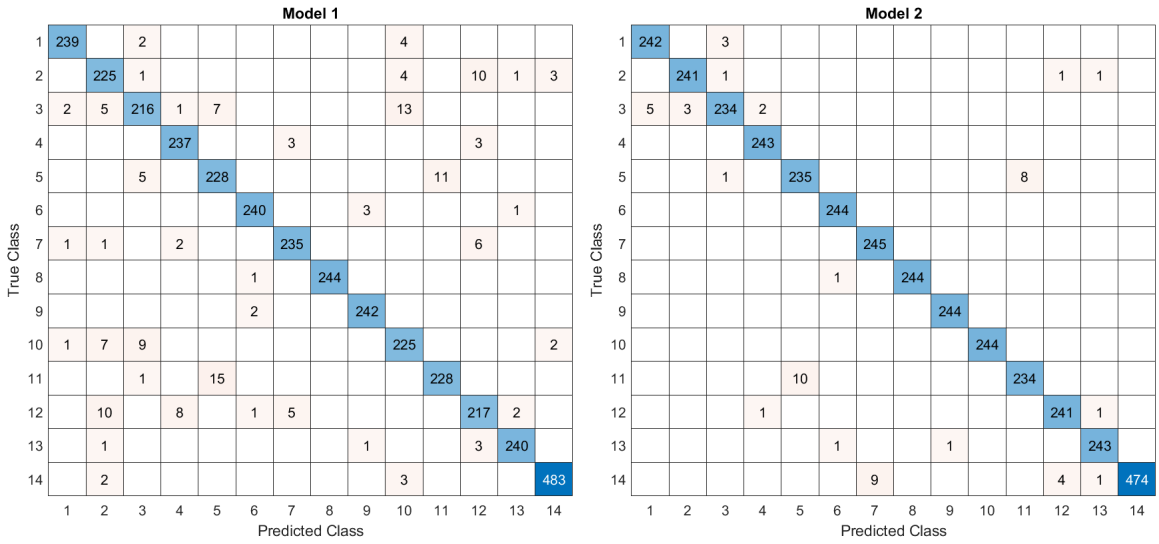


(a) (b)

Figure 17. Confusion matrix of classification results of neural network for Dataset 1

The artificial neural network performs well in both cases. The difference between models in Dataset 1 was not expected as adding more features worsened the accuracy. This might be due to the features not representing the class accurately and being too similar.

Classification accuracy for Dataset 2 improved with Model 2. The remaining errors were



(a) (b)

Figure 18. Confusion matrix of classification results of neural network for Dataset 2

focused on only to few classes. Between classes 5 and 11, there are 17 misclassified samples which could indicate that those samples are very similar among themselves. Also, a few samples were misclassified between classes 1 and 2.

7 Conclusions

The goal of this thesis was to introduce classical and machine learning bearing fault detection methods. Two classical and three machine learning methods were chosen to demonstrate how they perform in fault classification. The chosen machine learning methods were decision tree, K-nearest neighbor, and artificial neural network. With these methods, original signal data can not be used without preprocessing. Long signals were divided into smaller samples and time- and frequency-domain features were extracted from each sample.

Classical methods like envelope analysis and FFT were able to separate inner and outer faults from the signal of normal bearing. The method performed better with the CWRU dataset than with the LUT dataset, which might be due to the nature of the data and the noise produced by the simulation. The SNR can greatly impact the ability to extract the fault frequencies with the envelope spectrum. Therefore higher SNR value can improve the fault detection from the noise.

All three methods produced similar results. Dataset 1 had an average of 100% classification accuracy. The result was expected as the dataset is easy to separate and good features are easy to extract. Dataset 2 is more complex because it has more classes and fewer samples representing each class. The best classification accuracy was achieved by ANN with 98.5% accuracy. However, the difference in the accuracies to other methods is very small. The overall classification accuracy achieved for Dataset 2 was 98%, which is less than more advanced methods are able to achieve but good for simple models.

Each chosen method requires feature extraction which can be done in multiple different ways. The quality of the feature depends on how well it can show the differences between classes. Therefore badly chosen features might not be able to classify the data at all. The peak frequency used as a frequency feature didn't show significant differences between faults which could mean that it is not a good feature. Still, a decision was made to include it in the experiments as it can have a positive effect on the classification.

The best results in previous studies with the CWRU dataset are achieved with deep learning methods up to 99.8% classification accuracy. Convolutional neural networks and auto-encoders are proven to be efficient methods that do not require data preprocessing which eliminates the need for manual feature extraction.

Recent work has been focusing on a wide range of deep-learning methods and they have

produced good results. As presented in this thesis, even simple methods can give good results with less complex calculations. In future work, these methods could be improved by feature extraction and selection. Also, the features and quality of the CWRU dataset have not been explored thoroughly for example with feature ranking. Dividing long time series into a set of samples is the first step in the experiments and therefore the impact of different sampling methods and styles needs to be studied.

REFERENCES

[1] Case Western Reserve University bearing data center. `https://engineering.case.edu/bearingdatacenter`. Accessed: 21.08.2023.

[2] E. Alpaydin. *Introduction to machine learning*. Adaptive computation and machine learning series. 2014.

[3] V. Avina-Corral, J. Rangel-Magdaleno, C. Morales-Perez, and J. Hernandez. Bearing fault detection in adjustable speed drive-powered induction machine by using motor current signature analysis and goodness-of-fit tests. *IEEE transactions on industrial informatics*, 17(12):8265–8274, 2021.

[4] H. Balderston. The detection of incipient failure in bearings. *Material Evaluation*, 27:121–218, 1969.

[5] J. Brownlee. A gentle introduction to the rectified linear unit (ReLU). https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/, 2019. Accessed: 25.09.2023.

[6] W. Caesarendra and T. Tjahjowidodo. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines*, 5:21, 2017.

[7] J. Chen, C. Lin, B. Yao, L. Yang, and H. Ge. Intelligent fault diagnosis of rolling bearings with low-quality data: A feature significance and diversity learning method. *Reliability engineering system safety*, 237:109343, 2023.

[8] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, and H. Qin. Deep neural networks-based rolling bearing fault diagnosis. *Microelectronics and reliability*, 75:327–333, 2017.

[9] R. Euldji, M. Boumahdi, and M. Bachene. Decision-making based on decision tree for ball bearing monitoring. In *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, pages 171–175, 2021.

[10] B. Geropp. Envelope analysis - a signal analysis technique for early detection and isolation of machine faults. *IFAC Proceedings Volumes*, 30(18):977–981, 1997.

[11] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily, and A. Abdellatif. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Engineering Journal*, 14(4):101945, 2023.

[12] M. Heideman, D. Johnson, and C. Burrus. Gauss and the history of the fast Fourier transform. *Archive for History of Exact Sciences*, 34:265–277, 1985.

[13] A. K. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

[14] S. Jawad and A. Jaber. Rolling bearing fault detection based on vibration signal analysis and cumulative sum control chart. *FME Transactions*, 49, 2021.

[15] G. Jia, S. Yuan, and C. Tang. Fault diagnosis of roller bearing based on PCA and multi-class support vector machine. In *Computer and Computing Technologies in Agriculture IV*, pages 198–205, 2011.

[16] H. Kumar and G. Upadhyaya. Fault diagnosis of rolling element bearing using continuous wavelet transform and K-nearest neighbour. *Materials today : proceedings*, 2023.

[17] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138:106587, 2020.

[18] W. Li, M. Qiu, Z. Zhu, B. Wu, and G. Zhou. Bearing fault diagnosis based on spectrum images of vibration signals. *Measurement Science and Technology*, 27:035005, 2016.

[19] H.-C. Lin, Y.-C. Ye, B.-J. Huang, and J.-L. Su. Bearing vibration detection and analysis using enhanced fast Fourier transform algorithm. *Advances in mechanical engineering*, 8(10), 2016.

[20] R. Lior and M. Oded Z. *Data Mining With Decision Trees: Theory And Applications.* Number v. 69 in Series in Machine Perception and Artificial Intelligence. 2008.

[21] R. Lopez-Gutierrez, J. d. J. Rangel-Magdaleno, C. J. Morales-Perez, and A. García-Perez. Induction machine bearing fault detection using empirical wavelet transform. *Shock and vibration*, pages 1–12, 2022.

[22] D. Neupane and J. Seok. Bearing fault detection and diagnosis using Case Western Reserve University dataset with deep learning approaches: A review. *IEEE Access*, 8:93155 – 93178, 2020.

[23] M. S. Nixon, A. S. Aguado, and M. S. Nixon. *Feature extraction image processing for computer vision*. 3rd ed. edition, 2012.

[24] C. A. of Engineering. Deep learning: History and state-of-the-arts. In *The Development of Deep Learning Technologies*, pages 1–11. 2020.

[25] R. B. Randall and J. Antoni. Rolling element bearing diagnostics—a tutorial. *Mechanical systems and signal processing*, 25(2):485–520, 2011.

[26] H. Shao, H. Jiang, Y. Lin, and X. Li. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mechanical Systems and Signal Processing*, 102:278–297, 2018.

[27] A. Sharma, R. Jigyasu, L. Mathew, and S. Chatterji. Bearing fault diagnosis using weighted k-nearest neighbor. pages 1132–1137, 2018.

[28] L. Xiao, X. Yang, and X. Yang. A graph neural network-based bearing fault detection method. *Scientific Reports*, 13, 2023.

[29] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler. Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8:29857–29881, 2020.

[30] Z. Zhang, Y. Wang, and K. Wang. Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *Journal of Intelligent Manufacturing*, 24, 2013.

[31] J. Zhao, X. Sun, J. Wang, Z. Zhou, Y. Muhamedsalih, F. Gu, and A. D. Ball. Fault diagnostics of ac motor bearings based on envelope analysis of vibration residual signal. In *Proceedings of TEPEN 2022*, pages 889–899, 2023.