University of South-Eastern Norway

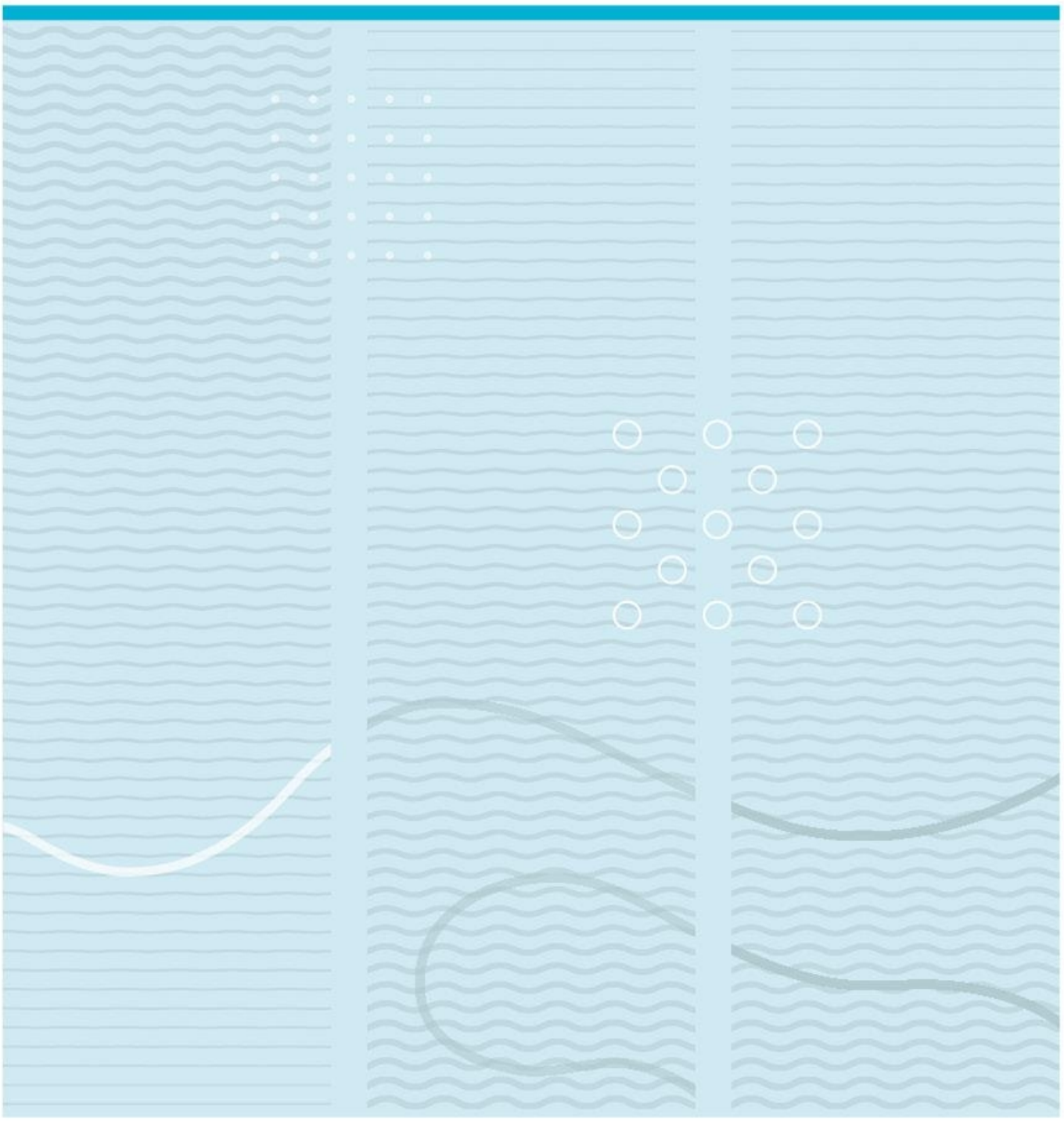
Faculty of Technology, Natural Sciences and Maritime Sciences

Study programme: CS5000

Autumn 2023

Zahra Dirin Samimifard

Using 2D CNN models with mid-level fusion-based approach for multi-classification of bearing faults



University of South-Eastern Norway

Faculty of Technology, Natural Sciences and Maritime Sciences
Institute of
PO Box 235
NO-3603 Kongsberg, Norway

http://www.usn.no

This thesis is worth 60 study points

Abstract

In order to increase the efficiency within the life cycle of the equipment, predictive maintenance has evolved as a popular research area. Bearings are one of the key components in rotating machinery which facilitate rotational movements. These elements are prone to defects and these defects can be traced using the signals that are generated by the bearings. These signals can be analyzed and provide varied characteristics that can be used as the key elements for predictive maintenance. In this thesis, a more efficient method has been proposed to classify various bearing faults using different deep learning models such as AlexNet, LeNet, NASNet, ResNet, EfficientNet, and VGG16. The classification has been done by using two different datasets called NASA and CWRU which are publicly available. Both datasets contain raw bearing vibration signals and we have generated varied 2D images from the raw vibrational signals, including frequency spectrums, spectrograms, and polar spectrums. Then, we developed and evaluated the performance of AlexNet, LeNet, NASNet, ResNet, EfficientNet, and VGG16 models based on 2D features. We have used different performance metrics such as F1-score, accuracy ROC curve, etc. In the NASA dataset, we observed the highest classification performance in the VGG16 model that was trained using polar spectrums with an accuracy of 97.12% and an F1-score of 0.965. In the CWRU dataset, AlexNet using spectrogram demonstrated the highest accuracy of 96.8% and F1-score of 0.968. In the next step, we provided a dual-channel mechanism by incorporating statistical features into our model using mid-level fusion. We have proved the efficiency of the dual-channel deep learning models and demonstrated the significance of the dual-channel EfficientNet to provide 100% accuracy using spectrogram and frequency spectrum in bearing anomaly classification. We have tried this method on both CWRU and NASA datasets, and significantly we could observe that dual channel methods demonstrated remarkable classification performance in comparison with the single channel models. Based on our results, we believe that different sectors that provide predictive maintenance can benefit from our results. By incorporating our proposed solution, they are able to enhance their fault classification performance. Add to the point, this thesis provides a comprehensive comparative study on bearing fault classification using different deep learning models, different channels, and different features which can serve as a foundation for other scholars to modify and evaluate their ideas.

Acknowledgements

I would like to express my deepest gratitude towards my supervisor Vimala Nunavath for her invaluable guidance and support. Her experience shaped the foundation of my research and helped me overcome the challenges I faced. I am also thankful to my co-supervisor Yapi Donatien Achou for helping me with the necessary domain knowledge in the scope of this thesis. Finally, I send my sincere gratitude to all my professors and faculty members at USN.

Acronyms

PM Predictive maintenance

CM Condition monitoring

AI Artificial Intelligence

ML Machine learning

DL Deep learning

CNN Convolutional Neural Networks

ANN Artificial Neural Networks

STFT Short-Time Fourier Transformation

FT Fourier Transformation

BPFI Ball Pass Frequency Inner race

BPFO Ball Pass Frequency Inner race

BSF Ball Spin Frequency

WT Wavelet Transformation

DCNN Deep convolution neural network

GAN Generative Adversarial Network

RNN Recurrent Neural Network

LSTM Long Short Term Memory

Contents

List of Figures

List of Tables

1 Introduction

Rolling element bearings are widely used in rotating machinery like helicopters, locomotives, airplanes, pumps, etc. They facilitate rotatory and horizontal movements among the equipment, thereby, they are an integral element to establish optimal performance. Motors are the supplier of almost 50% of energy for industrial applications and bearings are responsible for approximately 40% of their corresponding failure (Zeidan, Paquette, et al., 1994). A bearing consists of the inner and outer ring, a roller element, and a cage. Inner and outer rings provide a pathway for the rolling elements and the cage holds an even space between the bearing to guide them within the raceways(B. Zhang et al., 2008). Factors like excessive loads, peeling, sub case fatigue, misalignment, inclusion, and humidity can cause anomalies in the inner race, outer race, and rolling element of the bearing (Jin, Sun, Que, Wang, & Chow, 2016).

These anomalies interfere with the health of the equipment and cause several negative consequences including increased downtime, a plummet in efficiency and revenue, and in severe conditions catastrophic incidents, and human injury. In order to prevent these incidents, condition-monitoring techniques like vibration signals, acoustic measurements, and temperature measurements can be used to monitor the health of the equipment (Shah & Patel, 2014). The failure or malfunction in the bearing can affect the current vibration and produce abnormal signals, which can be used as an indicator to distinguish the defects. Thereby, through monitoring and analyzing these signals, we can predict and prevent failures by applying preventive maintenance.

Traditionally, equipment maintenance was performed after the occurrence of incidents, or equipment breakdown. The vast application domain of the bearings and the importance of their health condition gears scientists toward innovative solutions to detect and classify these anomalies or faults by using signal processing techniques (Neupane & Seok, 2020). With technological advancements and a massive amount of research being done using Artificial Intelligence (AI) in several domains such as healthcare (Amann et al., 2020), wood industry (Baesler, Moraga, & Ramis, 2002), oil and gas (Aung, Mikhaylov, & Aung, 2020), predictive maintenance using AI has also been evolved to provide operational efficiency (Matzka, 2020). Deep Learning (DL) methods, by providing automatic feature extraction from the raw data, modeling complex relations within the data, and their tolerance towards redundant features like noise, provide higher classification accuracy, enhance the efficiency of the equipment's life cycle, and mitigate the chance of failure (Neupane & Seok, 2020; Saufi, Ahmad,

Leong, & Lim, 2019a).

Previous solutions for bearing fault classification have been leveraging the hidden features among certain single types of data, like time domain features (Nayana & Geethanjali, 2017), frequency domain features (Hakim et al., 2022), or 2D representations of the signal (T. Lu, Yu, Han, & Wang, 2020), and have not been evaluating the full potential characteristics hidden in other forms of data. For example, by using time domain features, we may overlook frequency-based characteristics, or frequency domain features are not able to evaluate the time-based components. The unique architecture of CNN eliminates the application of feature-extraction methods which demonstrates its superiority over other traditional machine-learning techniques. Signals can be turned into 2D illustrations using various transformation techniques including Fourier Transformation (Bagchi & Mitra, 2012), Short-Time Fourier Transformation (Griffin & Lim, 1984), Principal Component Analysis (Castells, Laguna, Sörnmo, Bollmann, & Roig, 2007), Histogram of Oriented Gradients (Muthamizh Selvan & Rajesh, 2012), Wavelet transformation (Osadchiy, Kamenev, Saharov, & Chernyi, 2021), polar spectra (Lyzenga, 2017), etc.

Thereby, these demonstrations have the potential to confer advantages to the classification performance using different variations of 2D CNN. Hence, the objective of this thesis is to develop an innovative solution using deep learning models to classify bearing faults. The innovative solution includes a 2 - channel fusion model with regard to 2D visualization like polar spectrum, frequency spectrum, and the spectrogram for one channel, and statistical features for the second channel.

1.1 Motivation and Problem Statement

Anomaly detection and classification of the bearings is a novel research area as it provides operational efficiency, asset availability, safety, and cost-effectiveness of the equipment. Bearings are vital parts and the main source of failure in rotary machines like airplanes, automobiles, turbines, and generators (Jin et al., 2016). Material properties, lubricant properties, bearing size, the number of rolling elements, speed of bearing, and installation are among the factors that can impact the useful life cycle within this component. The faults like material fatigue, plastic deformation, and uneven rotation can stem from excessive load, wear, and lubrication (B. Zhang et al., 2008). Reaching a high accuracy in anomaly detection and classification of the equipment reduces downtime, increases efficiency, and prevents catastrophic failures.

Rolling elements like bearings are one of the key components for providing rotation between two parts within the equipment, and their failure accounts for one-third of the failures in rotating machinery (Neupane & Seok, 2020). The failure of an intact functional routine in this element results in the full breakdown of the rotating machinery. It is vitally important to uphold the integrity of the equipment to ensure operational efficiency especially when it deals with human safety or has the potential of significant financial loss. The defects within this element can arise in the inner race, outer race, and rolling element parts. Detecting these defects in an early stage can eliminate downtime and prevent incidents.

DL models have been widely used in anomaly classification of the bearings. Using the available data, these algorithms try to find current patterns and make corresponding predictions (Neupane & Seok, 2020). DL algorithms are able to define higher-level and lower-level categories with higher efficiency and accuracy (Neupane & Seok, 2020). They are comprised of multiple processing layers and are able to get insights into multiple levels of abstraction. The initial insights from the very first layers will be sent to the next ones obtaining a higher complexity until the final layer, in which the classification is generated (Neupane & Seok, 2020). The quality of the input data affects the classification results, thereby ensuring that the input data is capable of providing necessary information is challenging and vital for a robust classification. Therefore, in this thesis, dual-channel DL models are built for the classification of various bearing faults.

1.2 Research Questions

1. Are dual-channel CNNs suitable for multi-classification of bearing faults?

   To answer this research question, in this thesis, dual-channel CNN models will be developed using spatial and statistical features for the classification of bearing faults. To do so, we use separate data pipelines for each type of channel. In the 1D channel, statistical features will be fed. Whereas, 2D representations will be fed to the 2D CNN channel. The developed dual-channel CNN models use two datasets such as NASA and CWRU.

2. Does the dual-channel CNN model perform better than single-channel CNN models?

   In order to compare the performance of dual-channel models with single-channel 2D CNN models for the multi-classification of the bearing faults, we developed both types of models and

later compared their performance.

3. Does the dual-channel CNN model perform better than state-of-the-art algorithms?

   To answer this research question, the results of developed 2D CNN models are compared with the state-of-the-art results.

1.3 Thesis Goals

To achieve the main goal of this thesis, it is further divided into sub-goals as follows:

- Examine the state-of-the-art research within the field of anomaly classification and further improve it by introducing dual-channel 2D CNN models.
- Develop dual-channel CNN models such as EfficientNet, AlexNet, NasNet, ResNet, VGG16, and LeNet for the classification of varied faults within the bearing using spatial and statistical features.
- Validate the robustness of the proposed models by evaluating the performance of the model on two datasets and comparing their performance.

1.4 Research Approach

In this thesis, as illustrated in Figure 1, we have followed an applied research approach to provide applicable solutions for a challenge within predictive maintenance.

As per applied research, we started by defining our research focus which was evaluating the classification performance of the fusion EfficientNet in anomaly classification of the bearings. Then we provided a comprehensive literature review to evaluate how current methodologies and deep learning models have contributed to the performance of anomaly classification of the bearing. It was a critical step to get acquainted with the state-of-the-art methodologies and find the gaps within the scholar's previous research. In the next steps, we decided on the models, features, performance metrics, and datasets that could be used in our research.

We determined, AlexNet, LeNet, NASNet, ResNet, EfficientNet, and VGG16 as our main models, Polar spectrums, Spectograms, frequency spectrum, and statistical features as our evaluation features that we could extract from both NASA and CWRU datasets. These datasets contained the vibration

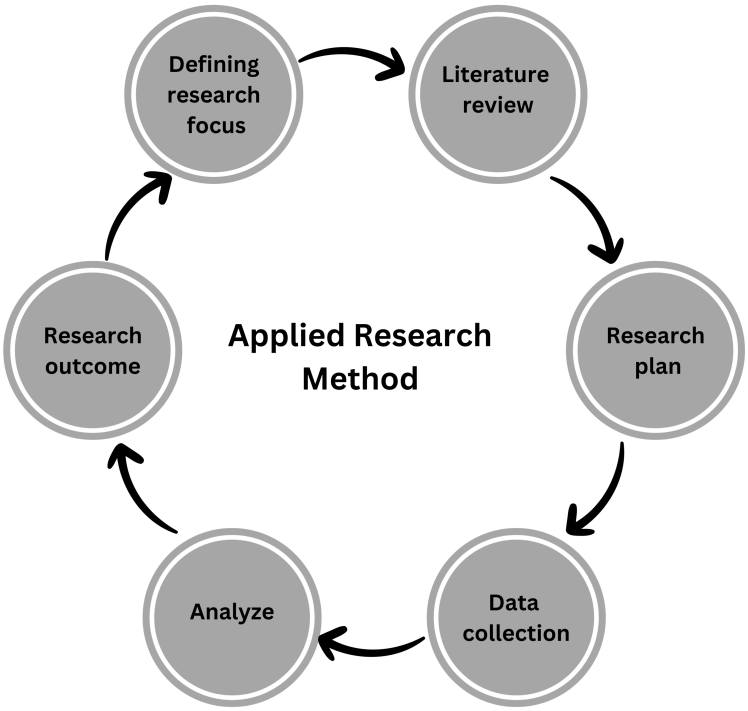


Figure 1: Applied research approach.

signals derived from the bearings which have been prone to different fault conditions including inner race defect, outer race defect, and rolling element defect. Afterwards, we analyzed the transformed vibration signals and evaluated the performance of the target model and architectures based on the determined performance metrics, and provided a comparative analysis to showcase the fusion methods and the significance of the fusion EfficientNet on the bearing anomaly classification

1.5 Assumptions and Limitations

This subsection outlines the assumptions we made and the limitations we encountered in this thesis.

1.5.1 Assumptions

1. We assume that the datasets have been generated through consistent environmental conditions during the capturing period.
2. The CWRU dataset consists of varied severities of the different fault types, and we assume that the severity has remained constant during each experiment.
3. We assume that mid-level fusion can adequately capture the features from both channels.

1.5.2 Limitations

1. The highest challenge that we faced during the experiments was our limited computational resources. Deep learning architectures require a considerable amount of computational resources. Therefore, we were not able to apply more complex hyper parameter tuning or use a higher number of epochs. Add to the point, The NASA dataset approximately consists of 700 million data points. The considerable magnitudes of data within this dataset, significantly diminish the processing speed to generate 2D images.

2. The NASA dataset used in this thesis includes unlabeled data. Therefore, in order to perform classification, we need to turn it into a labeled form. However, due to the presence of the noise within the signals, precisely pinpointing the initial defective data point is challenging.

3. By removing the noises, we might remove useful information or anomalous data points that would benefit the classification performance.

4. In our dual-channel model, we train our fusion model based on different data types of 2D images and statistical features. Therefore, the model complexity increases and results in a higher training time.

1.6 Thesis Contributions

The main contributions of this thesis are:

- Development of different deep 2D CNN models for classification of bearing faults using 2-channel fusion model.
- Evaluation of different deep 2D CNN models' performance on NASA and CWRU datasets.
- Comparing the performance of dual channel models with their peer single channel models.
- Comparing the efficiency of the dual channel EfficientNet model with the state-of-the-art result.

1.7 Thesis Outline

The rest of the thesis is organized as follows:

Chapter 2: This section provides the background information needed in order to understand the theories, technologies, and domain.

Chapter 3: This section presents a comprehensive literature review of prior research done on using DL for multi-classification of bearing faults.

Chapter 4: This section explains the pre-processing techniques that are used before feeding the data into 2D CNN models, the details of both NASA and CWRU datasets, and the architecture that we used to conduct several experiments for the multi-classification of bearing faults.

Chapter 5: This section presents the experimental results we have achieved and the discussion regarding the results.

Chapter 6: This chapter concludes the thesis by providing a summary of the work done in the thesis. Also, outline the potential work that could be done to our research in the future to achieve better desirable results.

2 Background

In this section, we provide an overview of foundational theory on the chosen domain and the Deep Learning models that have been used to answer the research question of this thesis.

2.1 Condition monitoring and predictive maintenance

Sensors play an important role in Industry 4.0, and by using them, the system can be more digitalized, and automation will become possible. The main approach in Industry 4.0 is to increase efficiency, and sensors have made it possible. They link multiple devices and enable machine-to-machine and machine-to-human interactions. Using computational power in their systems, these sensors have become more intelligent and are capable of data calculation. Industry 4.0 evolved with the intelligent process, by incorporating sensor technology, big data, cloud computing, AI, intelligent robotics, smart valves, and automated smart control applications (Javaid, Haleem, Singh, Rab, & Suman, 2021).

Sensors continuously assess conditions and take necessary actions about them. These counteractions efficiently provide analytical concepts that eliminate human error and increase efficiency within the system. They provide valuable knowledge, boost production, advertise, sell, compete, reduce costs, and enhance business performance and overall productivity. In the industry sections, sensors have been incorporated to solve various business challenges including predictive maintenance, building automation, asset monitoring, and process automation. They are a vital part of software intelligence and have changed the manufacturing process to ensure high productivity. Sensors collect data using a transducer and use the computational power to produce insight and transfer them to be processed in information systems and workflows. They have evolved in a way that they are capable of self-assays and auto-calibration nowadays (Javaid et al., 2021).

Maintenance is divided into three main categories:

1. Corrective maintenance: In this maintenance strategy, an action will be taken after an error or system failure has occurred. (Davari et al., 2021)

2. Preventive maintenance: This category encompasses the efforts to retain the integrity of certain conditions by providing systematic inspection, detection, and prevention of failures. It is divided into three groups: scheduled maintenance, condition-based maintenance, and Predic-

tive maintenance(Compare, Baraldi, & Zio, 2019)(Davari et al., 2021)

3. Predictive maintenance: For PM, there exist three approaches: model-based prognosis, knowledge-based prognosis, and data-driven prognosis. Data-driven PM method, which is the backbone of our research, incorporates the industrial data with computational methods to produce analytical information about the equipment. In this approach a training model will be deployed based on the sensor data, then it performs classification. The bearing's vibration shall provide us with helpful information about the health of the equipment and can be analyzed using signal processing methods (Davari et al., 2021).

2.1.1 Bearing Fault Structure

A bearing consists of the inner race, outer race, rolling element, and the cage which can be seen in Figure 2. Any damage within these parts can cause the corresponding failure which we state as an inner race defect, outer race defect, and rolling element defect. These varied defective components generate a diverse range of frequencies. A healthy bearing generates regular vibration signals which are generated from forces between different components and the rotational movement of the bearing. The continuous vibration structure in a healthy bearing profoundly differs from the nonuniform, discontinuous structure of a defective bearing. Throughout every rotation cycle, the vibration signal is influenced by the defect each time the motion encounters the defective zone (B. Zhang et al., 2008).

During the life cycle of a bearing, different frequency components are generated and can be used to distinguish different defective parts within a bearing. Each defective part produces frequencies based on the operational characteristics of the bearing including the number of rolling elements, number of balls or rollers, diameter of a rolling element, pitch diameter of the bearing, and contact angle of the bearing (Nivesrangsan & Jantarajirojkul, 2018). These frequencies can be described as follows:

- **Ball Pass Frequency Inner race:** BPFI corresponds to the frequency that the bearing generates when there is an error in the inner section of the bearing. It shall be calculated using the following formula (Nivesrangsan & Jantarajirojkul, 2018)(Samimifard, 2023):

$$\mathrm{BPFI} = \frac{N \cdot Z \cdot \mathrm{RPM}}{60} \cdot \left(1 + \frac{d}{D} \cdot \cos\alpha\right) \tag{1}$$

- **Ball Pass Frequency Outer race:** BPFO race corresponds to the frequency that the bearing gen-

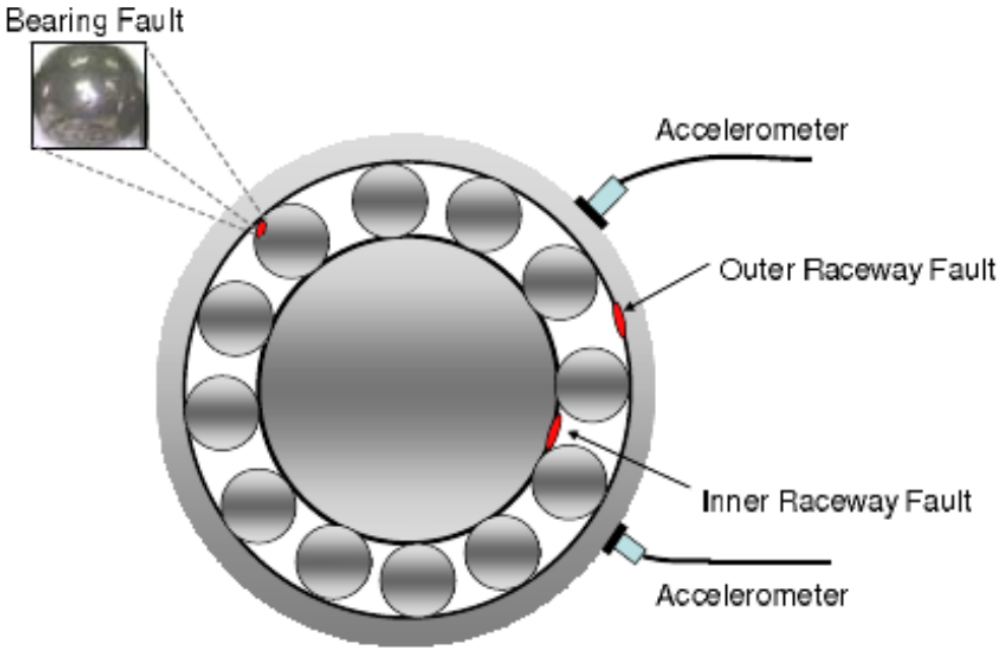


Figure 2: Bearing faults.(B. Zhang et al., 2008)

erates when there is an error in the outer section of the bearing. It shall be calculated using the following formula (Nivesrangsan & Jantarajirojkul, 2018)(Samimifard, 2023):

$$\mathrm{BPFO} = \frac{N \cdot Z \cdot \mathrm{RPM}}{60} \cdot \left(1 - \frac{d}{D} \cdot \cos\alpha\right) \tag{2}$$

- **Ball Spin Frequency:** BSF corresponds to the frequency that the bearing generates when there is an error in the rolling elements. It shall be calculated using the following formula (Nivesrangsan & Jantarajirojkul, 2018)(Samimifard, 2023):

$$\mathrm{BSF} = \frac{N \cdot Z \cdot \mathrm{RPM}}{120} \cdot \left(\frac{D^2 - d^2}{D^2}\right) \tag{3}$$

The variables for the three above formulas are explained as follows:

- N: Number of rolling elements
- Z: Number of balls or rollers
- RPM: Revolutions per minute of the bearing
- d: Diameter of a rolling element

- D: Pitch diameter of the bearing
- α: Contact angle of the bearing

2.1.2 Envelope analysis

Envelope analysis is a signal processing technique that is used to analyze the vibration signal generated from the bearings (B. Zhang et al., 2010). We can relate the frequency components derived from the envelope analysis to the frequency components of the bearings and determine the presence of specific defects. The extracted signal is a combination structure of the actual vibration signal and the broadband noise. The defects within the inner race, outer race, or rolling element can cause energy distribution and result in resonances in the bearing and related components when the rolling element passes through the spall. The bearing periodically generates these impulses during the continuous rotational movements and impulses are much shorter in comparison with the bearing's rotational cycle. These impulses affect a broad range of frequencies and cause varied resonances within the bearing. Envelope analysis is capable of revealing the faults within the bearing life cycle using the demodulating of the vibration signal (B. Zhang et al., 2008).

As illustrated in Figure 3, Envelope analysis as a multi-step procedure performs band pass filtering, followed by Hilbert transformation to generate the analytic signal. The envelope signal is then generated by taking the absolute value of the analytic signal. The envelope is then followed by the FT to extract the features as components for the experiments (B. Zhang et al., 2008, 2010)

Vibration signal: As the rolling element bearing rotates, the forces between components and external forces generate vibration signals. These vibration signals exist in any health condition of the bearing and demonstrate a signature based on the mode that the signals were collected (B. Zhang et al., 2008).

Band pass filtering: Vibration signals accompany background noise from the shaft and other components. Varied frequency bands including Low, middle, and high-frequency, reveal hidden characteristics of the defective component. Low-frequency bands that are below 1KHz, encompass fault frequencies related to processing equipment. Middle-frequency bands which are considered between 1KHz and 20KHz demonstrate the faults within the surface. However, the majority of the generated energy from a defective bearing is located within the high-frequency range which is above 20KHz.

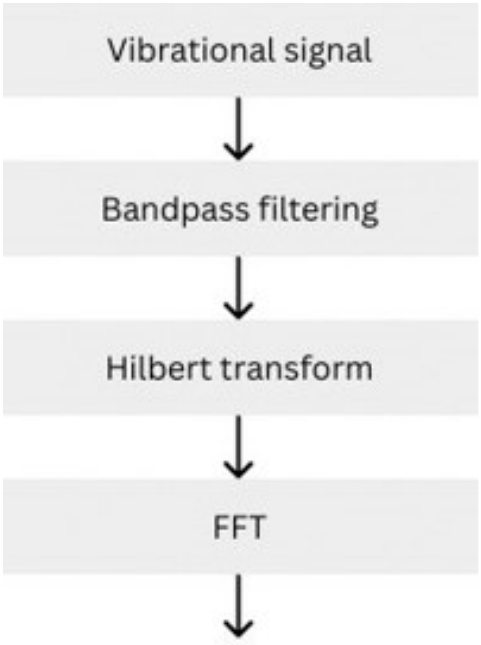


Figure 3: Demodulation flowchart.

Bandpass filtering filters the signal based on specific frequencies. In the bearing anomaly classification, we need the components around the resonant frequency that were triggered by the defective element (B. Zhang et al., 2008; Q. Lu, 2012).

Hilbert transform: A fault source modulates the vibration signal. Therefore, envelope demodulation can be used to extract amplitudes related to the fault frequency (Q. Lu, 2012). Hilbert transformation is used for developing analytic signals and the envelope signal is its absolute value (B. Zhang et al., 2008). An analytic signal is a complex signal whose imaginary part is the Hilbert transform (Feldman, 2011). In the following equations x(t) is the time domain signal and h(t) corresponds to the the Hilbert transformation formula(Q. Lu, 2012).

$$h(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(t)}{\lambda - t} dt = x(t) * g(t) \tag{4}$$

$$x(t) = -\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{h(\lambda)}{t - \lambda} d\lambda \tag{5}$$

The second equation is the inverse of the Hilbert transformation where(Q. Lu, 2012):

$$g(t) = \frac{1}{\pi t} \text{if } t \neq 0 0 \text{if } t = 0 \tag{3}$$

$h(\lambda)$is the result of applying the Hilbert transform to the x(t) and passing it through the filter as g(t). The magnitude of the analytic and the phase of the analytic are represented as the following equations respectively (Q. Lu, 2012):

$$\left|a(t)\right| = \sqrt{x^2(t) + (x'(t))^2} \tag{6}$$

$$\theta(t) = \arctan\left(\frac{x'(t)}{x(t)}\right) = 2\pi f_0 t + \varphi(t) \tag{7}$$

FFT: In the last step of the demodulation, to visualize the faulty component FFT is applied to the envelope signal to calculate the frequency spectrum (B. Zhang et al., 2008). This transformation has been widely used in frequency amplitude computation, and filtering operation. Applying FFT on real-valued information generates results with redundant imaginary parts. Fast Fourier Transformation optimizes trigonometric function calculation and uses a limited variety of discrete angles which as a result increases efficiency. FT allows the use of only half of the data. As demonstrated in equation 8, real parts represent even data values and imaginary parts represent odd data values (Bechhoefer & Fang, 2012).

$$F_n = F_n^e + e^{2\pi i n/N} F_n^o \tag{8}$$

where

$$F_n^e = \sum_{k=0}^{N/2-1} f_{2k} e^{2\pi i k/(N/2)} \tag{9}$$

$$F_n^o = \sum_{k=0}^{N/2-1} f_{2k+1} e^{2\pi i k n/(N/2)} \tag{10}$$

A vector data structure containing two arrays for even and odd data is then generated. A complex vector is also defined for real and imaginary data. FFT copies the real vector into the complex vector. In the next step, a decimation in frequency FFT is performed which applies the FFT to half of the complex data. Then A decimation in time FFT is applied to it. The transformations within Equation 9 and

Equation 10 are then recombined using Equation 8 to ensure a correct transformation (Bechhoefer & Fang, 2012).

2.2 Artificial Neural Network

ANNs mimic the human brain as an interconnection structure of neurons. As illustrated in Fig 4, in the first step, nodes or perceptrons process the received data using mathematical functions. Then, the output of the processed data will be sent to the next layer of perceptrons, and ultimately to the output layer. ANNs include a wide range of networks with unique architectures like perceptron networks, feed-forward layered networks, and recurrent neural networks. The efficiency of each network depends on its training. During the training process, the network learns to adjust its weights to minimize the differences between the predicted and the actual values to increase the accuracy (Kufel et al., 2023).

2.2.1 Deep Neural Networks

DNNs are a type of ANN with multiple hidden layers between the input and output layers. These hidden layers are the key components in the DNNs by extracting more features within the data. Each layer improves the refinement of the extracted feature and more layers empower the network's training capability. As demonstrated in Figure 4 when the data passes through the network containing multiple hidden layers, in each layer weight, biases, and activation functions are applied to refine the data. This network is capable of extracting complex features and patterns, which makes it a great tool for tasks like image processing, and natural language processing (Kufel et al., 2023).

2.2.2 Convolutional Neural Networks

DL models like CNN have been developed to process multi-dimensional data input and extract their features automatically using convolution structures. As illustrated in Figure 5, convolutional layers generate feature maps containing the pattern within the data and optimize the features during the training phase. To avoid data loss, padding is applied at the borders by increasing the input size with zero value. As the generated feature maps might encounter overfitting, pooling can be applied to reduce the spatial dimension of the feature maps. Finally, the data is passed through the fully connected layers for further classification (Z. Li, Liu, Yang, Peng, & Zhou, 2021).

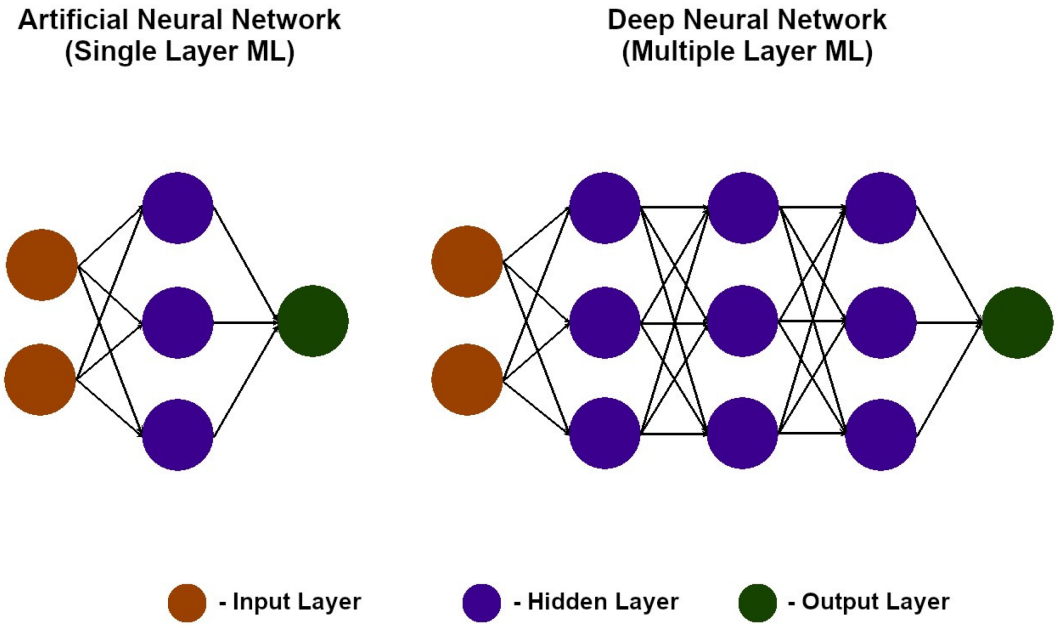


Figure 4: Representation of DNN and ANN. (Kufel et al., 2023)

Convolution layers: As illustrated in Equation 11, this layer is a combination of two function structures and demonstrates how one alters the shape of the other. In the image classification tasks, this procedure involves features and the feature detector. Figure 6 provides a visual representation of the equation, it demonstrates how the kernel moves through the input image and provides a feature map. The feature map encapsulates the vital information within the image and reduces the size of the image's dimensions (Haque, Verma, Alex, & Venkatesan, 2020).

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau)\, d\tau \tag{11}$$

Pooling layers: In the next layer to address the variability of the shapes, textures, and orientations of the 2D input, pooling is applied to the feature maps which were derived from the convolution layer. For example, different breeds of dogs in various orientations and positions fall into the dog category. To ensure consistency, pooling targets these variations by extracting dominant features from the feature map. Furthermore, this layer also reduces the dimension of the feature map by 75%, which downsizes the number of parameters before feeding them to the neural networks, and prevents overfitting (Haque et al., 2020).

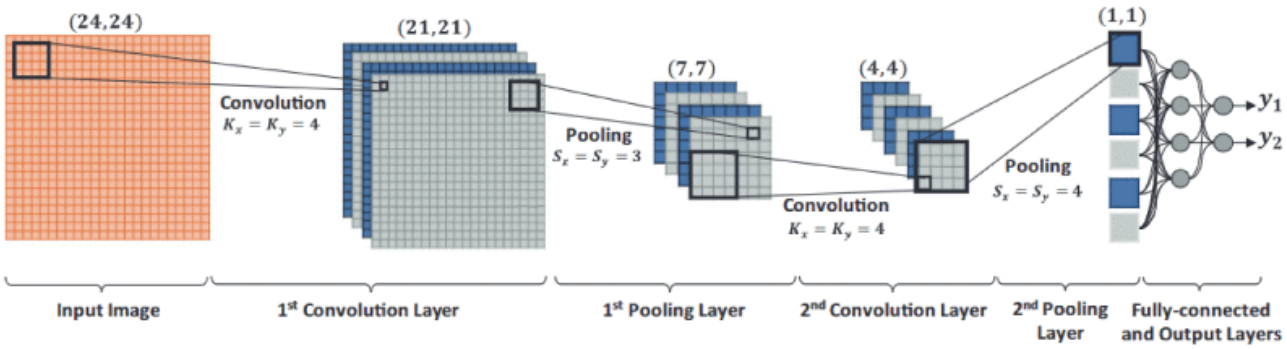


Figure 5: CNN Architecture. (Shahid et al., 2022)

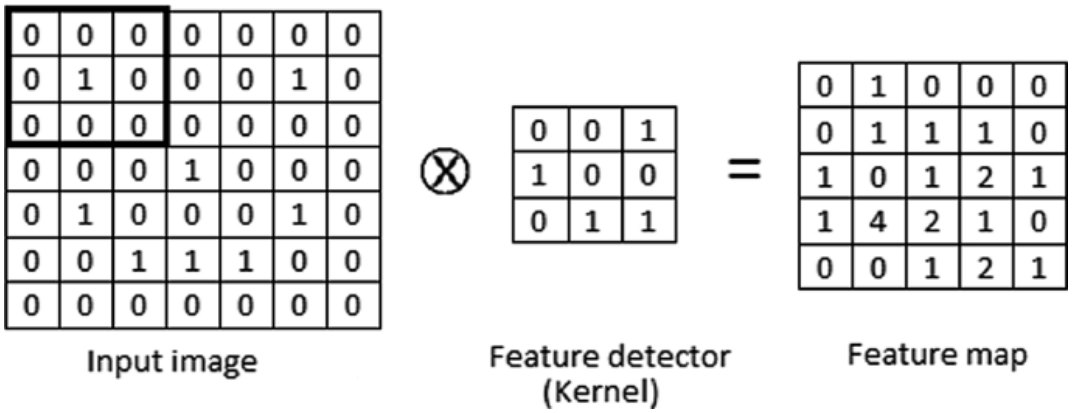


Figure 6: CNN Architecture. (Haque et al., 2020)

Flattening layers: Before feeding the feature maps into the neural networks, the 2D pooled features need to map into a 1D column. This can be done by taking the values from each map and then lining them up into a one-dimensional column(Haque et al., 2020).

Full Connection layers: The flattened layer is the input to the ANN and is fully connected to the next layer in the network. This link between the input features and their attributes in the corresponding layer increases the accuracy of the classification (Haque et al., 2020).

3 State of the Art

There is significant research that is going on in the areas of developing deep learning (DL) algorithms for bearing faults classification. In this section, we present some of the algorithms that have been developed by researchers in this field. The literature review is summarized and presented in Table 1.

Wang *et al* (H. Wang et al., 2022) studied the diagnosis of the bearing anomalies using BAS-SWD, and a multi-sensor data fused 1D-CNN which was originally based on VGG16. This method was conducted on non-stationary signals within the CWRU dataset. Initially, these signals were divided and normalized to ensure a standard input. A novel BAS-SWD algorithm was used for the feature extraction to decompose the pre-processed data into OCs, which are distinguishable data points with anomaly characteristics. The top 3 OCs with the highest envelope spectral kurtosis were used as the input of the model. The authors utilized the 1D CNN model and added a fused layer to combine data from the different sources. This multi-sensor fused model demonstrated a remarkable classification result with an average classification accuracy of 100%. The authors would like to challenge mixed fault diagnosis in gear, and bearing using this method in their future work.

Shao*et al* (Shao, Jiang, Zhang, & Liang, 2017) proposed a diagnosis method based on a convolutional deep belief network for bearing diagnosis of electric locomotives. In this approach, an auto-encoder compressed the vibrational data to reduce the massive amount of data and learn the lower-level features. No feature extraction had been deployed and the dataset was directly split into training and test datasets. Then, high-layer features were extracted using a convolutional deep belief network with the use of Gaussian visible units by employing a two-dimensional structure and periodic characteristics of the input data. These binary units employed an exponential moving average to improve the performance of the model. Finally, Softmax was used for fault classification. The authors in this research employed this method on an electric locomotive bearing dataset with three faults and eight operating conditions including normal condition, slightly outer race fault, severe outer race fault, roller element fault, inner race fault, outer race, and inner race compound fault, inner race and roller compound fault and outer race and roller compound fault. Each working condition included 400 samples and each sample had 1024 data points. The result of this experiment presented an average accuracy of 97.4375% which is higher than traditional methods like SVM, BPNN, and ANFIS. Add to the point, the comparison between 200 training samples and 300 training samples demonstrated that the results were always over 95%. Therefore, it was concluded that this method can avoid over-

fitting. This proposed method was also validated on Bearing 3 of Testing 1 of the NASA dataset using 50 random samples and results demonstrated 100% accuracy. However, this accuracy shall be validated on the other two NASA datasets which include different fault types as well.

Wang *et al* (P. Wang et al., 2017) proposed a novel approach to gearbox fault detection by converting the time series data into graphical images. The integration of WT and DCNN detects and classifies the severity of the fault. WT can represent different resolutions in frequency and time domains, which can effectively detect faults from the early stage. Also, due to the window size variation, different faults can be illustrated. DCNN by analyzing the brightness of the specific pixels of the side bands detects underlying features and classifies them. The authors also used three different severity levels in their analysis, and for each one, two data series were collected of which 60 were used for training and the rest for the test. After converting the vibration signals into 2D images they used them as the inputs of the DCNN. The training was carried out for 50 epochs. The time taken to train the network was 4.25 s, and for classifying the test data was around 0.001 s. The network performed a classification accuracy of 99.58% Applying this method on other datasets to check the validation shall be considered for future work.

Jiang*et al* (Jiang et al., 2019) proposed an anomaly detection approach based on GAN architecture to overcome challenges within an imbalanced dataset. Industrial applications demonstrate imbalanced time series data where normal conditions are well higher than abnormal ones, which results in algorithms like SVM, and CNN can barely keep the accuracy as they perform on class-balanced properties. During an imbalanced condition, the classifier focuses on the majority class, and the detection gears toward normality. Thereby, GAN was proposed to resolve this challenge using training a generator and a discriminator for the detection. The generator employed an encoder-decoder-encoder three-sub-network, based on deep convolutional generative adversarial networks, and then a feature extractor was inserted between data and GAN to reduce the training time. During the training process, feature extraction on normal data was conducted, and then by using an anomaly detector the researchers obtained data distributions and potential representative modes of these features. This method was applied to vibration signals in both a laboratory-generated dataset, and rolling-bearing data from the CWRU dataset. The result of this method interestingly achieved the highest level of accuracy by 100% for both datasets and demonstrated that it can significantly distinguish the normal samples from abnormal ones. The method was implemented on a one-dimension dataset. Applying other time series data shall be taken into account for further analysis.

Sayanjit*et al* (S. S. Roy et al., 2020) presented an auto-correlation-aided feature extraction method for the diagnosis and classification of bearing defects. Auto-correlation was implemented on the healthy and faulty vibration signals to check their similarity. The best combinations could be recognized using several statistical, Hjorth, and non-linear feature parameters concerning the recursive feature elimination technique. Then the top 5 features were fed into a random forest classifier. In this approach, the noise in the preprocessing stage was eliminated because noisy signals can not impact the correlogram because their correlations are too small. The authors also evaluated this method in the CWRU dataset. Three classes of faults exist in this experiment, including an inner race with a 0.007-inch diameter, an outer race with a 0.014-inch diameter, and ball faults with a 0.021-inch diameter. The vibration signals were digitized at 12 kHz and 48 kHz. The signals were then partitioned into segments containing 6000 and 12000 data points. The results demonstrated that 36 features were extracted from vibration correlograms and then using recursive feature elimination reduction, the top 5 features were selected based on accuracy, sensitivity, and specificity for defect classification. The classification was made on Different Bearing Defect Sizes (12kHz), Different Bearing Defect Sizes (48kHz), Different Shaft Speeds (12kHz), and Different Shaft Speeds (48kHz). The overall classification results demonstrated that the model was able to perform the classification with 99.9% accuracy, 100% sensitivity, and 99.8% specificity for the 12kHz set and 99.2% accuracy, 100% sensitivity, and 99.5% specificity for the 12kHz set. The model was also validated on the NASA dataset with 97.9% accuracy, 99.4% sensitivity, and 95.8% specificity. The author used only Random Forest for the classification. Other algorithms can also be evaluated using auto-correlation-aided feature extraction to demonstrate a solid efficiency result from the proposed methodology.(Samimifard, 2023)

Ronak *et al* (Bhadra et al., 2018) developed a DL-based method including CNN for pattern recognition of the bearings faults. They employed the research on the NASA dataset to evaluate the results and compared it with the conventional machine learning methods. In conventional analysis, statistical features like RMS, skewness, and kurtosis are generated and normalized, and then k-NN, SVM, and RF are applied. On the other hand, CNN automatically extracts features in convolution layers from the raw signals. The authors also implemented both 1D matrix and 2D matrix in convolution layers. 1D matrix was also turned into a 2D matrix using the dimension of the nearest possible square. The results of deploying various algorithms presented 81.2% accuracy for KNN, 66.3% accuracy for SVM, 82.1% accuracy for RF, 93.3% accuracy for 1D-CNN, and 92.9% accuracy for 2D-CNN which demonstrated great performance for DL methods. The learning curves for 1D and 2D DL-based methods

demonstrated great results in achieving accuracy in classification after 900 epochs, which took 13 sec/epoch for 1D and 25 sec/epoch for 2D. Therefore, 1D performed better as it demonstrated higher accuracy and faster computational speed. In this research, the author evaluated the accuracy between different ML algorithms and CNN. Thereby, the evaluation between different DL algorithms shall also be considered.(Samimifard, 2023)

Mao *et al* (Mao et al., 2018) worked on detecting failures within bearings using Incremental Support Vector Machine and Stacked Auto-Encoder. In this combinational offline and online method, feature extraction using the faulty data, and the ISVM model was developed in the offline section, and in the online part, fault feature vectors were constructed using the network weight of offline SAE based on the new data. This method was implemented on the CWRU and IMS datasets. In the IMS dataset, 500 samples were extracted for each health condition with a learning rate of 0.2 and an offline training sample of 300. FFT was then used to develop frequency spectrum data from the raw signal and fed them to the Stacked Auto-Encoder. In the last step, traditional SVM and Incremental SVM were applied to the features. The classification results of this method illustrated 98.33% accuracy for the SVM, and 98.33% accuracy for the ISVM, which proved the feasibility of this method. A comparison between SVM and other classification algorithms concerning the Stacked Auto-Encoder and analyzing their performance shall be taken into account for the upcoming approaches.(Samimifard, 2023)

Roy *et al* (M. Roy et al., 2018) proposed an approach to an automated feature extraction method for online condition monitoring based on the stack of the traditional autoencoder and an online sequential extreme learning machine network. Three steps were involved in this procedure. In the first step, which was an offline approach, Autoencoder, which encompassed the same number of nodes in input and output layers of 4096 neurons, and a single hidden five-neuron layer were trained with smoothed input data. Next, online feature extraction was applied and these features were used as the input of OSELM with an input layer of 5 neurons, hidden layers with 10 neurons, and a single output node. The authors then applied this method to the NASA dataset and eliminated the noise by picking one in a row of file input features resulting in 4096 input features in each sample for each bearing. The result of this method demonstrated that it was capable of detecting faults with 100% accuracy. The authors suggest that as a further step, the classification performance needs to be evaluated using this method.

Purarjomandlangrudi *et al* (Purarjomandlangrudi et al., 2014) proposed a fault detection approach

called Anomaly detection, which used classification techniques. In this approach, the algorithm was trained with the normal data, and then, the learned algorithm was applied to a new dataset. The methodology was applied to the NASA dataset. The FFT spectrum was provided eight days before the element's full breakdown and represented major defects from day 5. Kurtosis and Non-Gaussianity Score were then extracted from the raw data with 50 ms length and 25 ms shift each. The anomalies were then detected using Gaussian distribution on extracted features, estimating parameters for a Gaussian, and Selecting the threshold. The result of this technique provided 95% detection accuracy which was higher than the accuracy resulting from SVM. Add to the point, this method provides early detection of 100h before the anomaly. The authors recommended applying this method to other rotating elements, or a comparison of this method with other available ml techniques to investigate the results for their complementary future work.(Samimifard, 2023)

Tao *et al* (T. Lu et al., 2020) proposed a bearing fault detection technique based on AlexNet using transfer learning. AlexNet which is a CNN architecture, is comprised of 5 convolutional layers and 3 fully connected layers followed by a softmax layer. Transfer learning is divided into a pre-trained network and the transferred network. The parameters in the former part were trained with available data and were used as feature extractors. Transfer learning overcomes the overfitting challenge by reducing parameters. This method is applied to the CWRU dataset. The preprocessing step comprised of generating a spectrogram by the Non-Uniform Fast Fourier Transform Hamming window, cutting spectrograms into 10ms duration using 5 pixels, and turning them 227*227*3 pixels which are input size of AlexNet. The result of this experiment demonstrated over 99.7% accuracy on classification, which demonstrated highly applicable classification performance.

Abbasi *et al* (Abbasi, Lim, & San Yam, 2019) worked on the PM of an Air booster compressor motor using RNN and LSTM. The involved parameters included current, active power, discharge temperature, air pressure, and vibration. Unlike traditional neural networks which are not suitable for time series prediction due to the independency of data with each other, RNN depends on the previous inputs. The data for this research was gathered within a two-year timeline. During this period the Air booster compressor motor failed 3 times for different reasons. In the preprocessing stage, the data were divided into healthy and unhealthy values containing shutdown, outliers, and out-of-range faults for the latter. Levenberg Marquardt and Bayesian Regularization algorithms were compared for the training. Training functions, the architecture of the model, and network parameters were the indicators of the model. The simulation results for the RNN-LSTM model presented that the Levenberg

Marquardt model generated lower RMSE and performed faster. The researcher trained the model with different epochs and the number of neurons in the hidden layer and the optimal number of neurons was 15. By increasing the number of neurons the RMSE increased. It was also increased after epochs reached 750 because the model was overfitted. The authors also presented that the model can predict potential faults in the ABC motor. The only disadvantage of this approach is the long training time. However, they could overlook this problem because network training was an offline procedure. An online approach shall be defined for further steps.

Wang *et al* (J. Wang et al., 2021) proposed a new multi-sensor information fusion method for fault classification using CNN. In this model, the time-domain vibration signals were turned into a two-dimensional matrix, which then was fed into CNN for classification. Each sampling corresponds to three data points representing data that were extracted from three different positions of the motor. To perform Signal-to-Image Conversion, a sliding window mechanism was adapted on 90 points with 45 points moving forward, and then (-1, 1) normalization was applied. This 180*3 image was then fed into CNN. The convolution layer contained 32 kernels with the zero-padding method to avoid dimension loss. In the max-pooling layer, a 2*2 window size was applied with the zero padding method. After this layer, a dropout layer was applied to avoid overfitting. The author has applied this method to CWRU, IMS, and the dataset from the designed bearing fault test rig to perform the classification. The result of this evaluation demonstrated 99.92% on CWRU, 99.68% IMS, and 99.25% on the test rig. In this research, a comparison between the accuracy of the deep learning algorithms was conducted only on the CWRU dataset. This evaluation shall be applied to two other datasets as well. This model was applied to the known types of faults. Therefore, to make this model applicable to the industry, the author recommended anomaly detection under different working conditions using the multi-sensor information fusion method. Also, a combination of this method with other methods is recommended.

Khorram *et al* (Khorram et al., 2021) presented an End-to-end 1-D CNN-LSTM deep learning method for the anomaly detection of the bearings. Time-series data contains local and global features. CNN extracted the local features and LTSM learned the long-term correlation between two variables and dealt with global features. To perform this combinational method, the dataset was transformed into equal-sized sequences before feeding to CNN. In CNN, a dropout layer was placed after each main layer to avoid overfitting, Batch-normalization layers were also applied to enhance accuracy and speed, and a sigmoid activation function was applied in the fully connected layer. The author

labeled the dataset into Healthy, Suspected, Inner-race-fault, and Rolling-element-fault and used 30 files for each class. The train test split was performed to be divisible by the batch size, thereby they suggested 25% for the training and 75% for the test. The result of this experiment presented 97.13% for the fault classification. This classification was performed on a limited dataset, thereby, applying it to the full dataset shall be taken into account. The author proposed implementing GRU-LSTM on the dataset and evaluating its performance. Also applying generative adversarial networks on small-scale datasets can be taken into account.(Samimifard, 2023)

Shao *et al* (Shao, Jiang, Wang, & Zhao, 2017) proposed a deep feature fusion method for anomaly detection. In this approach, the signals were divided into the training set and test set without any preprocessing. Deep auto-encoder as a combinational auto-encoder, used a denoising auto-encoder to provide low-layer feature learning and a contractive auto-encoder to learn deep features based on the previous step for the feature extraction. Thereby, the demand for manual feature extraction was eliminated. The locality-preserving projection was then used for their improvement. The result was then sent to Softmax for classification. This method was implemented on a Rotor fault test rig with seven working conditions. Each working condition contained 200 samples and each sample contained 1000 data points. The author used 70% of the samples for the training and the rest for the test. Three different experiments were conducted by the authors including Raw vibration data without manual feature extraction, 19 extracted features from eight frequency-band signals, and 19 extracted features from the most sensitive signals. The proposed method was then compared to the Standard deep DAE, Standard deep CAE, BP, and SVM in the three mentioned experiments. The results of this method presented 95.19% accuracy for experiment 1, 96.33% for experiment 2, and 97.10% for experiment 3, which proved the efficiency of this method over other evaluated methods and effectiveness for feature learning. The authors also suggested evaluating the accuracy and efficiency of the other deep learning algorithms based on this method.

Enshaei *et al* (Enshaei & Naderkhani, 2019) explored a method for anomaly detection of bearings within the induction machine using deep bidirectional LSTM. This method was then developed on the CWRU dataset and then divided by a 90% training set and a 10% test set. In this method, the patterns were expected to be classified to enable fault detection. The network architecture was comprised of 1 input layer, 1 Bi-LSTM layer with 100 neurons, 6 fully connected layers, a softmax layer, and 1 output layer. Hyperbolic tangent function and sigmoid as an activation function were considered in this approach. The classification results demonstrated that BiD-LSTM performed better in comparison with

the typical LSTM with a total accuracy of 98.3%. The result was also compared with the DNN and demonstrated 100% accuracy. However, the latency, which was considered an efficiency factor was increased as well. The authors believed that by increasing the number of LSTM layers, the accuracy could also increase. They also presented a hybrid network concerning the lowest latency to increase the accuracy of their future work.

Xiang *et al* (X. Li et al., 2019) proposed a method using Deep Transfer Learning to overcome the challenge of anomaly detection of unsupervised data. This method was conducted based on the knowledge extracted from different supervised data of the rotatory equipment on 4 different datasets including CWRU Dataset, IMS Dataset, Bogie Dataset, and Crack Dataset. Xavier normal initializer was used to obtain weights and biases and the BP algorithm for the parameters update. In this approach, IMS Dataset, Bogie Dataset, and Crack Dataset were considered learning sources, and Gaussian noises were deployed to produce source samples. The signal-to-noise ratio was then employed to produce the same amount of noisy data. The DNN was used for the transfer learning and the results of applying this method presented that higher accuracies would be achieved using more target-supervised data. Moreover, it represented that at least 70% accuracy could be achieved in extreme cases. In this research, the authors focused on the feasibility of the diagnosis using limited supervised data and proposed applying this method to real-industrial data for their future work.

Guo *et al* (Guo et al., 2018) demonstrated an approach to rotating element fault detection using deep convolutional transfer learning, and unlabeled data. The proposed method is comprised of two steps. The authors first used a 1-D CNN for feature extraction and condition classification. 16 layers including one input layer, six convolutional layers, six pooling layers, two fully connected layers, and one output layer were developed in this neural network. Then, domain-invariant features were applied to a 1-D CNN. Thereby, the deep convolutional transfer learning network was expected to perform unlabeled classification with regard to other labeled data. CWRU Bearing Dataset, IMS bearing dataset, and Railway Locomotive Bearing Dataset were used to develop six anomaly detection trials, and each transfer was repeated 10 times. The results of deploying this method demonstrated that the accuracy of the anomaly detection based on each transfer could reach over 82%, which demonstrated the effectiveness of this method. This method was then compared with no transfer learning, handcrafted-feature-based transfer learning, and state-of-the-art transfer learning including DDC and DANN, and demonstrated higher results in each trial. applying this method using online data during the second step can be taken into account for industrial validation.

Ramin*et al* (Rajabioun et al., 2023) proposed a method for bearing fault classification using a multi-sensor approach. 2D DL frameworks were also developed to perform the classification based on these multi-sensor data. In this approach, six different signals have been merged and created a matrix which can then be fed into the Neural networks. The authors have turned the 1D signals into the 2D matrix to enable the 2D CNN model to identify spatial patterns. The authors have applied this method to a costume set of 2 induction motors in one healthy state and 5 fault conditions and generated 300 experiments. The results demonstrate that by incorporating multi-sensors the results increased to 96% classification accuracy.

Sayanjit*et al* (S. S. Roy et al., 2020) presented an auto-correlation-aided feature extraction method for the classification of bearing defects. This model was implemented on the healthy and defective vibration signals to check their similarity. The best combinations could be recognized using several statistical, Hjorth, and non-linear feature parameters with regard to the recursive feature elimination technique. Then the top 5 features were fed into a random forest classifier. This method was evaluated in the CWRU case study and the results demonstrated that the model was able to perform the classification with 99.9% accuracy. The model was also validated on the NASA dataset with 97.9% accuracy. The author used only Random Forest for the classification. Other algorithms can also be evaluated using auto-correlation-aided feature extraction to demonstrate a solid accuracy result from the proposed methodology.

Although various scholars have extensively worked on anomaly classification methods and tried different DL algorithms, our research provides a unique perspective within the field. Most of the scholars have been working on single models, single features, or even one dataset, thereby, they were not able to provide a comprehensive analysis of how different models would differ in performance and how to increase the performance. In this thesis, we provide a comprehensive study on bearing anomaly classification using AlexNet, EfficientNet, NasNet, ResNet, and VGG16 and evaluate how different features like spectrograms, Frequency spectrums, and polar spectrums can affect performance. Add to the point, through incorporating the fusion methods and using two channels within the models, we can highlight the significance of this method, specifically on EfficientNet, and prove their efficiency in anomaly classification and their contribution in increasing the accuracy.

Author	Dataset	Feature	Model	Accuracy
Wang *et al* (H. Wang et al., 2022)	CWRU	highest envelope spectral kurtosis	Multi-sensor 1D fused VGG16	100%
Shao*et al* (Shao, Jiang, Zhang, & Liang, 2017)	Electric Locomotive Bearing Dataset	vibrational data	CDBN	97.43%
Wang *et al* (P. Wang et al., 2017)	Gearbox data	Wavelets	DCNN	99.58%
Jiang*et al* (Jiang et al., 2019)	CWRU	Time series data	GAN	100%
Sayanjit*et al* (S. S. Roy et al., 2020)	CWRU-NASA	Auto correlation on vibrational data	Random forest classifier	99.2%-97.9%
Ronak *et al* (Bhadra et al., 2018)	NASA	Statistical features	1D CNN - 2D CNN	93.3%- 92.9%
Mao *et al* (Mao et al., 2018)	CWRU	Frequency spectrum	ISVM	98.33%
Roy *et al* (M. Roy et al., 2018)	NASA	Auro encoder	OSELM	100%
Purarj *et al* (Purarjomandlangrudi et al., 2014)	NASA	FFT spectrum - kutrosis - Non Guassianility score	Gaussian distribution	95%
Tao *et al* (T. Lu et al., 2020)	CWRU	Spectogram	AlexNet	99.7%
Wang *et al* (J. Wang et al., 2021)	CWRU - NASA	signal to image	Multi sensor fusion usning CNN	99.92% - 92.68%
Khorram *et al* (Khorram et al., 2021)	NASA	time series data	CNN-LSTM	97.13%
Enshaei *et al* (Enshaei & Naderkhani, 2019)	CWRU	raw vibrational data	deep deep bidirectional long short-term memory	100%
Xiang *et al* (X. Li et al., 2019)	CWRU	deep transfer learning	DNN	70%
Guo *et al* (Guo et al., 2018)	CWRU	raw signal data	1D CNN	82%
Ramin*et al* (Rajabioun et al., 2023)	self generated data	raw signal	Multi sensor fusion	98%
Sayanjit*et al* (S. S. Roy et al., 2020)	CWRU - NASA	Auto correlation	random forest classifier	99.9% - 97.9%

Table 1: literature review summary

4 Research Methodology

In this section, we explain the research methodology employed to achieve the objective of this thesis. In the first subsection, we describe the datasets and used preprocessing techniques. Subsequently, we present the architectures of the used 2D-CNN models for the classification of the various bearing faults.

4.1 Datasets and Data pre-processing

In this thesis, two datasets were considered. The first dataset used is the CWRU dataset which was collected by Case Western Reserve University (CWRU) and the other one is the NASA dataset. Both datasets are publicly available and contain bearing vibration signals that accommodate different fault conditions.

4.1.1 CWRU dataset

As illustrated in Figure 7, Within this setup, bearings are connected to a shaft with a rotation speed of 1772 rpm. Different fault types including inner race, rolling element, and outer race in different sizes from 7mm to 40 mm are imposed on them using an electro-discharge machine. The data was recorded using accelerometers which were located at the drive end and fan end of the motor and underwent post-processing in Matlab. The vibrational signals were collected within different runs using 12000 and 48000 sampling rates and generated the dataset. The data within this dataset were in Matlab format. In this thesis, we are working on a drive-end subset of the CWRU dataset with a 12000 sampling rate that was captured from the bearing located at the drive end of the shaft. Table 2 provides an overview of the CWRU dataset. As illustrated in this table, different defect types of inner race defect, outer race defect, and rolling element defect and their corresponding defect size of 7mm, 14mm, and 21mm defect sizes have been involved in our experiment to perform classification on the three defect types (X. Zhang, Zhao, & Lin, 2021) .

4.1.2 CWRU Dataset preprocessing

The signals within the CWRU prelabeled dataset were captured using a 48000 sampling frequency. We have turned the raw signals into 1-second samples with 4800 data points. As illustrated in Figure

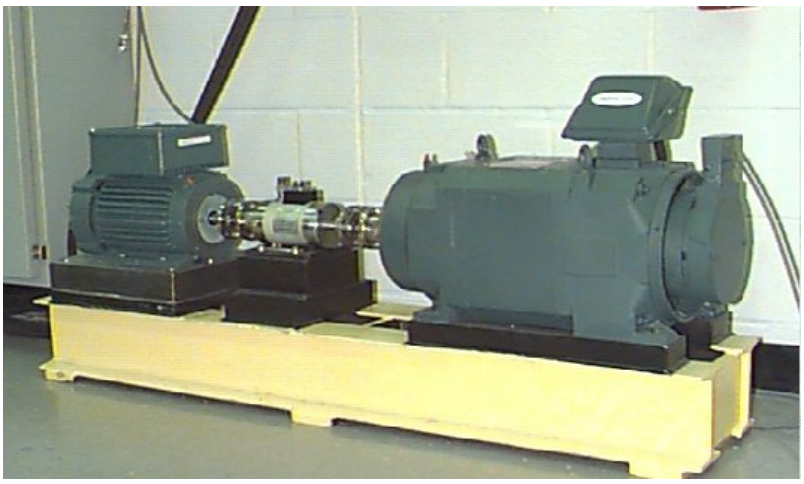


Figure 7: CWRU Bearing test rig (Saufi et al., 2019b).

Defect diameter	Defect	Number of data points
7 mm	Inner race defect	486224
7 mm	Outer race defect	486804
7 mm	Ball pass defect	487384
14 mm	Inner race defect	381890
14 mm	Outer race defect	484483
14 mm	Ball pass defect	486224
21 mm	Inner race defect	489125
21 mm	Outer race defect	489125
21 mm	Ball pass defect	486804

Table 2: The excerpt of CWRU Dataset.

8, we have made the division based on a 50% window size. In this demonstration, every second sample overlaps with the second half of the previous sample. Overlapping results in a more consistent sampling by capturing the details at the boundaries and also providing more samples to be used for classification in the neural networks.

4.1.3 NASA dataset

The NASA dataset collection setup can be seen in Figure 9. To collect the NASA data, four Rexnord ZA-2115 bearings were installed in one shaft with a Pitch diameter of 71.5mm, a Rolling element diameter of 8.4mm, 16 rolling elements per row, and a 15.17 contact angle. The shaft was driven by an AC motor and coupled by rub belts with 2000 rpm rotation speed and a radial load of 6000 lbs. The bearings were lubricated using an oil circulation system. The vibrational data were collected along the X and Y axis with a sampling rate of 20 kHz and 20480 data points for each sample. During this run-to-failure experiment, three faults were detected including roller element fault, inner race fault, and outer race fault (Qiu, Lee, Lin, & Yu, 2006).

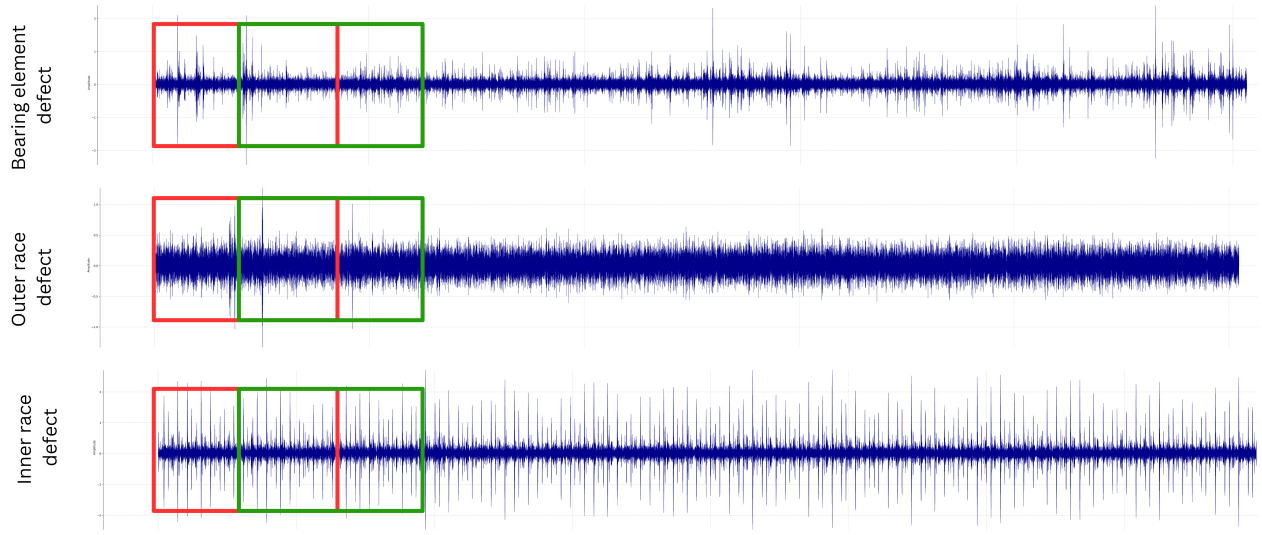


Figure 8: Time domain signals with 14mm fault size.

Three different datasets were produced with the following details mentioned in detail in Table 3. In this table, duration is the running time until the failure of the bearing. That means it represents a full life cycle of the bearing before it fails functional activity. IMS-Rexnord Bearing Data.zip file contains three different datasets. Table 3 provides some details about each dataset regarding the type of failure that occurred during this run-to-failure experiment and the duration of each experiment. Each file also is comprised of 1-second vibrational signals with 20,480 points. The interval column describes the interval time between two data samples and the fault section represents which fault in which bearing resulted in the bearing functional failure.

Dataset	Duration	Number of files	Channels	Interval	Fault
Dataset 1	34 days 12h	2156	8	First 43, 5min rest 10 min	Inner race Bearing 3 Roller element Bearing 4
Dataset 1	6 days 20h	98	4	10min	Outer race Bearing 1
Dataset 2	31 days 10h	4448	4	10min	Outer race Bearing 3

Table 3: NASA Dataset detail (Samimifard, 2023).

4.1.4 NASA dataset pre-processing

In the NASA dataset, the observations are not labeled with the state of the bearings. As the experiments have been conducted on the full life cycle of a bearing, we need to initiate the defective signals within them. We use the generated measurements of BPFI, BPFO, and BSF, as our expected

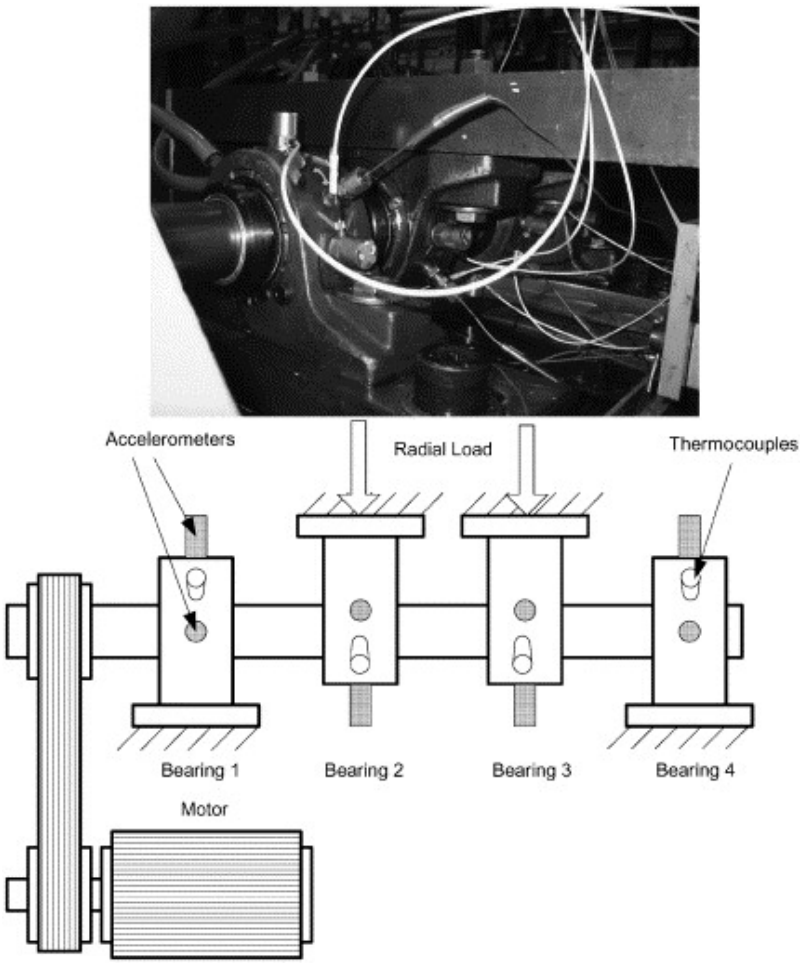


Figure 9: Bearing test rig. (Qiu et al., 2006)

frequencies, and the frequencies and amplitudes derived from the envelope of the signal, to detect the highest peak within each observation (Samimifard, 2023).

Figures 10, 11, and 12 illustrate, the amplitude of the highest peaks for all the observations within an experiment. A sudden increase in the amplitudes can highlight the presence of a defect. As the following high peaks demonstrations present, we can observe an initial point of 432 for the defective observations of bearing 1 in experiment 2 for the outer race defects, 2110 for the defective observations of bearing 3 in experiment 1 for the inner race defects, and 1237 for the defective observations of bearing 4 in experiment 1 for the rolling element defect. Using these initial points, we can label our observations with their corresponding defect type.

4.2 Feature Extraction

In order to turn the raw signal into a structure that can be efficiently used for classification, we have contributed by extracting statistical features and transforming them into spectral representation. Ex-

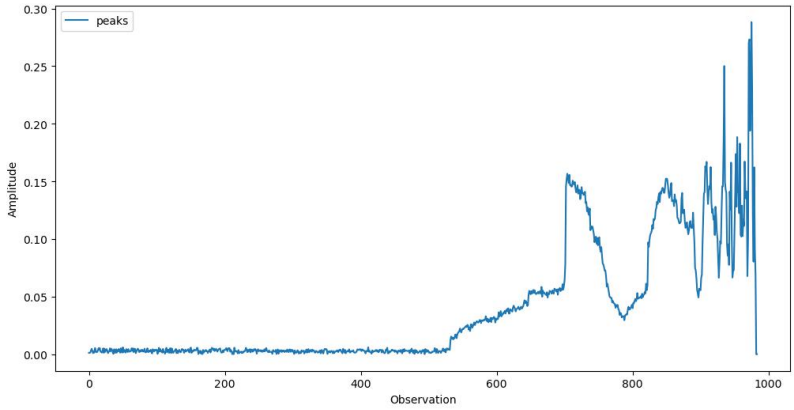


Figure 10: High peaks generated from bearing 1 in dataset2.

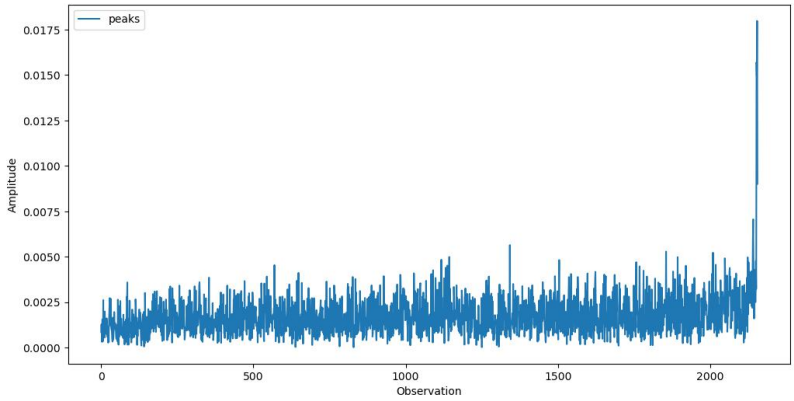


Figure 11: High peaks generated from bearing 3 in dataset1.

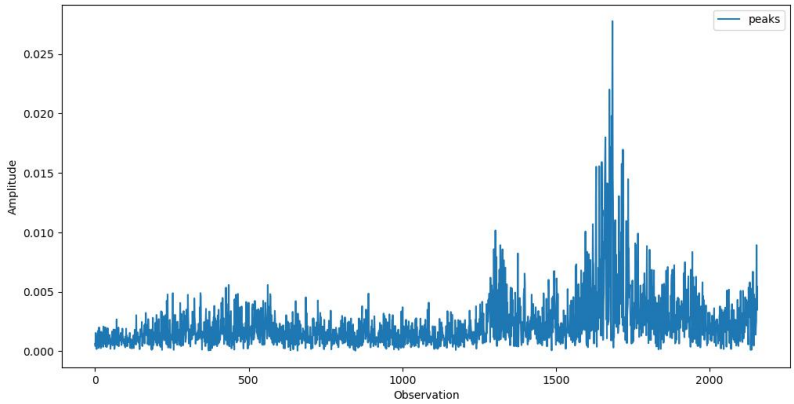


Figure 12: High peaks generated from bearing 4 in dataset1.

tracting the most important features enables us to leverage them to our benefit by reducing the number of data points, or revealing distinguishable patterns.

4.2.1 Statistical Features

Different statistical features have been used to provide in-depth insights about the state of the bearings.

- **Mean:** Demonstrates the average of the data points within the signal. Any change within this attribute over time can determine the presence of a defect within the bearing.

- **Skewness:** It evaluates the asymmetry within the signal and any metrics except zero represent an anomaly within the signal. (Sacerdoti, Strozzi, & Secchi, 2023)

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3 \tag{12}$$

- **Kurtosis:**This attribute measures the tail of a distribution. A high value in the measurement of the tail can indicate a defect within the bearing (Dey & Jana, 2022).

$$Kurtosis = \frac{\sum_{i=1}^{n} \left(x_i - \mu \right)^4}{(n-1)\sigma^4} \tag{13}$$

- **RMS (Root Mean Square):** It demonstrates the energy of a signal and can be used as an identifier of any damage within the bearing. In the following equation. (Dey & Jana, 2022)

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \tag{14}$$

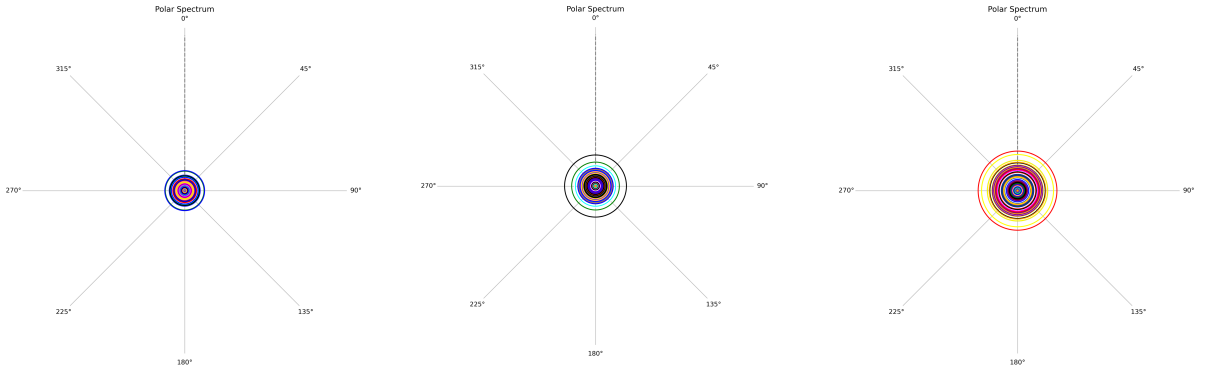
- **Standard Deviation:** This attribute identifies variability within a signal. An incremental value of Standard Deviation can represent an abnormal activity in the signal. It is represented as the following where each data point is defined as x_i, n is the number of data points, and μ is the mean value of all data points (Dey & Jana, 2022) :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \mu \right)^2}{n-1}} \tag{15}$$

4.2.2 Polar Spectrums

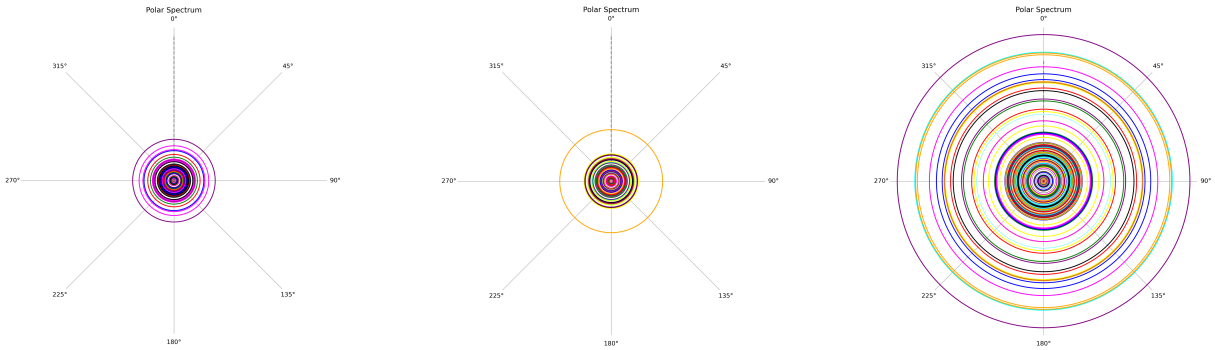
Fourier transformation illustrates the frequency content in a time domain signal. It provides the corresponding magnitude and the phase of the frequency. By mapping the frequency spectrums into polar coordinates, we can illustrate the magnitude and phase of the frequency using the radial distance and the angle respectively. It plots a circular representation using different amplitudes and their corresponding frequencies. It also provides us with information concerning both strength and timing which are pivotal for analyzing the signal behavior. Figures 13, 14, and 15 demonstrate the polar spectrums for different fault conditions and different severities within the bearing. As illustrated

in these Figures, polar spectrums can be determinative in identifying frequencies related to specific defects within the bearing.



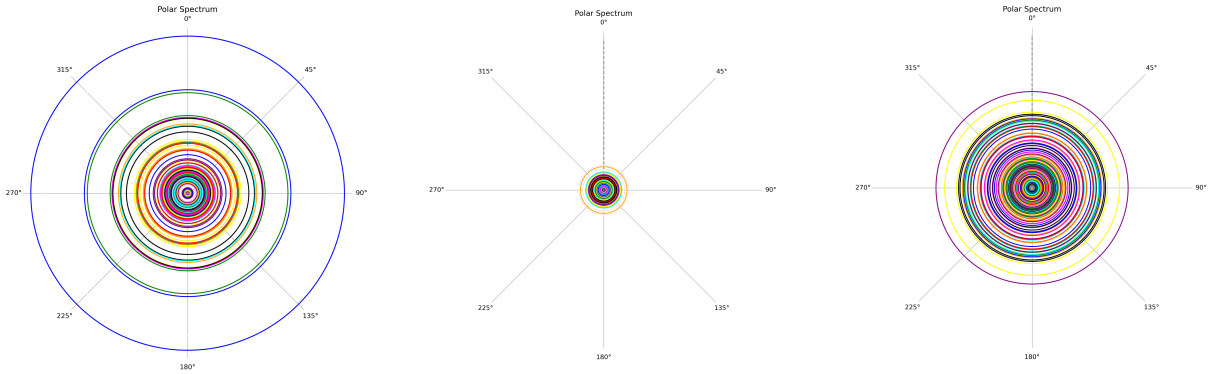
Ball Bearing defect (7mm) Ball Bearing defect (14mm) Ball Bearing defect (21mm)

Figure 13: Polar spectrums of ball bearing defect.



Inner race defect (7mm) Inner race defect (14mm) Inner race defect (21mm)

Figure 14: Polar spectrums of inner race fault.

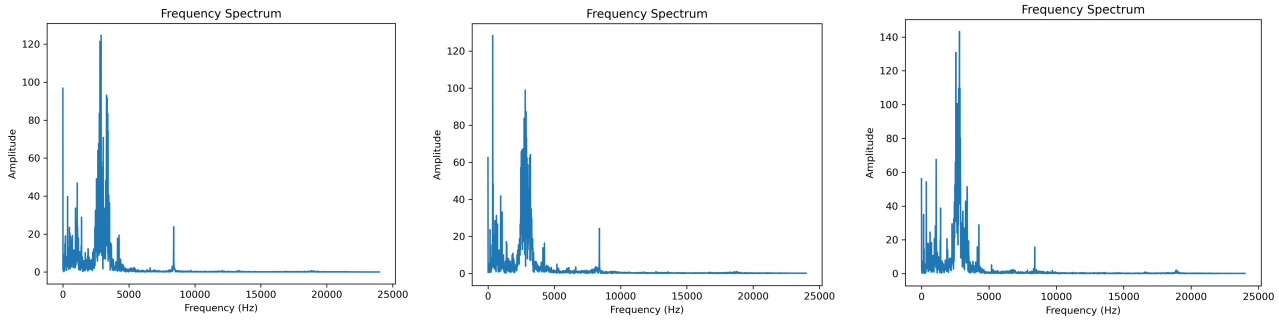


Outer race defect (7mm) Outer race defect (14mm) Outer race defect (21mm)

Figure 15: Polar spectrums of outer race fault.

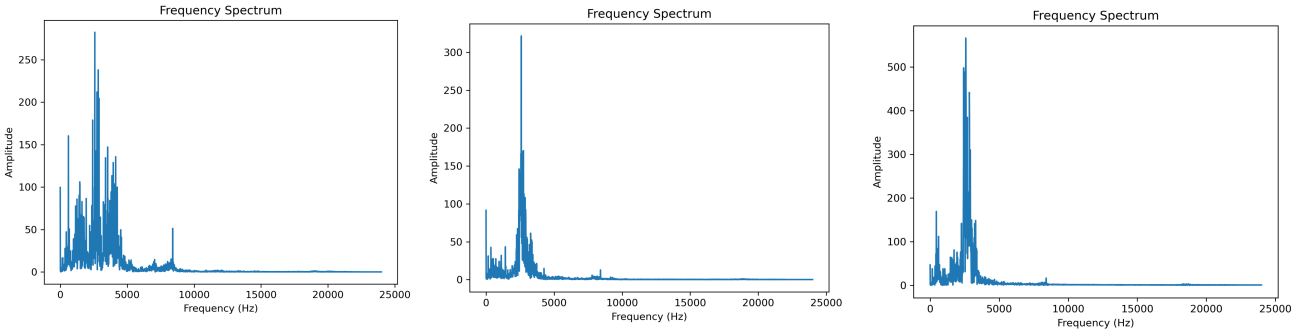
4.2.3 Frequency Spectrums

The frequency spectrum (FFT) illustrates the signal frequencies and their corresponding amplitudes. Using FFT, we convert time-domain signals into the frequency domain to uncover distinct patterns within the signal. Figures 16, 17, and 18 demonstrate the frequency spectrum for different fault conditions and different severities within the bearing. Add to the point, by analyzing features like the frequency and amplitude of peaks within the frequency spectrums we are able to identify the defects within the signal.



Ball Bearing defect (7mm) Ball Bearing defect (14mm) Ball Bearing defect (21mm)

Figure 16: Frequency spectrums of ball bearing defect.



Inner race defect (7mm) Inner race defect (14mm) Inner race defect (21mm)

Figure 17: Frequency spectrums of Inner race defect.

4.2.4 Spectogram

Spectrograms are visualizations that demonstrate how frequencies change over time. In this 2D visualization, the X-axis corresponds to the time, the y-axis corresponds to the frequency, and the color of each point is the amplitude of the frequency. It is a combination of time-domain and frequency-domain representations. Figures 19, Figure 20, and Fig 21 demonstrate the frequency spectrum

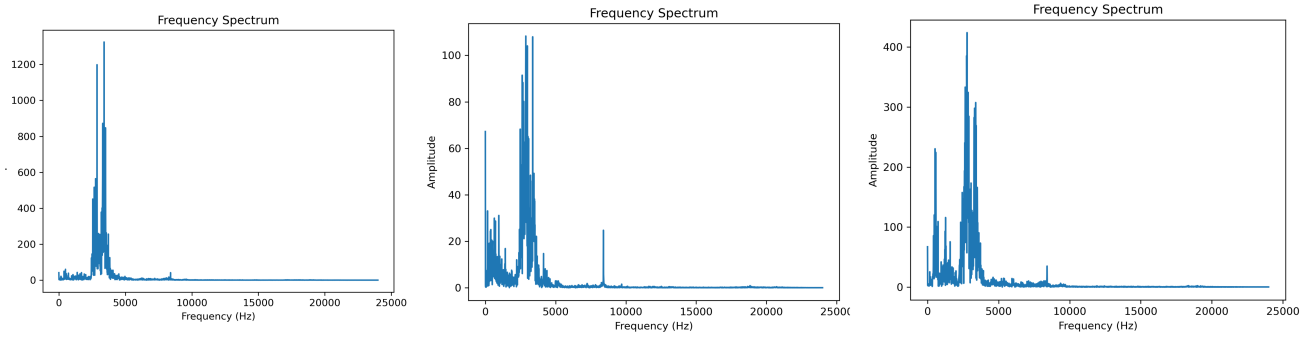


Figure 18: Frequency spectrums of Outer race defect.

for different fault conditions and different severities within the bearing. In these representations, brighter regions demonstrate higher amplitudes and darker regions demonstrate lower amplitudes. Add to the point, vertical and horizontal patterns. Add to the point, it may contain vertical and horizontal patterns that may demonstrate sudden events or consistency over time.

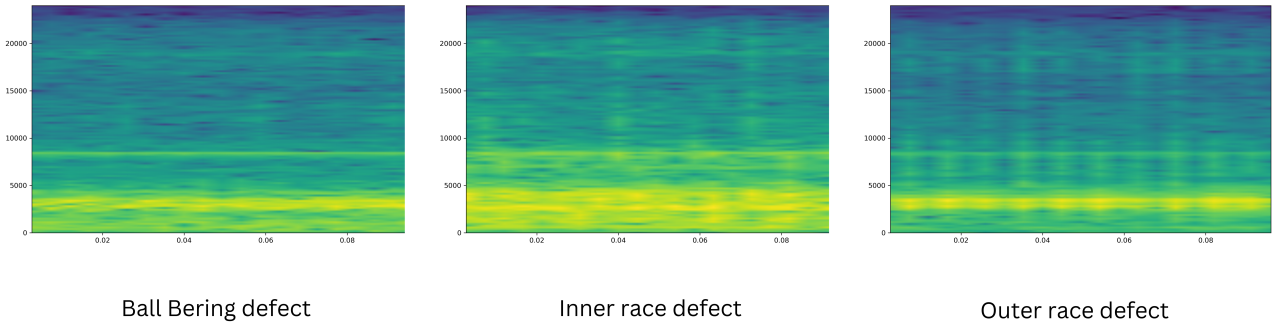


Figure 19: Spectogram of ball bearing defect.

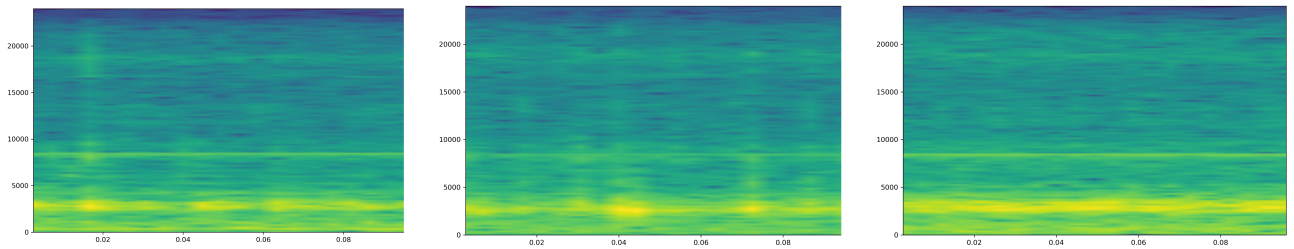


Figure 20: Spectogtam of Inner race defect.

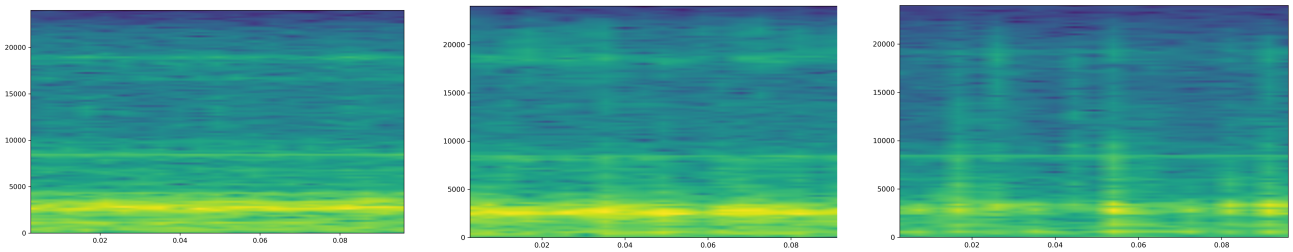


Figure 21: Spectogram of Outer race defect.

4.3 Network Architectures

The six different 2D CNN models have been utilized for the classification of the bearing faults. The experimental setup for all the used CNN models are described below.

4.3.1 AlexNet

Alexnet is a deep learning model that has revolutionized the Convolutional Neural Network. As illustrated in Figure 22, this model consists of 5 convolutional layers. The first convolutional layer filters the input image with 96 kernels of size 11×11 and a ReLU activation function. A local response normalization is applied to improve the performance of edge detection and a max pooling layer is applied to minimize the spatial size. The subsequent two convolutional layers include 256 kernels of size 5×5, and 384 kernels of 3×3 size respectively with ReLU activation, and the kernel sizes for Conv4, and Conv5 are 384 and 256. Then the max pooling layer is again applied and the layers are then flattened.

In the next step, three fully connected layers are integrated where the first two have 4096 neurons with a dropout layer to avoid over-fitting. The last fully connected layer with 256 kernels of size and softmax activation function demonstrates the probability over the predictable classes (Alzubaidi et al., 2021; Dhillon & Verma, 2020).

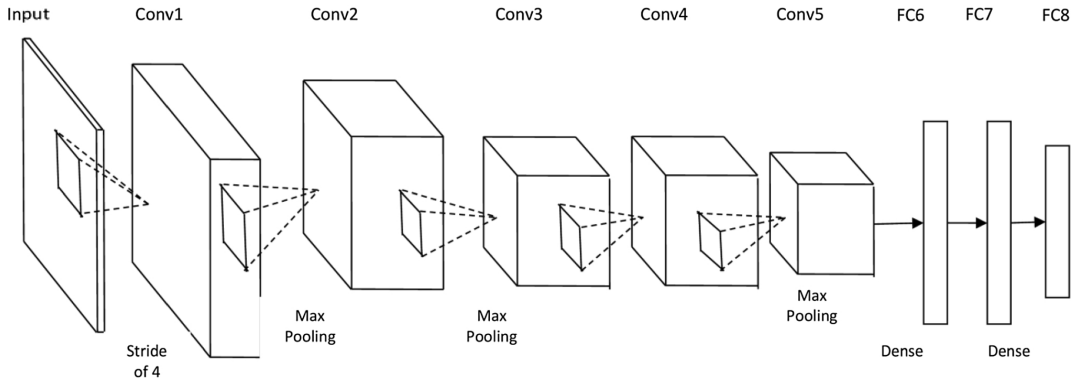


Figure 22: Alexnet Architecture(Dhillon & Verma, 2020).

4.3.2 LeNet

As illustrated in Figure 23, the first convolutional layer applies 6 filter kernels of size 5×5 and a ReLU activation function. In order to decrease the dimensions of the feature map a pooling layer is applied using 2×2 window and stride of 2. Then next convolutional layer applies 16 filter kernels of size 5×5 and a ReLU activation function. Spatial dimensions within the feature map are again decreased using a pooling layer with 2×2 window and flattened into a 1D vector, and then fed into 2 fully connected layers with 120 and 84 neurons and ReLU activation function. A dropout is then applied for the regularization and the last fully connected layer performs the classification (Patel, 2020; Dhillon & Verma, 2020).

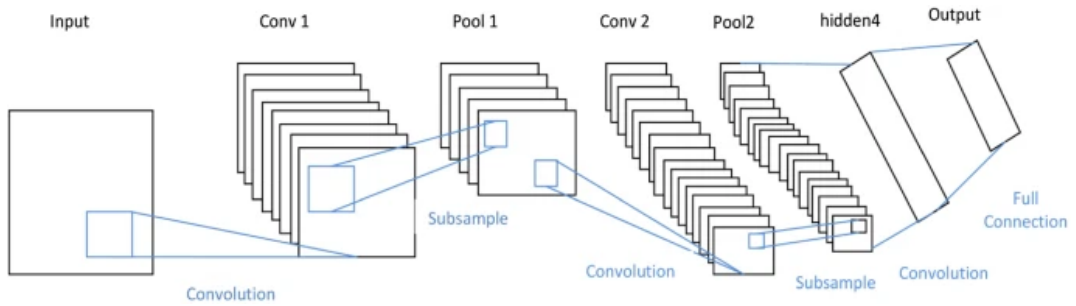


Figure 23: LeNet Architecture(Dhillon & Verma, 2020).

4.3.3 NASNet

This architecture has been designed to automatically select the most efficient model with the highest performance. As illustrated in Figure 24, this method is implemented based on normal cells and reduction cells in which the first one preserves the spatial dimension and the latter halves the dimension. Each one of these cells processes the data using specific convolutions, identity mappings, and pooling methods. As cells are the building blocks of NASnet, the input goes through a series of these two cell types, and their feature maps are then concatenated to provide more informative features from the raw data. These features are then passed through the Global Average Pooling layer to produce a 1D vector. The 1D vector is then sent to the dense layer with 128 neurons and the ReLU activation function. A dropout layer is then applied and the processed features have resulted in the classification using a dense layer and softmax activation function (Radhika et al., 2020; Dhillon & Verma, 2020).

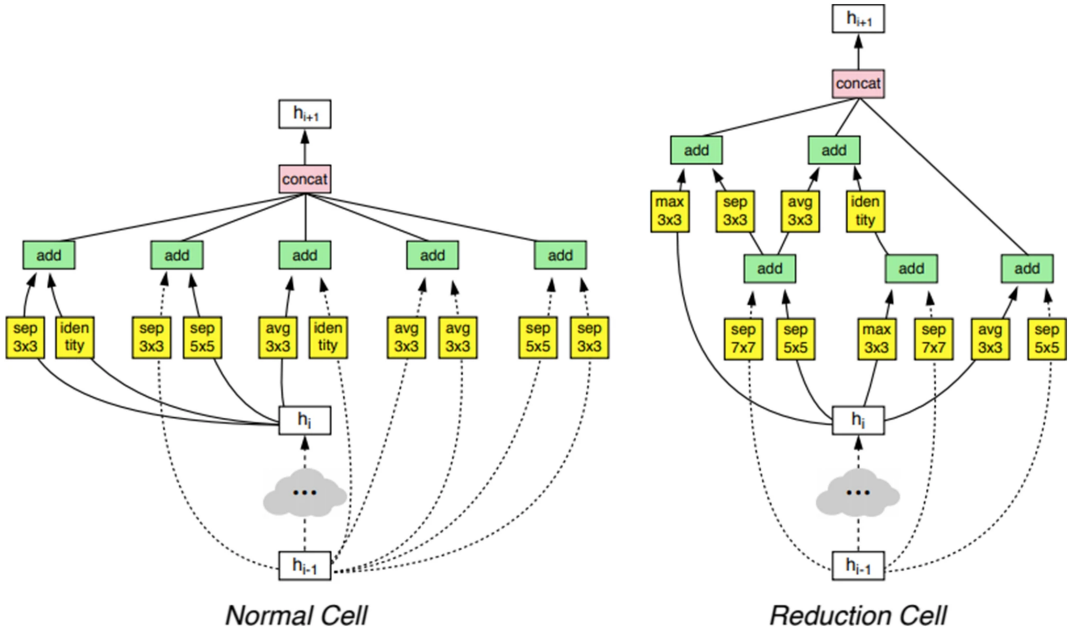


Figure 24: NASnet Architecture(Dhillon & Verma, 2020).

4.3.4 ResNet

The Residual Network has been introduced to minimize the vanishing gradient problem by skipping certain layers. In this architecture, the first convolutional layer captures the main features and then is followed by a pooling layer for spatial dimension reduction. ResNet is comprised of a series of residual blocks and each block is also constructed using a few convolutional layers. Our specific ResNet model, which is Resnet50 has 49 convolutional layers. In the last step, global average pooling is ap-

plied to reduce the spatial dimension of the feature map. The last layer is a fully connected layer demonstrating the probability over the classes using the softmax activation function (Alzubaidi et al., 2021; Patel, 2020; Sapijaszko & Mikhael, 2018).

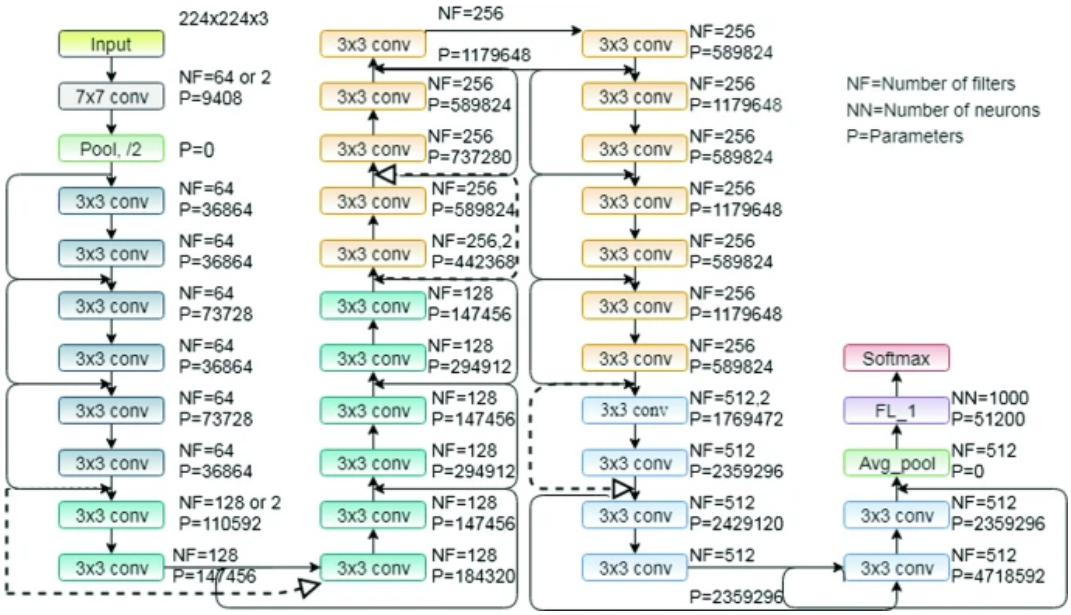


Figure 25: ResNet Architecture(Radhika et al., 2020).

4.3.5 VGG16

As illustrated in Figure 26, this network with a deep architecture is constructed with 13 convolutional layers and utilizes 3×3 filter with a stride of 1 and ReLU activation function. 5 max-pooling layers with a stride of 2 have been applied among convolutional layers to reduce the spatial dimensions of the feature map. Three fully connected layers are also included in this architecture using 4096 units for the first two and 4096×3 units for the last one. In the last fully connected layer a softmax activation function is applied to evaluate the probability over the distinct classes (Radhika et al., 2020).

4.3.6 EfficientNet

This model is an improved version of CNN by the width, depth, and resolution of the network. In this context, depth is the number of layers, width is the number of channels in each layer and resolution is the quality of the image. As illustrated in Figure 27, this base model is comprised of 7 blocks. The first convolutional layer has a 3×3 kernel. The next layer, which is a Mobile Inverted Bottleneck Convolution is equipped with a 3×3 kernel and scale factor of one. Three blocks with a 3×3 kernel and scale factor of 6 and two blocks with a 5×5 kernel and scale factor of 6 are included in this architecture

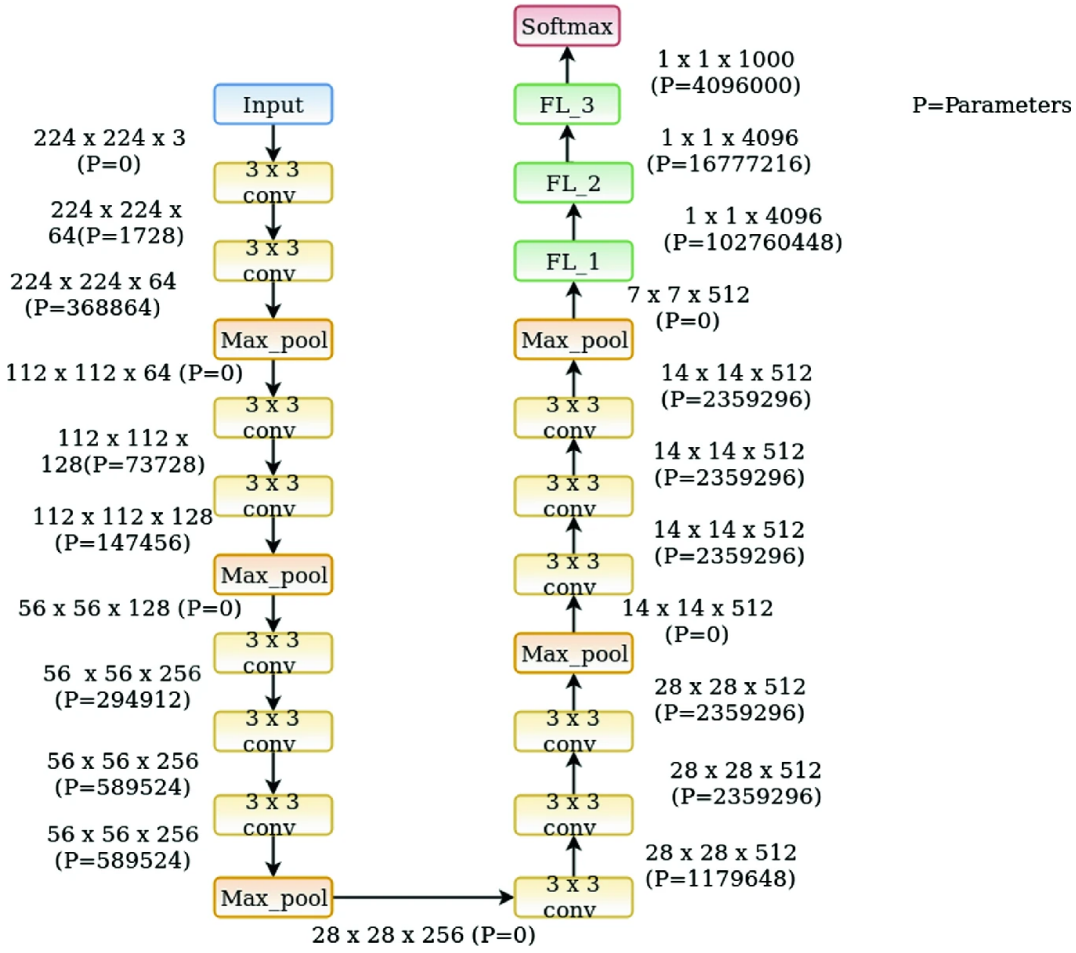


Figure 26: VGG 16 Architecture(Radhika et al., 2020).

(Goutham, Sameerunnisa, Babu, & Prakash, 2022). After the base model, a global average pooling is applied to reduce the dimension of the spatial features. Then it will be forwarded to a fully connected layer with 128 neurons and a ReLU activation function. In the next step, a drop-out layer is applied to avoid overfitting, and in the final step, a fully connected layer and a softmax activation function the distribution over classed will be generated.

Figure 27: EfficientNet Architecture(Bazi et al., 2019).

4.3.7 Fusion Technique

As demonstrated in Figure 28, the fusion method is empowered with features from two distinct data sources. These two channels are fed with statistical features and varied 2D representations. In the first channel, the statistical features pass through the statistics branch with a 1D CNN architecture. This channel is comprised of 2 convolutional layers, 2 max-pooling layers, 1 flattened layer, and 1 fully connected layer with ReLU activation function. The next branch extracts the features within the 2D representation using a 2D CNN architecture as the images pass through this channel. The extracted features within these two channels are then concatenated and fed to a fully connected layer with 256 neurons and a ReLU activation function following a dropout layer. In the last step, a final fully connected layer with a softmax activation function is applied to predict the distribution over the distinct classes.

4.4 Model Training Framework

4.4.1 Data preparation

The model development framework initiates with operations that effectively transform the data before feeding them to the neural networks. Statistical data are fed to the 1D branch, thereby, they need to undergo certain steps that turn them into a format that can be interpreted by the 1D CNN model. These steps start with normalizing the various statistical features derived from the bearing signals using StandardScaler which is provided by Scikit-Learn library.

The second channel interprets the 2D images extracted from the signal including the frequency spectrum, polar spectrum, and spectrogram. These images need to be resized into a standardized configuration before being sent to each variance of 2D CNN architecture. Add to the point, data augmentation techniques including rotations, width, height shifts, and horizontal flips are also applied to enhance the diversity patterns. Further, the associated labels related to each input are then turned into integer format using LabelEncoder and then transformed into one hot encoded structure.

4.4.2 Model training

In order to evaluate the performance of the model, the available data needs to be split into two distinct units. The train-test split empowers the model to be able to evaluate the effectiveness of the

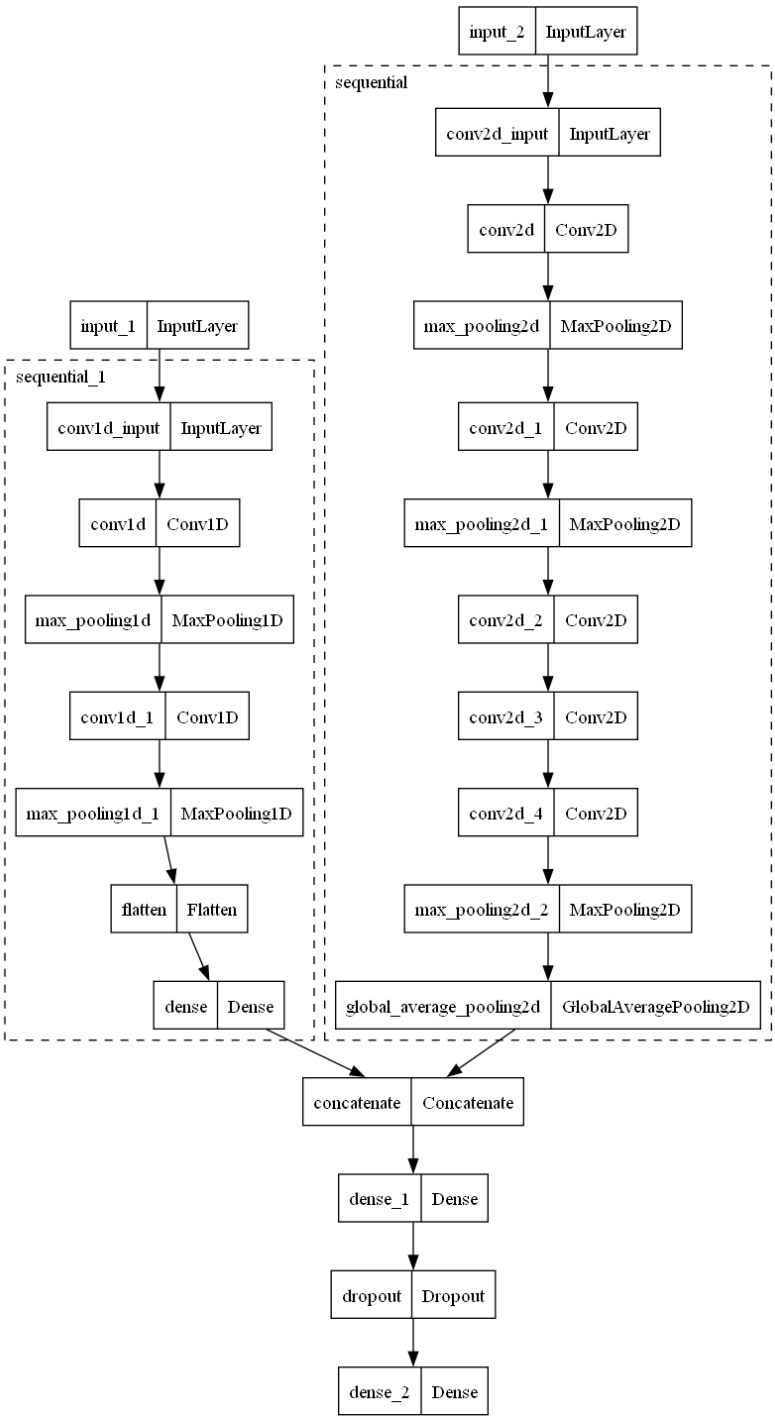


Figure 28: General structure of the fusion method using AlexNet and 1D CNN.

model on a new set of data, therefore we split our data using an 80:20 ratio, in which 80% of our data is used to train our model and 20 % of the data is used for testing the model. We are dealing with an imbalanced set of data faults, and we have to address this challenge by calculating the weight of each class and train our model using the frequency of the weight. By using this technique we can make sure that the model will not rely on the larger set of the defects type, and consider equality while training.

We used hyperparameter tuning to find the model with the highest performance. Using hyperparameter tuning We evaluated the model by different combinations of epochs and batch sizes. We have used batch sizes of 32 and 64, and in order to avoid overfitting, we have considered small sizes of epochs of 20 and 30. During the training process, we used the Adam optimizer to adjust the learning rate during the training. As we have to perform multiclass classification, we have used sparse-categorical-cross entropy as our loss function. In order to evaluate the performance, we have determined accuracy as our metric which provides us with insight into the efficiency of the model. To effectively avoid overfitting, we also have implemented.

4.4.3 Validation and Testing Strategy

During the training process, we allocated 20% of the training data for the validation which is evaluated at the end of each epoch. Evaluating this distinct set of data enables us to assess the performance of the model during the training set and avoids overfitting by applying early stopping in our model with the patience set to 3 epochs. After the model is trained, we need to test the performance of the test set. The following performance metrics are then calculated to assess the efficiency of the model.

- **Accuracy**: As demonstrated in the following formula, accuracy is the ratio of true predictions to all true and false results (Latif, Zou, Idrees, & Ahmad, 2020).

$$\text{Accuracy} = \frac{T_{\text{Pos}} + T_{\text{Neg}}}{T_{\text{Pos}} + T_{\text{Neg}} + F_{\text{Pos}} + F_{\text{Neg}}} \tag{16}$$

- **Precision**: As described in the formula, precision is the ratio of the true positives to the sum of true positives and false positives (Latif et al., 2020).

$$\text{Precision} = \frac{T_{\text{Pos}}}{T_{\text{Pos}} + F_{\text{Pos}}} \tag{17}$$

- **Recall**: The following recall formula, calculates the ratio of the true positives to the sum of true positives and false negatives(Latif et al., 2020).

$$\text{Recall} = \frac{T_{\text{Pos}}}{T_{\text{Pos}} + F_{\text{Neg}}} \tag{18}$$

- **F1-Score**: This metric demonstrates the balance between precision and recall by calculating

their weighted average (Latif et al., 2020).

$$F1\text{-Score} = \frac{2 \times (Precision + Recall)}{Precision + Recall} \tag{19}$$

- **Confusion matrix**: This 2D matrix demonstrates the true label and the predicted label derived from the classifier using the following algorithm. In the following formula r and c represent the row and column, m is the total number of samples within the test set, and x_i is the input to the classifier (Heydarian, Doyle, & Samavi, 2022). Figure 29, as a sample, demonstrates the confusion matrix derived from the dual channel AlexNet using spectrograms on the CWRU dataset.

$$M(r,c) = \sum_{i=1}^{m} (I(y_i = r)\ I(h(x_i) = c)), \tag{20}$$
$$\forall r, c \in \{0, \ldots, q-1\}$$

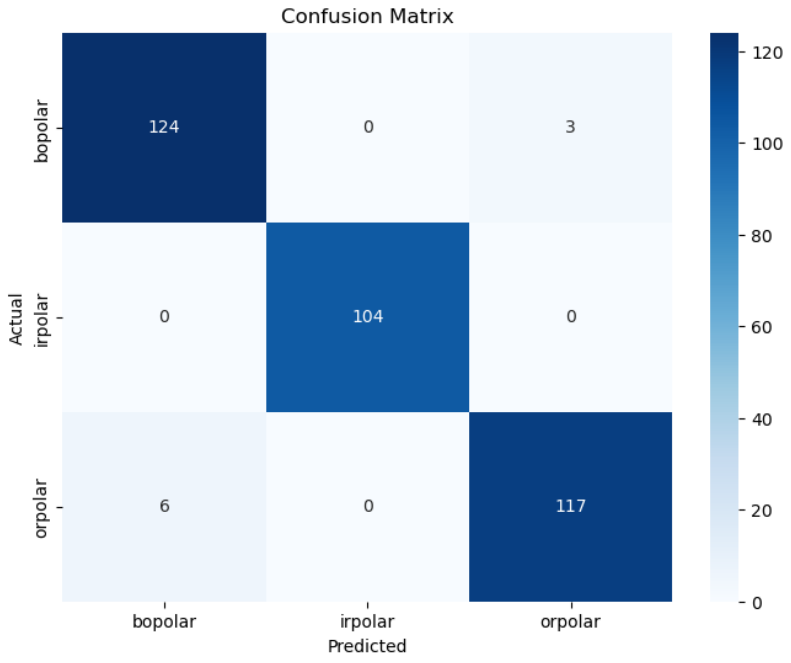


Figure 29: Dual channel AlexNet using spectrograms on the CWRU dataset.

- **ROC-auc curve**: Receiver Operating Characteristics is the probability curve and Area Under The

Curve demonstrated the capability of the model to distinguish among the separate classes by plotting true positive rates on the y-axis and false positive rates on the x-axis. A result in the vicinity of 1 for AUC demonstrates a great classification result and as the result gets closer to 0, it demonstrates a poor classification insight(Narkhede, 2018).

4.5 Computing resources

4.5.1 Hardware Configuration

The efficiency and performance of the generated models are in a tight relation with the underlying infrastructure. We have trained our models using a high-performance processor of Intel(R) Core(TM) i7-10750H, which is equipped with a base clock frequency of 2.60GHz. This system was also empowered with NVIDIA GeForce GTX 1650 Ti Discrete 4GB GDDR6 DirectX 12 GPU. This particular type of GPU comes with 4GB memory and is supported by DirectX 12 to provide advanced graphical performance. Other hardware configurations include 16GB DDR4-2933 RAM and the storage system of Two drives: M.2 2242 SSD 512GB, and 2.5 HDD 1TB

4.5.2 Software and libraries

For the experiments within this thesis domain, we have used a system with the operating system of Windows 11 Pro. The development took place under Python programming language using version 3.9.13. As implementing deep learning models has been the main focus of this thesis, We have used TensorFlow 2.11.0 deep learning library to create my models. We also have leveraged other libraries including sys 3.9.13, cv2 4.8.0, Numpy 1.26.0, Pandas 1.4.4, Matplotlib 3.5.2, sns 0.11.2, Scikit-Learn 1.0.2, OS. Time, and Collections.

5 Experiments and Results

This section begins by describing the experimental setup employed for the classification of various bearing defects. Subsequently, we discuss the obtained results through a series of experiments conducted on the two datasets. The link to the generated development codes is located in the Appendix section of this thesis.

5.1 Experiments

In order to perform several experiments, we employed the following experimental setup and outlined the hyperparameter tuning procedure.

5.1.1 Experimental setup

The experimental setup for the NASA dataset has been done as follows. NASA dataset contains the raw vibration signals of the bearings. These signals encompass the information on the defects that occurred on the Inner race, Outer race, and rolling element part of the bearings. After the demodulation process, we turned the raw vibration signals of the defects into their corresponding 2D visualizations including Frequency Spectrums, Polar Specrums, and Spectrograms.

The experimental setup for the CWRU dataset has been done as follows. The CWRU dataset contains the raw vibration signals of the bearings in the Matlab format. These signals encompass the information on the defects that occurred on the Inner race, Outer race, and rolling element part of the bearings. After splitting the vibrational signals into equal bunches, we turned the fragmented raw vibrational signals of the defects into their corresponding 2D visualizations including Frequency Spectrums, Polar Specrums, and Spectrograms.

In the next step, we split the dataset into training and testing datasets using an 80:20 ratio, where 80% of the dataset was assigned for the training and 20% of the dataset was reserved to test. Add to the point, as we are dealing with an imbalanced set of defects we have calculated the weights for each class to train our model based on the specific class weights. We then evaluated their classification performance using LeNet, NASNet, AlexNet, EfficientNet, and VGG16. Using mid-level fusion methods, we implemented dual feature channels, where one channel was implemented using 2D

representations and 2D CNN-based methods, and the other channel was implemented using statistical features and 1D CNN. The final classification was then performed based on the two channels.

5.1.2 Hyperparameter tuning

In order to find the most optimal model structure, we have followed the same hyperparameter tuning structure on both datasets. We have evaluated the performance of the model based on different epochs of 20 and 30 and different batch sizes of 32 and 64. The model was then trained under different combinations of these epochs and batch sizes and for each combination, the evaluation was conducted to provide the test accuracy and F1-score. The best hyperparameter was then identified based on the combination that provided the highest accuracy. We have also used early stopping with patience 3 during the training phase to avoid unnecessary runs after no improvements within 3 epochs.

5.2 Results and discussions

5.2.1 Experimental Results on the NASA dataset

The obtained classification results using various 2D CNN models on the NASA dataset are presented in the tables from Table 4 to Table 9.

- **Single Channel Models**

  In this section, we present the results obtained based on several feature extraction methods using single-channel 2D CNN models on the NASA dataset.

  - **Polar Spectrums**

    The classification results using the polar spectrum with a single channel model are presented in Table 4. From Table 4, it can be observed that out of all the developed models, VGG16 obtained a higher classification accuracy of 97.2% and with an F1-score of 0.96.Whereas, the second best-obtained result is using AlexNet with an accuracy of 96.16% and with an F1-score of 0.951.

As illustrated in the ROC curves within Figure 30, we can observe that VGG16 and ResNet demonstrate a greater performance within the model to provide true predicted classes.

In order to understand the more detailed view of the few experiments, confusion matrices are obtained. The confusion matrices are shown in Figure 31 and it can be observed from the confusion matrices 31 that AlexNet provides 116 true predicted labels out of 267 and LeNet, 274 out of 287. From this, it can be observed that there is a challenge for classification between inner race defects and rolling element defects. Adding to the point, 111 rolling element defects have been misclassified as inner race defects in VGG16. We also identified 182 misclassifications out of 287 using EfficientNet. ResNet on the other hand, demonstrated a significant classification performance with minimum false predicted labels of 8 out of 112. As we are trying to minimize equipment downtime and increase their lifespan, we need our model to be able to provide more accurate assessments.

Model	Accuracy	Class Precision	Class Recall	F1-Score
AlexNet	0.9616	0.03 1. 1.	1. 0.989 0.1	0.951
EfficientNet	0.45	0.03 1. 0.88	0. 1. 0.	0.525
LeNet	0.96	0. 0.92 0.96	0. 1. 0.96	0.95
NASNet	0.82	0. 0.82 0.	0. 1. 0.	0.74
ResNet	0.928	0. 1. 0.6	0. 1. 1.	0.90
VGG16	0.972	0.05 1.0 1.0	1.0 1.0 0.4	0.96

Table 4: Single Channel using Polar Spectrum.

- **Spectograms**

In this, we present the classification results when using spectrograms along with various 2D-CNN models. The obtained classification results are presented in Table 5. From the table, it can be observed that the EfficientNet model got the classification result of 90% accuracy with an F1-score of 0.89, which was significantly higher than other models within the table. The second best-achieved result is using VGG16 with an accuracy of 76% and with an F1-score of 0.72.

The roc curve is shown in Figure 33 where the graph shows that EfficientNet and VGG16 demonstrate high true positives and low false positives.

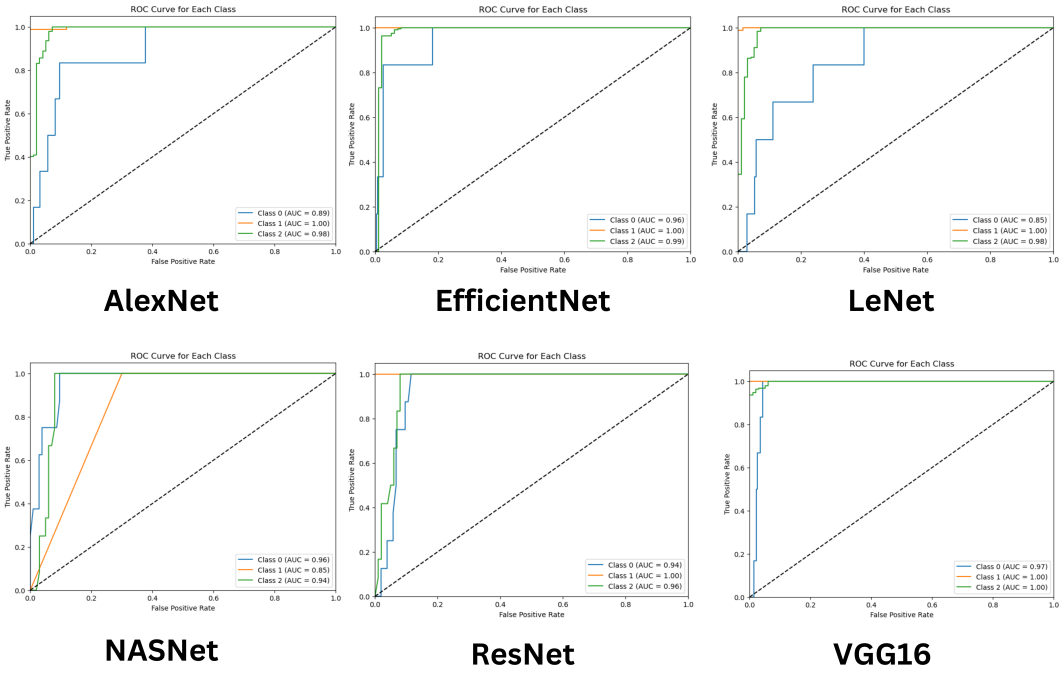


Figure 30: Single channel roc curve using polar spectrum.

.

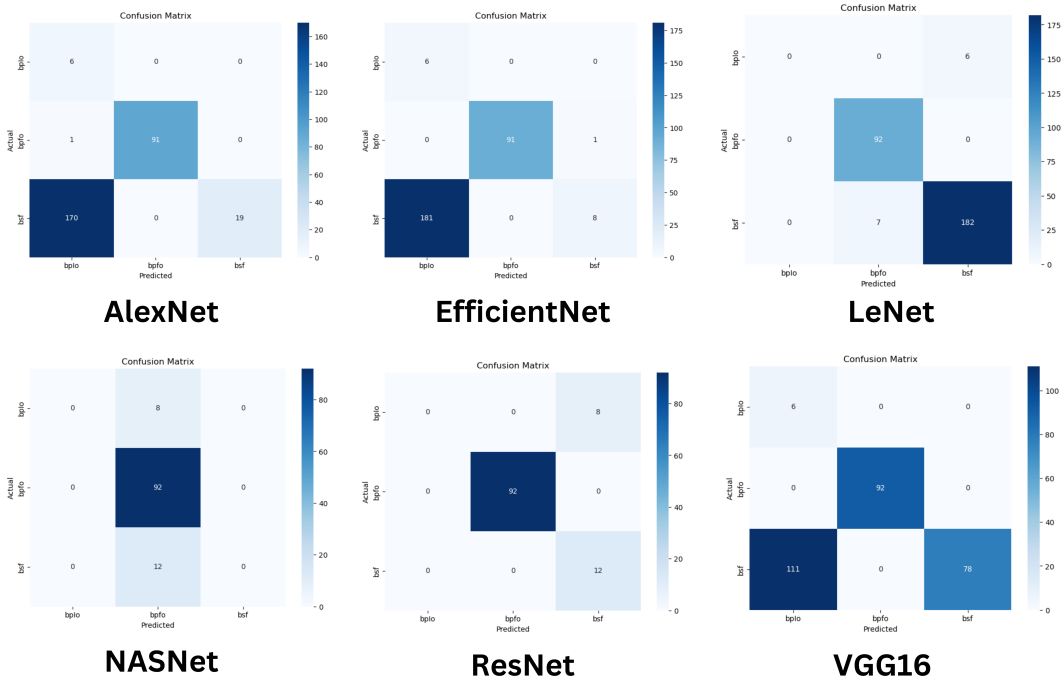


Figure 31: Single channel confusion matrix using polar spectrum.

.

In order to understand the more detailed view of this experiment, confusion matrices are obtained. The confusion matrices are presented in Figure 32 and it can be seen that EfficientNet classifies 71 out of 287 correctly and can efficiently categorize the inner race fault types. AlexNet, and NASNet on the other hand, depict a great deal of misclassification.

Whereas VGG16 demonstrates better performance in the confusion matrix by classifying 110 true labels out of 287.

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.66	0. 0. 0.65	0. 0. 1.	0.53
EfficintNet	0.90	0.027 1. 0. 8	1. 0.7 0.	0.89
LeNet	0.32	0.02 0.96 0.	1. 0.88 0.	0.26
NASNet	0.65	0.02 0. 0.56	1. 0. 0.095	0.52
ResNet	0.66	0. 0.32 0.75	0. 0.98 0.015	0.53
VGG16	0.76	0. 0.34 0.94	0. 1. 0.095	0.72

Table 5: Single Channel using Spectrograms.

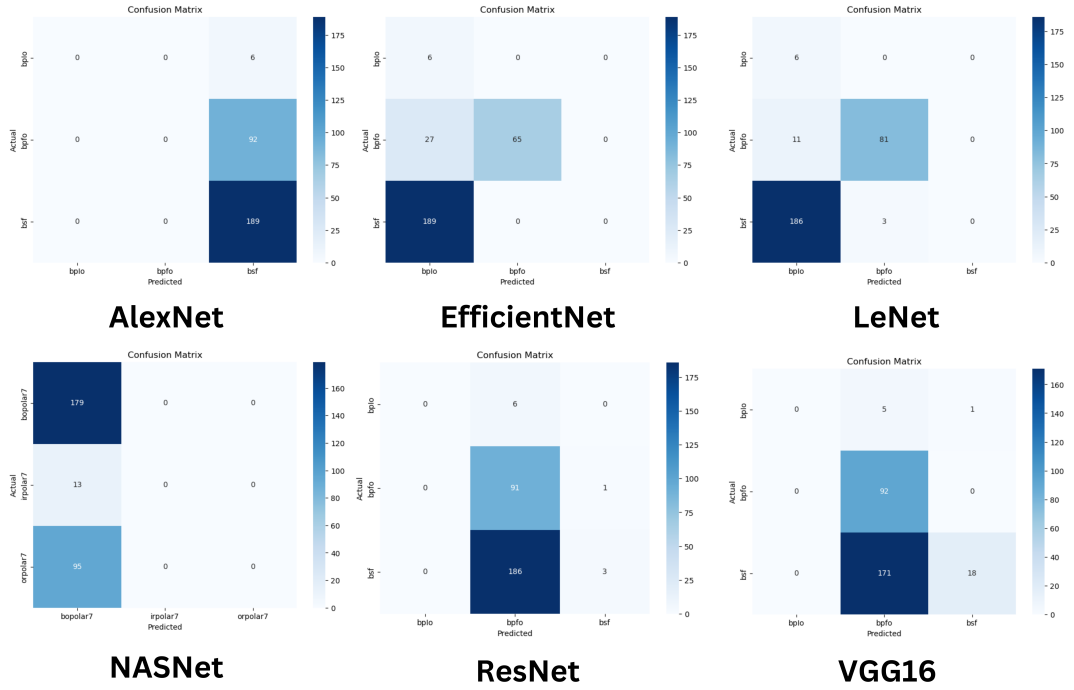


Figure 32: Single channel confusion matrix using Spectogram.

.

- **Frequency Spectrums**

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.65	0. 0.32 0.	0. 1. 0.	0.52
EfficintNet	0.9	0.02 1. 0.	1. 0.70 0.	0.89
LeNet	0.32	0.02 0.96 0.	1. 0.88 0.	0.29
NASNet	0.65	0.02 0. 0.56	1. 0. 0.09	0.52
ResNet	0.66	0. 0.32 0.75	0. 0.98 0.015	0.53
VGG16	0.49	0.07 0.47 0.84	0.66 0.89 0.29	0.48

Table 6: Single Channel using Frequency Spectrums.

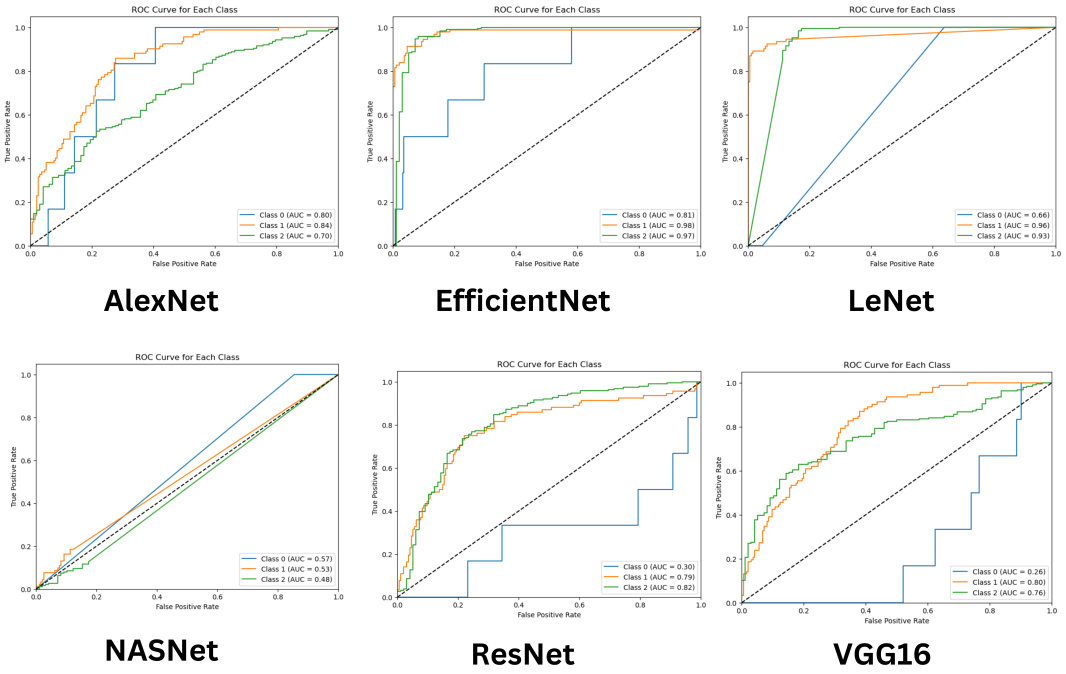


Figure 33: Single channel roc curve using spectrogram.

.

The classification results using the frequency spectrum with a single channel model are presented in Table 6. From the table, it can be seen that EfficientNet provides the highest accuracy of 90% for the classification of bearing faults using frequency spectrums with an F1-score of 0.89. Whereas, the ResNet model achieved the second-best classification accuracy of 65% accuracy and an F1-score of 0.53. From this, it can be observed that there is a huge gap between the accuracy of the EfficienNet and other models when we evaluate using frequency spectrums. This gap may signify the advanced architecture of the EfficientNet in network scaling to increase the accuracy.

The AUC curve is plotted for this model and shown in Figure 35. From the graph, as we can see, EfficientNet achieved a significant classification capability with AUC values of 0.81,0.89, and 0.97. This means, EfficientNet achieved higher performance.

In order to understand the more detailed view of this experiment, confusion matrices are obtained. The confusion matrices are presented in Figure 34. From the confusion matrices, we can observe that there are 71 true predictions out of 287 for EfficientNet, 87 True predictions out of 287, and 147 out of 287 for VGG16. We can also perceive that the various models like AlexNet, NASNet, ResNet, and even EfficientNet all had problems

with the classification of rolling element defect types. This fact makes them unreliable in performing accurate predictive maintenance in real-world scenarios.

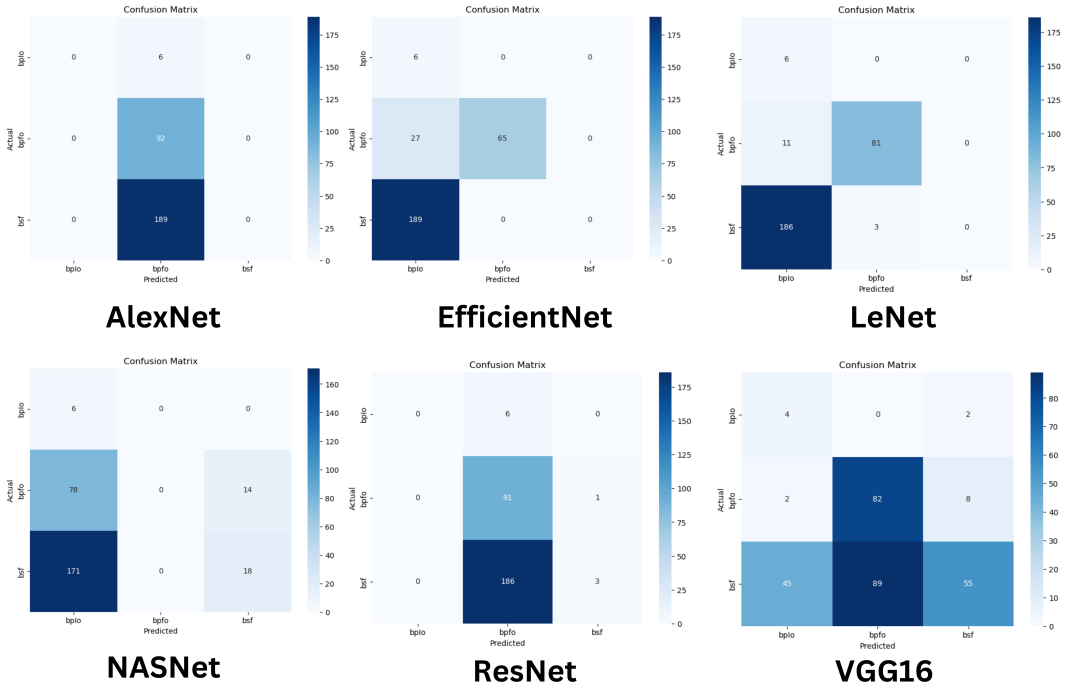


Figure 34: Single channel confusion matrix using Frequency Spectrum.

.

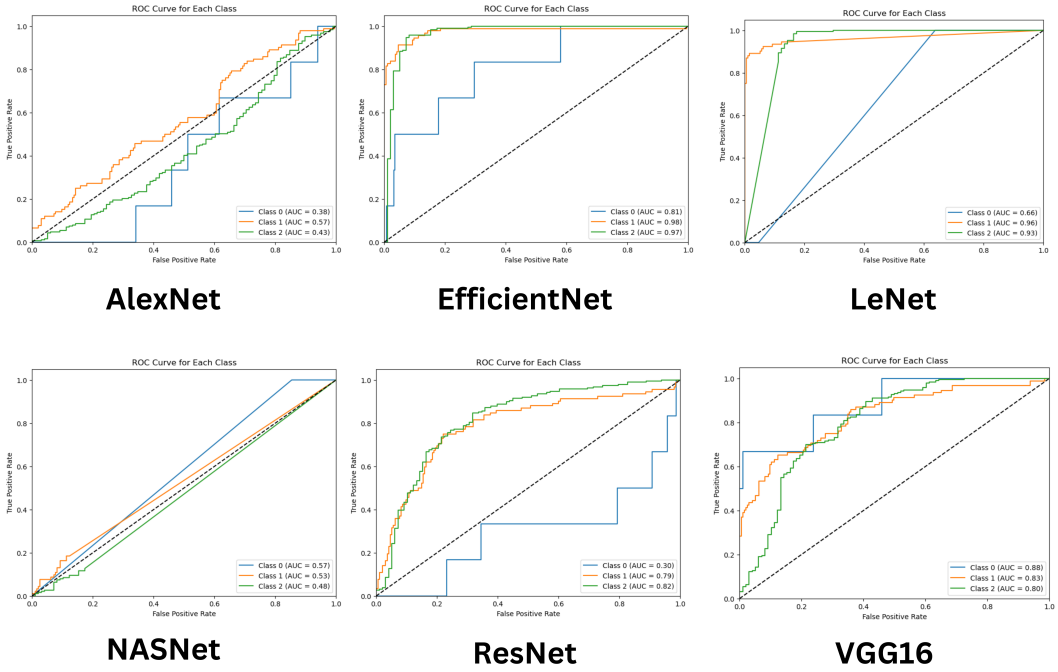


Figure 35: Single channel roc curve using Frequency Spectrum.

.

- **Dual Channel Models** In this, we present the results obtained based on several feature extraction methods using dual channel 2D CNN models on the NASA dataset.

  - **Polar Spectrums**

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.99	0.99 0.91 0.98	1. 0.84 0.98	0.99
EfficintNet	0.99	0.99 1. 1.	1. 0.92 1.	0.99
LeNet	0.993	0.99 1. 0.98	0.99 1. 0.98	0.992
NASNet	0.331	0. 0. 0.33	0. 0. 1.	0.164
ResNet	0.996	0.99 1. 1.	1. 0.92 1.	0.996
VGG16	0.989	0.98 0.9 .98	0.99 0.76 1.	0.982

Table 7: Dual Channels using Polar Spectrum and statistical features.

In this, we present the classification results when using polar spectrum and statistical features along with various dual-channel 2D-CNN models. The obtained classification results are presented in the Table 7. From the table, it can be observed that the ResNet model achieved the highest classification performance with an accuracy of 99.6% and with an F1-score of 0.99. All the other models except NASNet have also demonstrated high classification probabilities of over 98% and an F1-score of over 0.98.

The ROC curves are obtained for all the experiments using the polar spectrum in Figure 37. The graph demonstrates that except NASNet, all the other models have resulted in high AUC values above 0.9. EfficientNet and ResNet demonstrate a significant AUC value for all classes which is representative of the correct classification performance of the models.

In order to understand the more detailed view of this experiment, confusion matrices are obtained. The confusion matrices (CM) are presented in Figure 36. From these CMs, we can observe that, all the models i.e., ResNet, AlexNet, EfficientNet, VGG16, and LeNet demonstrate robust classification performance. EfficientNet, and ResNet both contained only one misclassified defect, and LeNet, AlexNet, and VGG16 2,3, and 4 misclassified defects respectively.

  - **Spectograms** In this, we present the classification results when using spectrograms and statistical features along with various dual-channel 2D-CNN models on the NASA dataset.

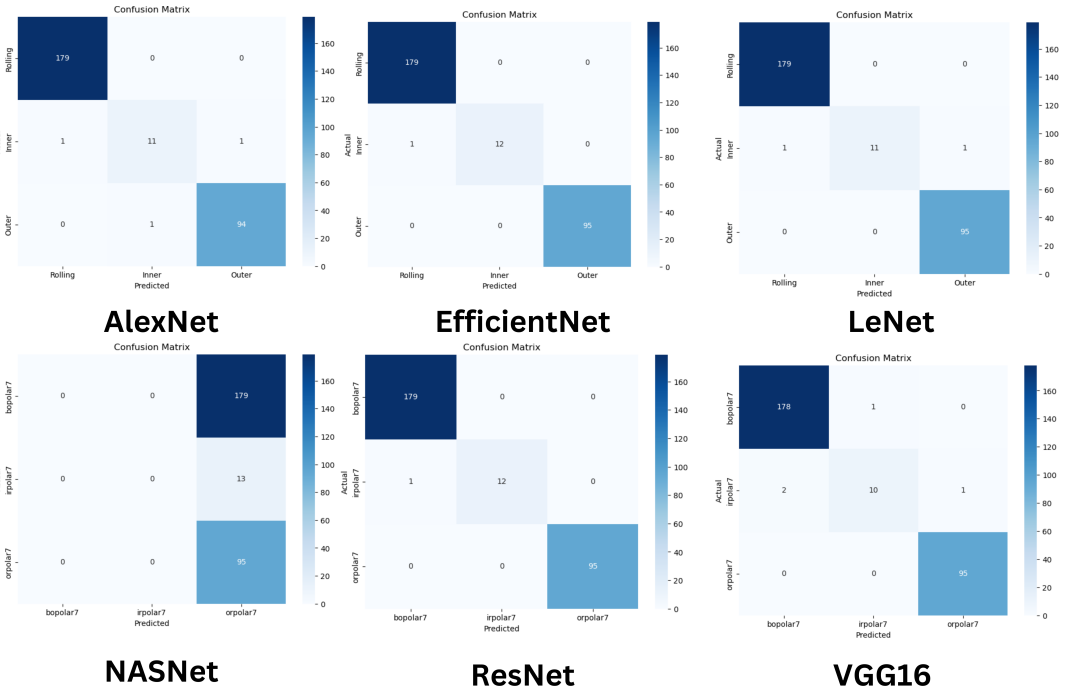


Figure 36: Dual channel confusion matrix using polar spectrum

.

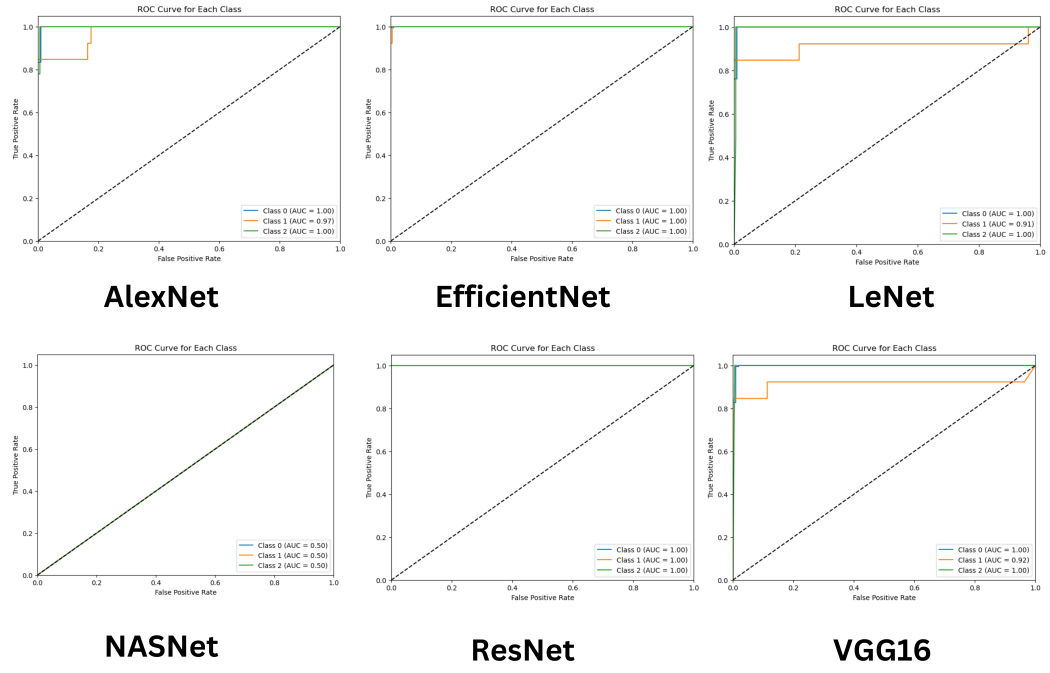


Figure 37: Dual channel roc curve using polar spectrum

.

With regard to the classification results of the fusion model, the obtained results are tabulated in Table 8. From the table, it can be seen that EfficientNet demonstrates a significant accuracy of 100% with the F1-score of 1. Whereas, models AlexNet, LeNet, ResNet, and

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.993	0.99 1. 0.98	1. 0.84 1.	0.992
EfficintNet	1	1. 1. 1.	1. 1. 1.	1
LeNet	0.993	0.99 1. 0.98	1. 0.84 1.	0.992
NASNet	0.641	0.62 0. 0.	1. 0. 0.	0.628
ResNet	0.996	1. 1. 0.98	1. 0.92 1.	0.996
VGG16	0.986	0.99 0.68 0.98	0.97 0.84 1.	0.984

Table 8: Dual Channels using Spectograms and statistical features.

VGG16 also demonstrate high classification results above 98% and with an F1-score of above 0.98.

When the AUC curves are plotted, the graphs obtain the values of 1 for different classes within the ROC curve. Within the graph 39, EfficientNet demonstrates the efficiency of the model and its ability to make true predictions. All the other models except NASNet provide the AUC value of 1 for the outer race and rolling element defect classes and above 0.91 for the inner race defect class.

In order to understand the more detailed view of this experiment, confusion matrices are obtained. The confusion matrices (CM) are presented in Figure 38. From CMs, we can observe that the EfficientNet model was able to perform the classification with zero misclassified instances. Following EfficientNet, ResNet with one misclassified instance, and LeNet and AlexNet both with two misclassified instances have been the highest efficient models.

- **Frequency spectrums** In this, we present the classification results when using frequency spectrums and statistical features along with various dual-channel 2D-CNN models on the NASA dataset.

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.993	0.99 1. 0.98	1. 0.84 1.	0.992
EfficintNet	1	1. 0.92 0.98	01. 0.92 0.98	1
LeNet	0.993	0.98 0.83 0.98	0.99 0.76 0.98	0.992
NASNet	0.331	0. 0. 0.33	0. 0. 1.	0.16
ResNet	0.996	1. 1. 0.98	1. 0.92 1.	0.996
VGG16	0.993	0.99 1. 0.98	1 0.84 1.	0.992

Table 9: Dual Channels using frequency spectrums and statistical features.

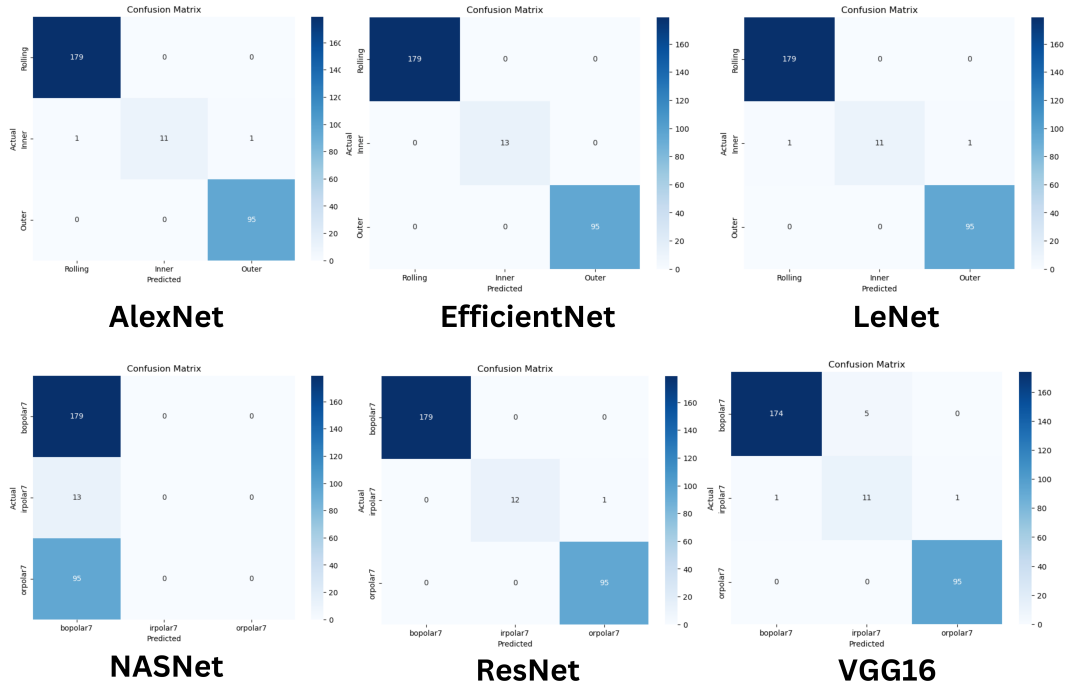


Figure 38: Dual channel confusion matrix using Spectograms.

.

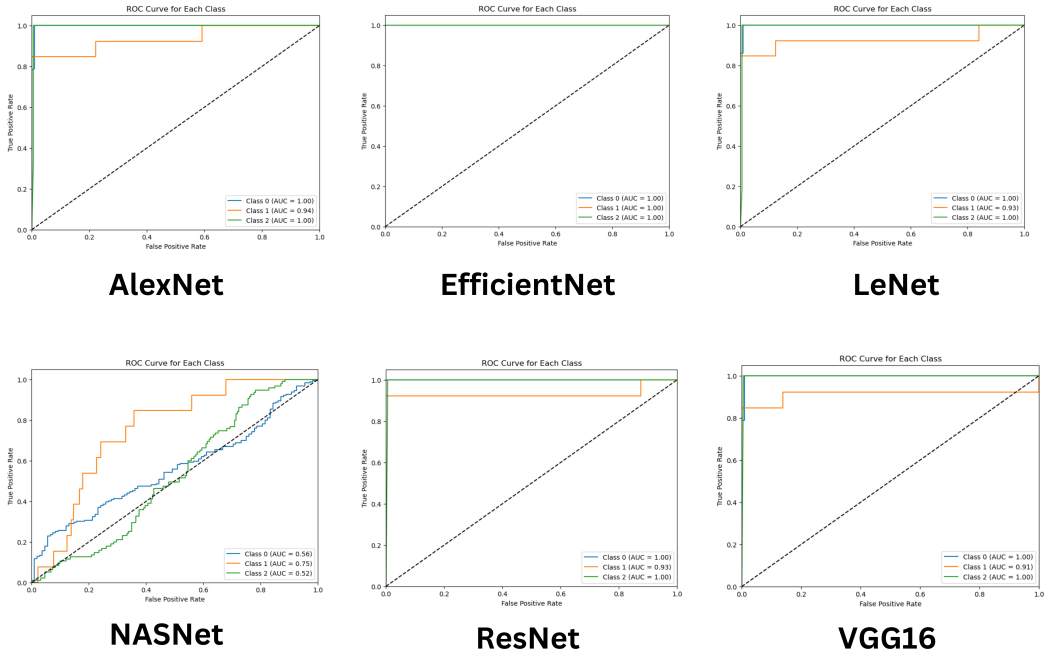


Figure 39: Dual channel roc curve using Spectograms.

.

With regard to the classification results, the obtained results of the fusion model are tabulated in Table 9. From the table, it can be seen that the EfficientNet model achieved the highest accuracy of 100% in the classification of varied bearing faults. Whereas, ResNet, AlexNet, LeNet, And VGG16 models also obtained the accuracy of the classification above

99% accuracy.

The ROC curves are obtained for all the dual Channels using frequency spectrums and statistical features in Figure 41. From the graphs, it can be seen that Alexnet demonstrated great performance in correctly classifying the defects by providing AUC values of 1 for all classes. Also, AlexNet, ResNet, VGG16, and LeNet, all provide AUC values over 0.9.

In order to understand the more detailed view of this experiment, confusion matrices are obtained. The confusion matrices (CM) are presented in Figure 40. From the CMs, we can observe that there is only one misclassified defect in ResNetet, and 2 for EfficientNet, alexNet, and VGG16, and 5 for LeNet.

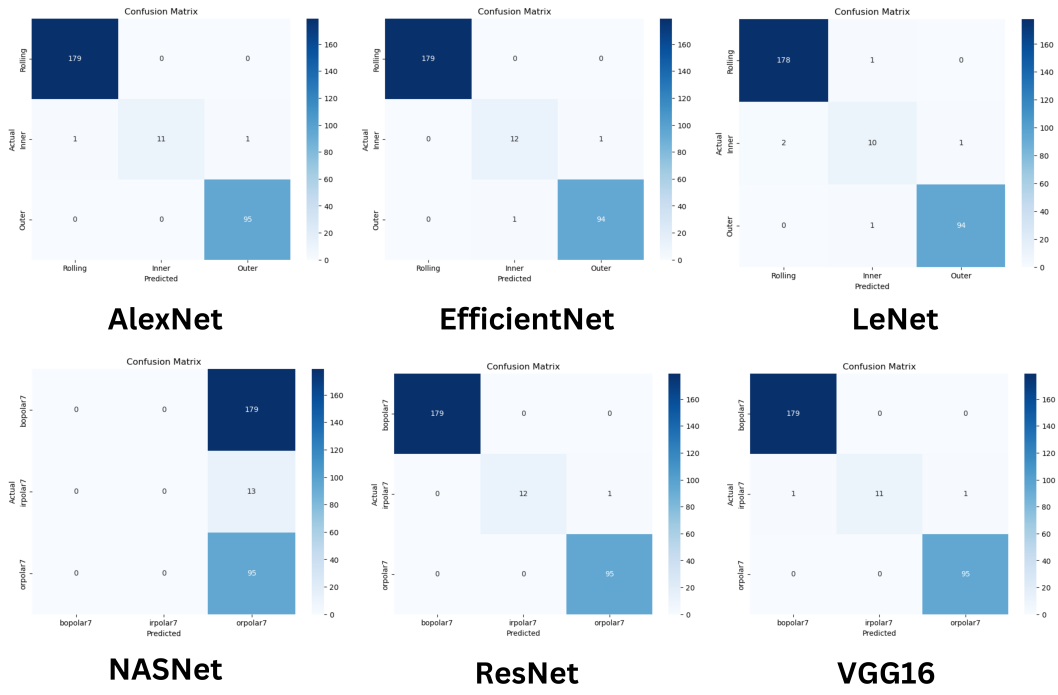


Figure 40: Dual channel confusion matrix using Frequency spectrum.

.

5.2.2 Experimental Results on CWRU dataset

The obtained classification results using various 2D CNN models on the CWRU dataset are presented in the tables from Table10 to Table 15.

- **Single Channel Models** In this, we present the results obtained based on several feature ex-

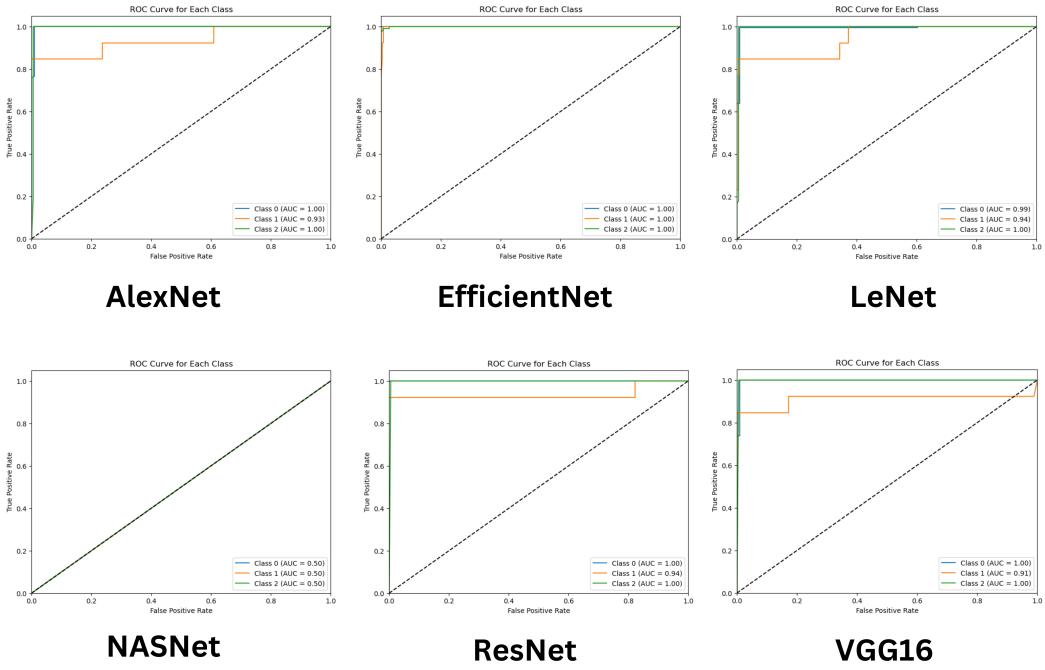


Figure 41: Dual channel roc curve using Frequency spectrum.

.

traction methods using single-channel 2D CNN models on the CWRU dataset.

- **Polar Spectrums**

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.55	0.64 0.36 0.66	0.42 0.63 0.54	0.553
EfficintNet	0.724	0.70 0.62 0.82	0.77 0.80 0.56	0.722
LeNet	0.635	0.58 0.77 0.65	0.81 0.36 0.68	0.621
NASNet	0.641	0.62 0.63 0.66	0.72 0.47 0.07	0.6636
ResNet	0.682	0.58 0.66 0.79	0.79 0.5 0.67	0.67
VGG16	0.677	0.57 0.63 0.87	0.0.83 0.625 0.51	0.674

Table 10: Single Channel using Polar Spectrum.

Several 2D CNN models have been trained on the CWRU dataset. The obtained classification results on the single channel using the polar spectrum are presented in Table 10. In the table, we can observe that the EfficientNet model resulted in the highest classification accuracy of 72.4% with an F1-score of 0.722. AlexNet also demonstrated the second-best performance model by obtaining the classification accuracy of 55% .

The ROC curves are obtained for all the models for this experiment and the graphs are depicted in Figure 43. In the graphs, EfficientNet has demonstrated a better performance at truly predicting the defect types with the AUC values ranging from 0.88 to 0.911.

In order to understand the more detailed view of this experiment, confusion matrices are obtained. The confusion matrices (CM) are presented in Figure 42. Within the matrices, we can observe sparse misclassification among all the models in this experiment.

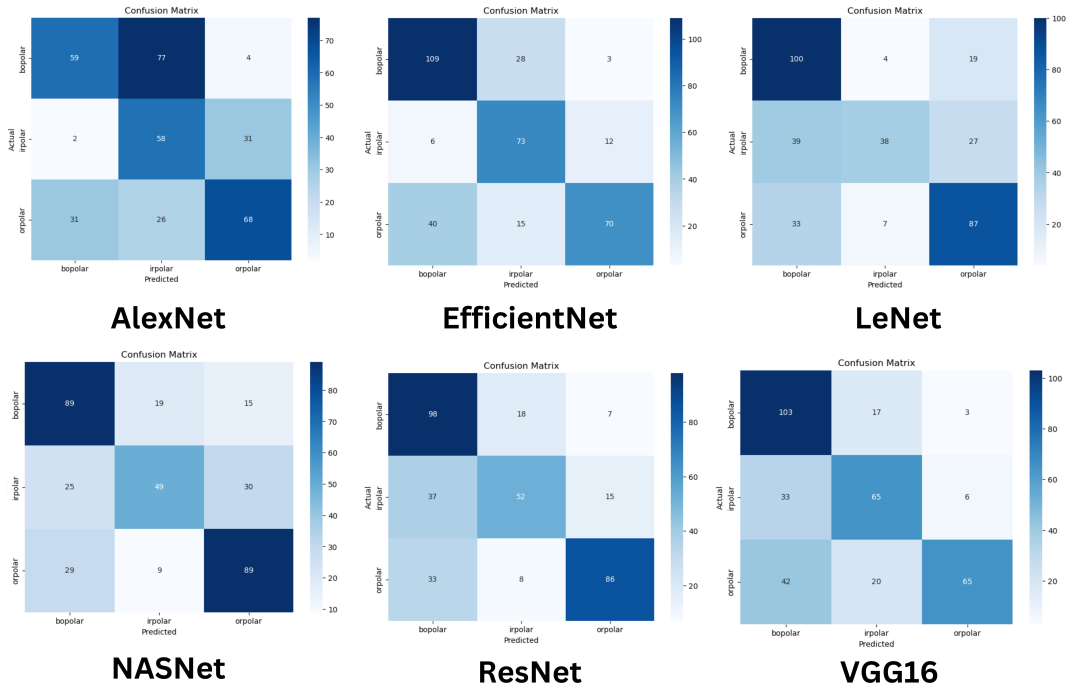


Figure 42: Single channel confusion matrix using polar spectrum.

.

– **Spectograms**

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.968	0. 0.29 0.	0. 1. 0.	0.968
EfficintNet	0.838	0.69 0.96 0.97	1. 0.90 0.62	0.831
LeNet	0.889	0.78 0.98 0.957	0.951 0.826 0.881	0.891
NASNet	0.711	0.92 1. 0.47	0.18 0.57 1.	0.716
ResNet	0.683	0.55 0.80 0.97	0.97 0.48 0.56	0.678
VGG16	0.723	0.724 0.64 0.8	0.74 0.99 0.42	0.722

Table 11: Single Channel using Spectrograms.

With regard to the classification results within Table 11, AlexNet results in the highest classification accuracy in comparison with another model with an accuracy of 96.8% and F1-score of 0.968. EfficientNet and LeNet also demonstrated good results of 83.8% accuracy and 0.831 F1-score for EfficientNet, and 88.9% accuracy and 0.891 F1 score for the LeNet

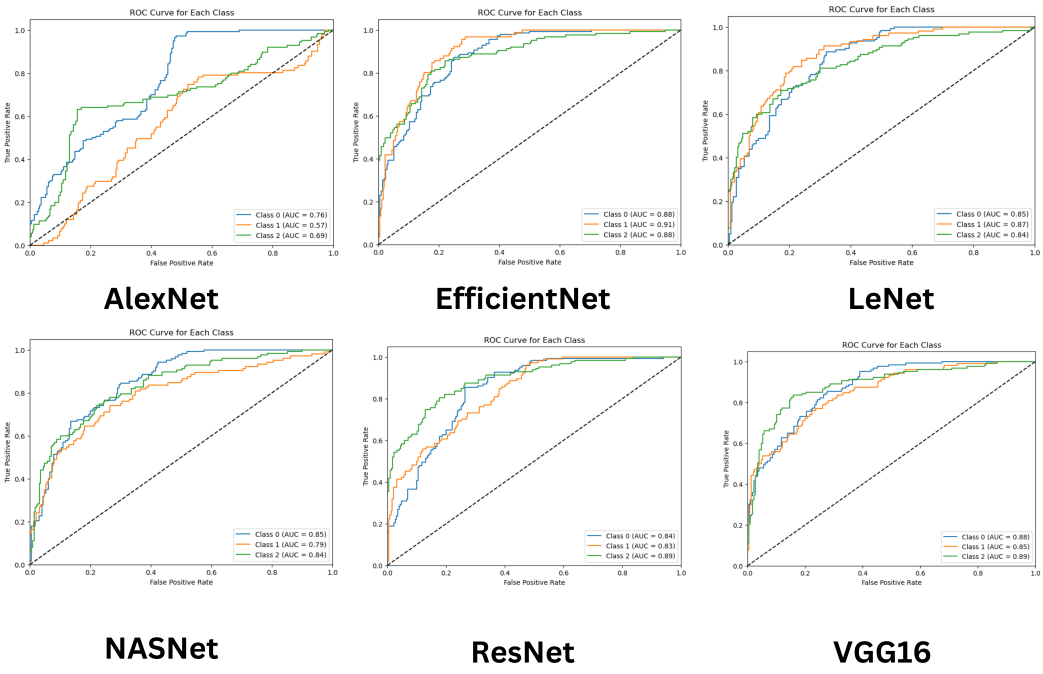


Figure 43: Single channel roc curve using polar spectrum.

.

We can also observe a great classification performance by evaluating the high AUC graphs within the EfficientNet in Figure 44. LeNet and ResNet also provided AUC values over 0.92. All other models have resulted in AUC values over 0.83 and demonstrated an excellent performance to provide true predictions

As we can observe in the confusion matrices in Figure 44, provide a decent classification by providing 297 correct classifications out of 354 defect types. AlexNet, LeNet, resNet, and VGG16 also result in 93,39, 112, and 105 misclassification instances respectively

- **Frequency Spectrums**

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.55	0.64 0.36 0.66	0.42 0.63 0.54 .	0.553
EfficintNet	0.641	0.49 0.90 0.55	0.50 0.27 0.85	0.643
LeNet	0.853	0.83 0.97 0.77	0.71 0.97 0.88	0.851
NASNet	0.358	0.26 0.1 0.16	0.47 0.076 0.070	0.216
ResNet	0.539	0.5 0.65 0.47	0.55 0.68 0.40	0.536
VGG16	0.79	0.80 0.81 0.69	0.40 0.91 0.96	0.78

Table 12: Single Channel using Frequency Spectrums.

With regard to the classification results within Table 12, we can observe that LeNet pro-

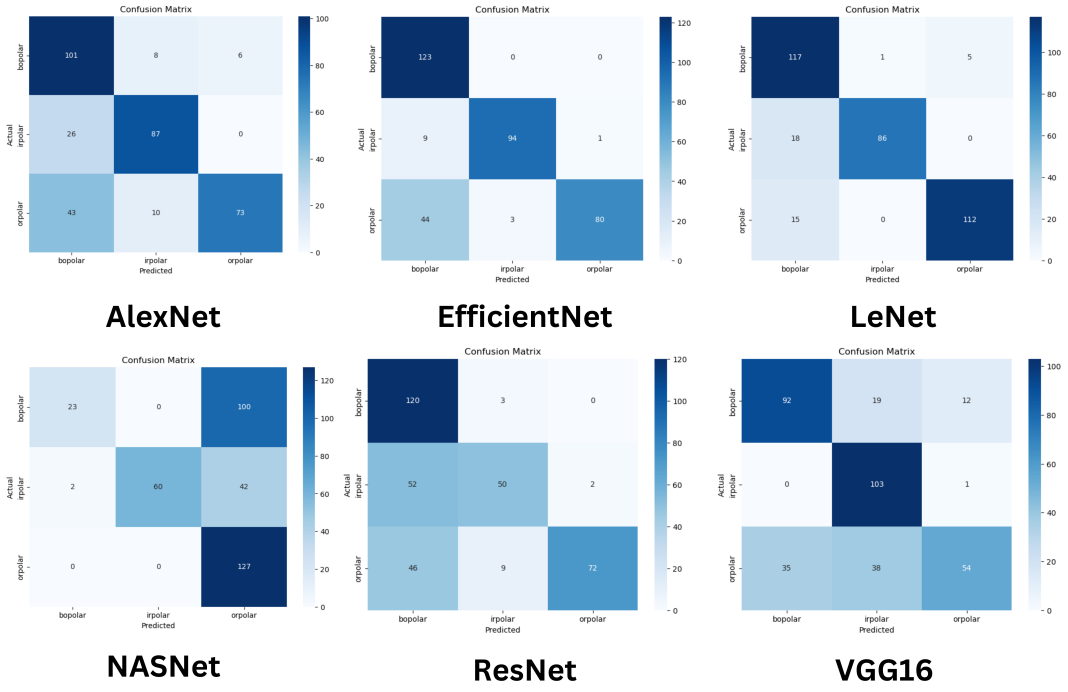


Figure 44: Single channel confusion matrix using spectogram

.

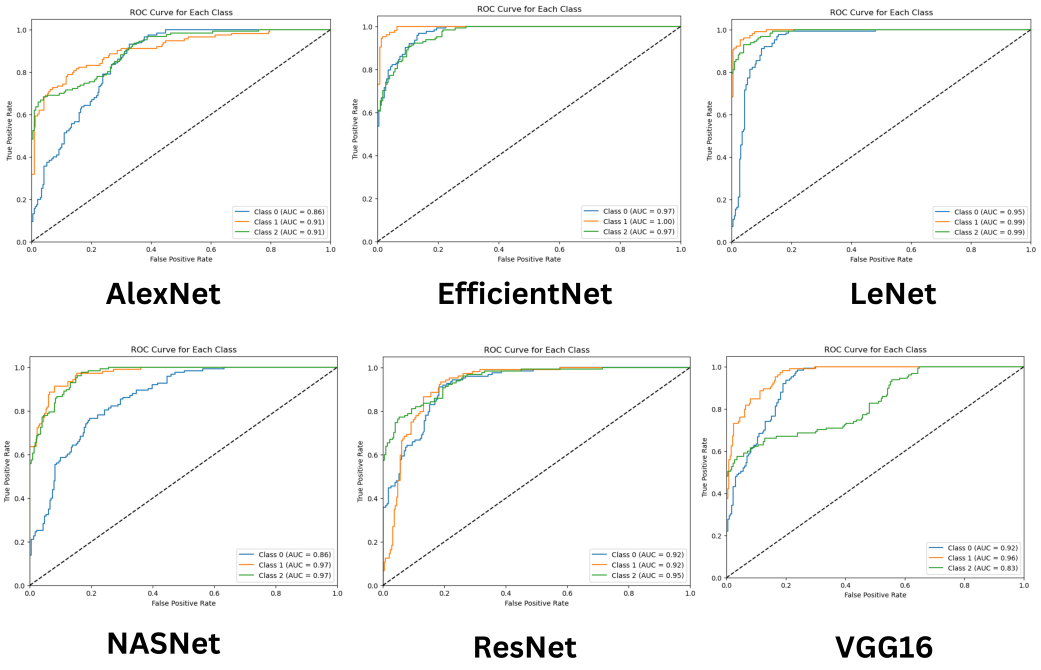


Figure 45: Single channel roc curve using spectogram

.

vided the highest classification accuracy of 85.3%, and F1- score of 0.851 in comparison with other models in this experiment. VGG16 also provided a good accuracy of 79% which is well higher than other models within this experiment.

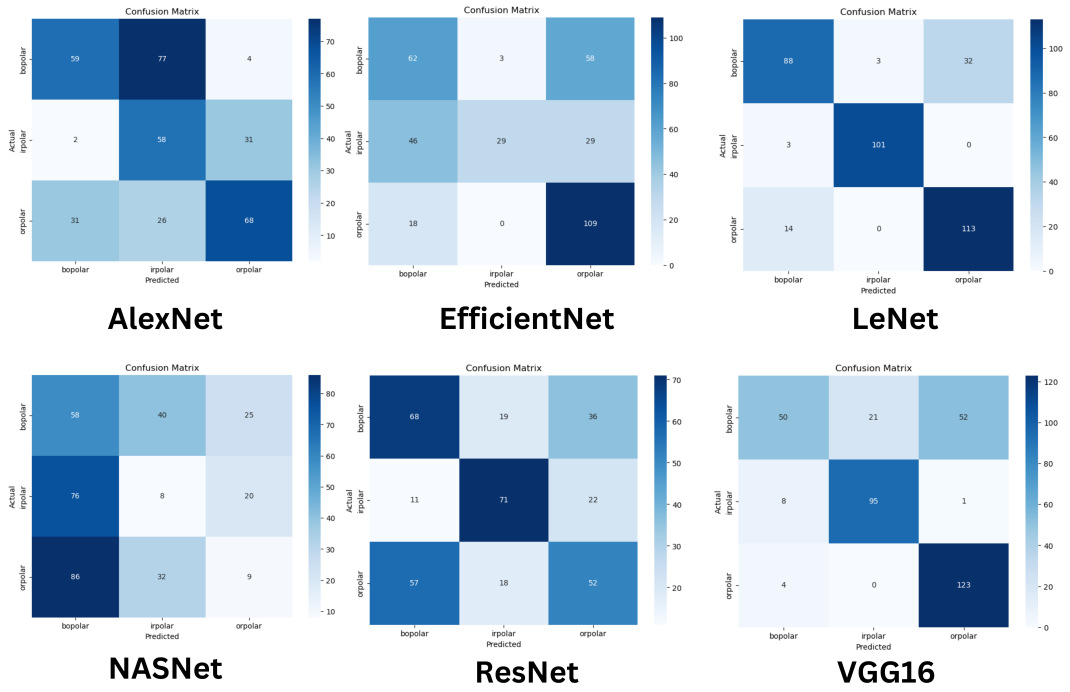


Figure 46: Single channel confusion matrix using frequency spectrum

.

Add to the point, the AUC values within the ROC curves in Figure 47 represent that LeNet has the highest capability of true prediction in comparison with other models within the experiment. We could also observe that VGG16 and EfficientNet both provide promising capability for true predictions as well.

the confusion matrices in Figure 46, demonstrate the classification performance of the models. As these matrices represent, AleXnet, EfficientNet, LeNet,resNet, and VGG16, result in 171, 150, 52, 163, and 86 misclassified instances out of 354 respectively

- **Dual Channel Models**

  - **Polar Spectrums**

    With regard to the classification results within Table 13, we can observe that AlexNet and EfficientNet provide the highest accuracy of 98.3% and F1-score of 0.983. ResNet, VGG16, and LeNet provided high accuracy of, 97.7%, 97.4%, and 95.7% respectively.

    Add to the point, the AUC values within the ROC curves in Figure 49, prove the effectiveness of this model in providing true classification. EfficientNet and ResNet, both demon-

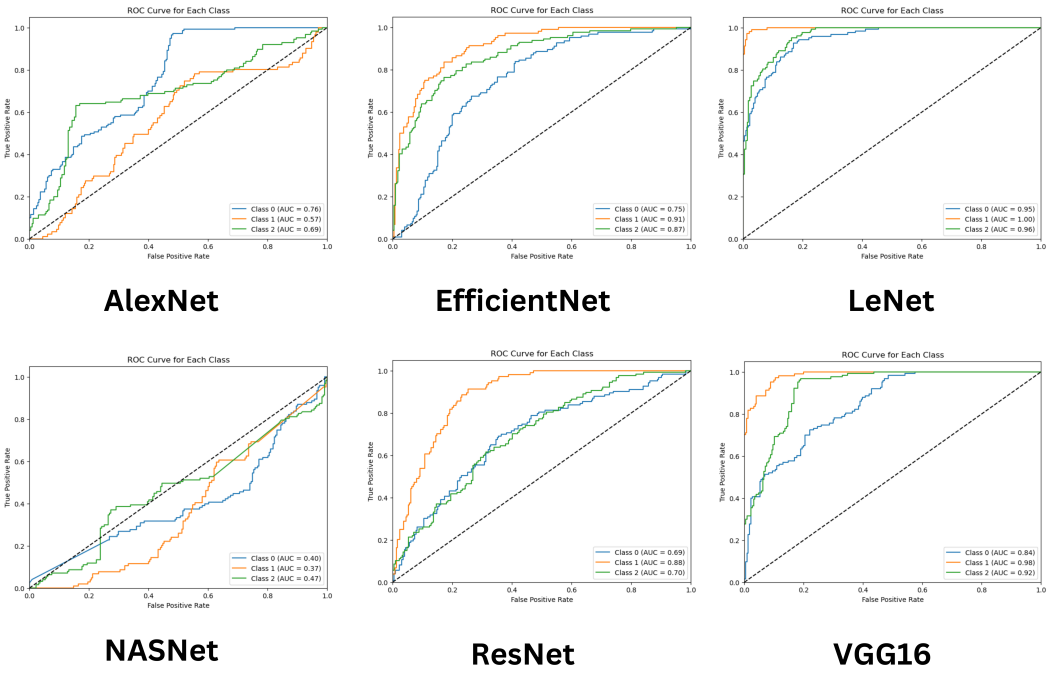


Figure 47: Single channel roc curve using frequency spectrum

.

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.983	0.98 1. 0.96	0.96 1. 0.98	0.983
EfficintNet	0.983	0.98 0.98 0.97	0.96 1. 0.98	0.983
LeNet	0.957	0.95 0.99 0.93	0.92 1. 0.95	0.957
NASNet	0.596	0.51 0. 0.54	0.86 0. .617	0.579
ResNet	0.977	0.97 1. 0.96	0.96 1. 0.97	0.977
VGG16	0.974	0.98 1. 0.9453125	0.94 1. 0.98	0.974

Table 13: Dual Channels using polar Spectrum and statistical features.

strate significant classification performance by AUC values of 1 for all the fault types.

As we can observe within the confusion matrices in Figure 48, all the models except NASNet, demonstrate a minor misclassification. As we can see, this misclassification for AlexNet, EfficientNet, and LeNet, has been mostly centralized on the mismatch among the defects within rolling element defect and outer race defects fault types.AleXnet, EfficinetNet, LeNet, ResNet, and VGG16,resulted in 6,7,7,8, and 9 misclassified instances respectively

- **Spectograms**

  With regard to the classification results within Table 14, we can observe that EfficientNet, and ResNet, both demonstrate significant accuracy of 100% with the F1-score of

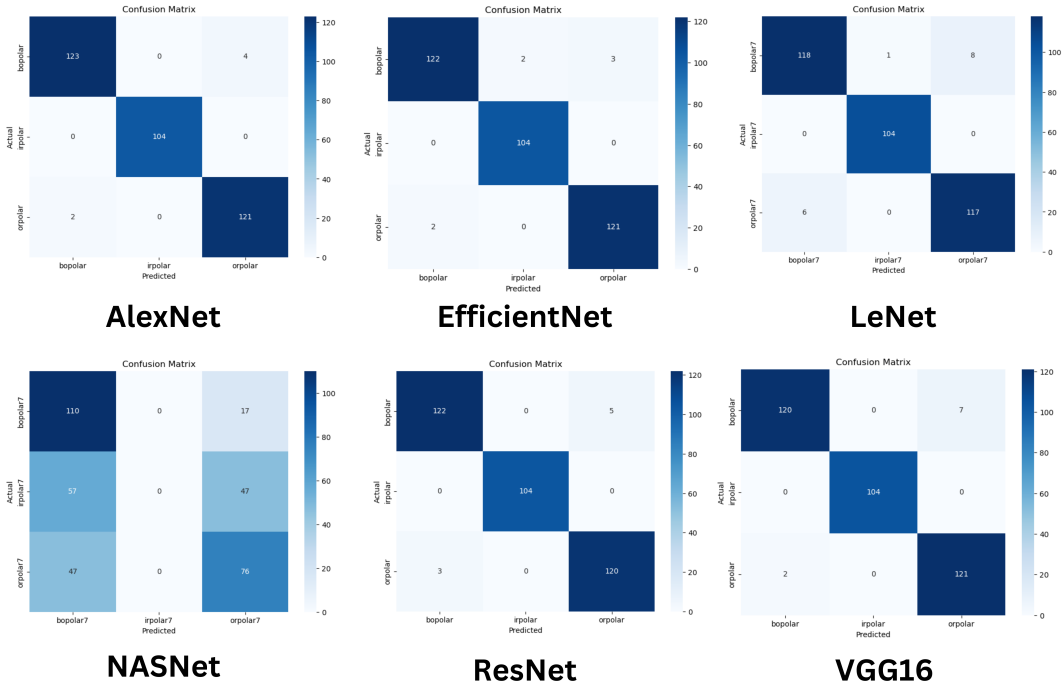


Figure 48: Dual channel channel confusion matrix using polar spectrum

.

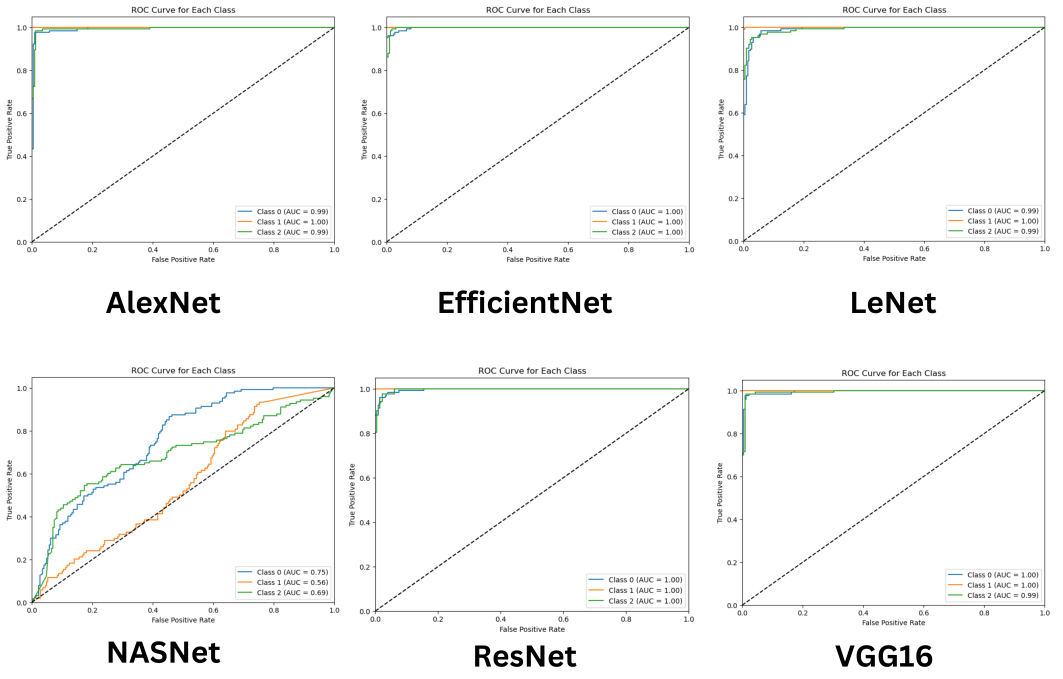


Figure 49: Dual channel channel roc curve using polar spectrum

.

1. AlexNet, LeNet, and VGG16 also both provide a high classification accuracy of 99.1%, 99.7%, and 99.6% with the F1-score of 0.991, 0.997, and 0.996 respectively.

Add to the point, the AUC values within the ROC curves in Figure 51, demonstrate sig-

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.991	0.1 0.99 0.98	0.97 1. 0.1	0.991
EfficintNet	1	1. 1. 1.	1. 1. 1.	1
LeNet	0.997	0.1 1. 0.98	0.98 1. 1.	0.997
NASNet	0.361	0.36 0.2 0.	0.99 0.01 0.	0.20
ResNet		1. 0.97 1.	0.97 1. 1.	0.991
VGG16	0.996	0.98 .99 0.93	0.92 1. 0.98	0.996

Table 14: Dual Channels using Spectograms and statistical features.

nificant AUC values of 1 for AlexNet, EfficientNet, LeNet, and ResNet which proves the efficiency of these models in identifying true labels.

As we can see in the confusion matrices in Figure 50, EfficienetNet performs perfectly in classifying different fault types. AlexNet, LeNet, and ResNet, on the other hand, demonstrate minor misclassified instances of 6, 15, and 8 within fault types respectively.

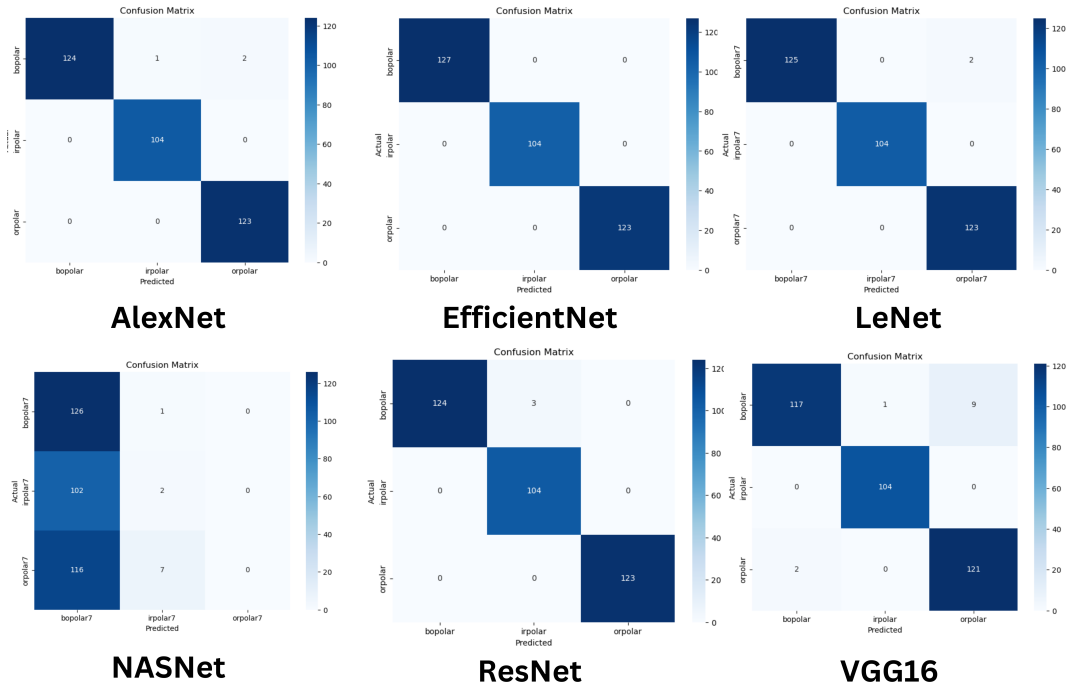


Figure 50: Dual channel confusion matrix using spectograms spectrum

.

– **Frequency spectrums**

With regard to the classification results within Table 14, we can observe that EfficientNet was able to provide significant classification results of 100% accuracy with an F1-score of 1. AlexNet and ResNet also demonstrate great classification accuracy of 97.7%, with an

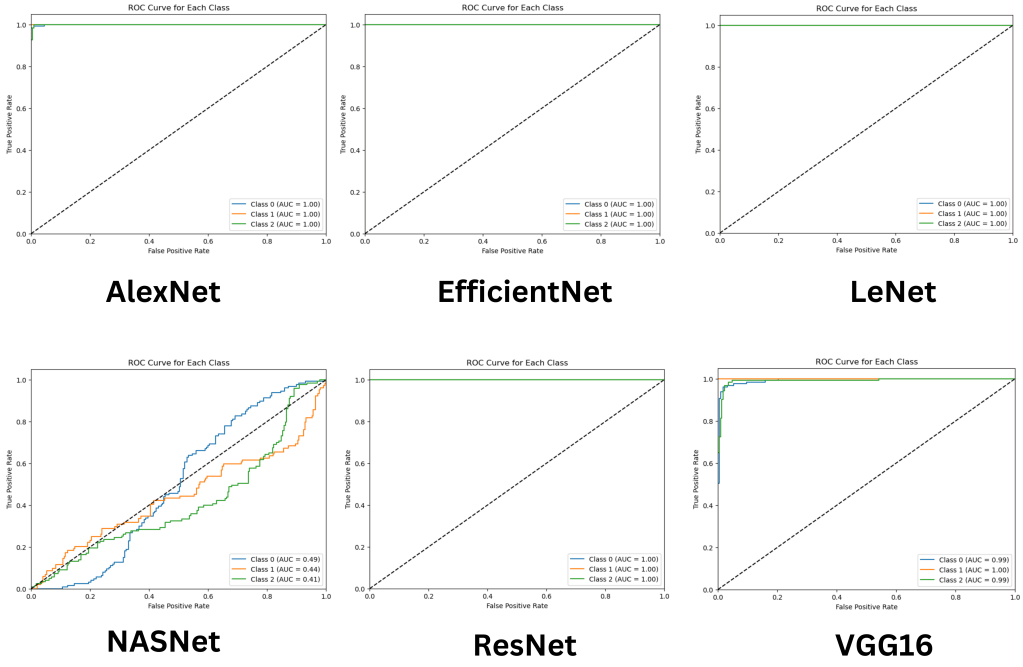


Figure 51: Dual channel roc curve using spectograms

.

Model	Accuracy	Class Precission	Class Recall	F1-Score
AlexNet	0.977	0.98 0.98 0.96	0.95 1. 0.98	0.977
EfficintNet	1	1. 1. 1.	1. 1. 1.	1
LeNet	0.974	0.96 1. 0.95	0.96 1. 0.96	0.974
NASNet	0.358	0. 0. 0.34	0. 0. 1.	0.17
ResNet	0.997	1. 1. 0.99193548	0.99 1. 1.	0.997
VGG16	0.971	0.99 1. 0.93	0.92 1. 0.99	0.971

Table 15: Dual Channels using frequency spectrums and statistical features.

F1-score of 0.977.

As illustrated in the ROC curve in Figure 53, we can obtain that AlexNet, EfficientNet, and ResNet, all demonstrate significant AUC values of 1 which proves the efficiency of these models in identifying true labels.

As illustrated in the confusion matrices in Figure 52, AlexNet, EfficientNet, LeNet, and ResNet, demonstrate significant classification performance where there was 0 misclassification in EfficientNet, 8 for AlexNet,9 for LeNet, and 1 for ResNet

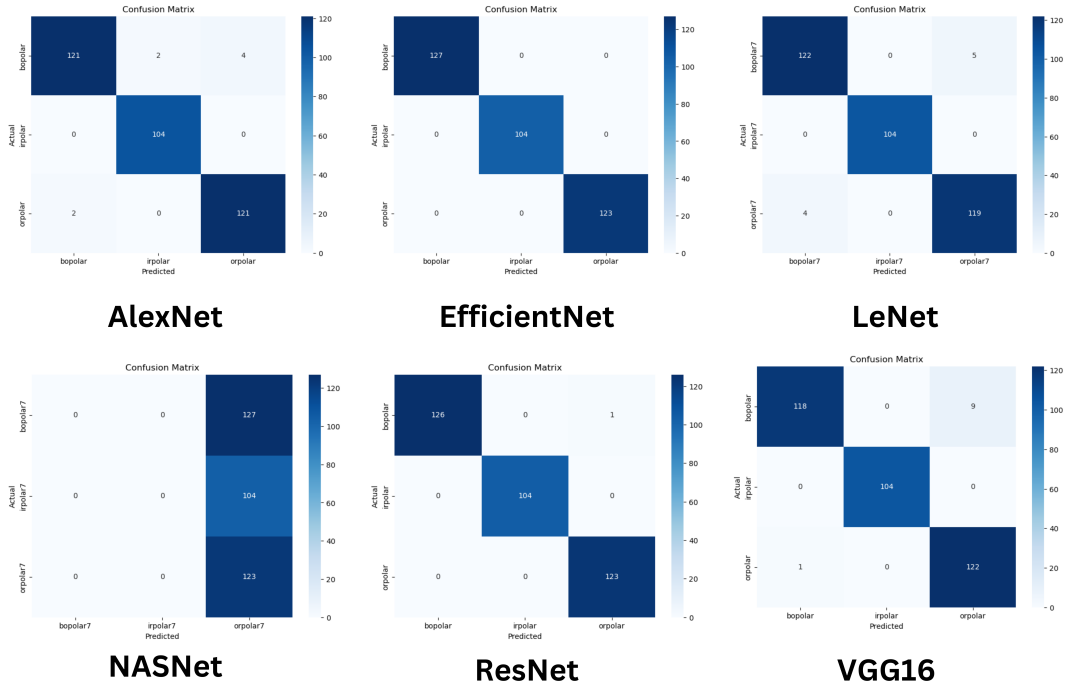


Figure 52: Single channel confusion matrix using frequency spectrum

.

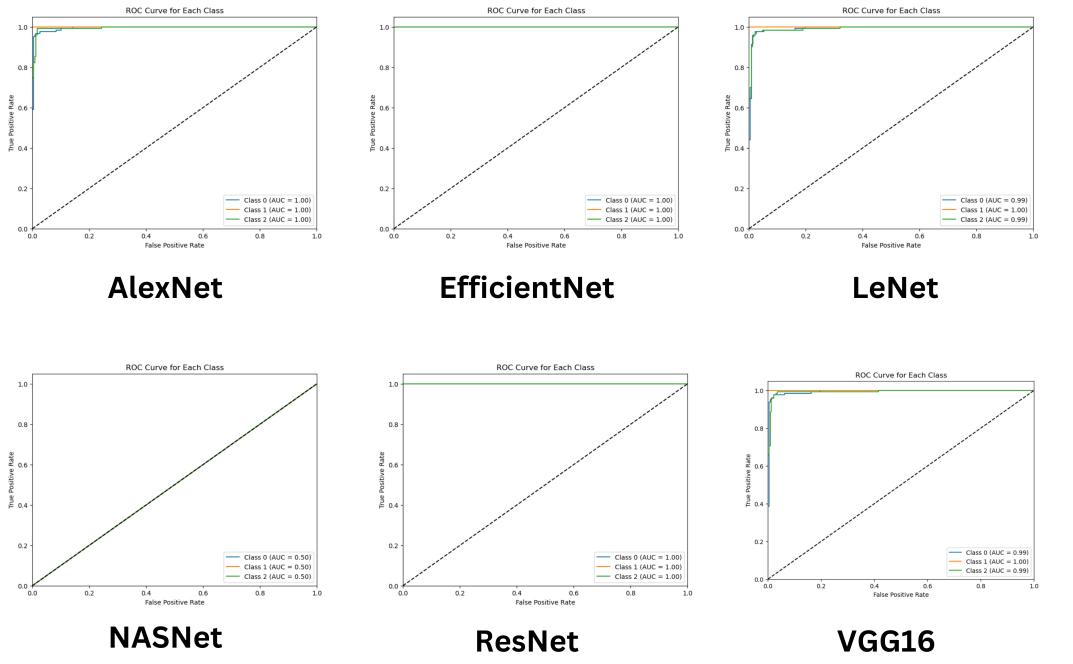


Figure 53: Single channel roc curve using frequency spectrum

.

5.3 Summary of the results and discussions

To summarize the results of the 2D CNN models, the results are as expected. Tables 16, and 17 provide a comprehensive insight into the efficiency of the different developed deep learning methods on a

single-channel and dual-channel approach based on the CWRU and NASA datasets and using Polar spectrums, frequency spectrums, and spectrograms feature extraction techniques. Given the results derived from these models, we can justify the superiority of the dual-channel models in comparison with their corresponding single-channel methods. We can observe a significant positive change in the classification performance of All the models except NASNet architecture.

Within the CWRU dataset single channel analysis, we can observe that spectrograms have contributed to providing higher classification accuracy in comparison with other 2D features that were used in the experiments. On the other hand, in the NASA dataset, polar spectrums have been significantly contributing to revealing features that can provide higher classification performance.

In the CWRU dataset, among all the evaluated models, Multi-channel EfficientNet has provided the highest accuracy of 98.3 %, 100%, and 100% respectively using polar spectrums, spectrograms, and frequency spectrums. Within this dataset, we also could observe that a multichannel model ResNet model using a spectrogram can also provide the highest accuracy of 100%. In the NASA dataset as well, EfficientNet demonstrated a significant classification performance of 99.65%, 100%, and 100% respectively using polar spectrums, spectrograms, and frequency spectrums. By comparing the results we can obtain that in both datasets, Multi EfficientNet using spectrograms and frequency spectrums has outperformed other classification models. In the CWRU dataset, we can obtain the classification time using the best hyperparameters for spectrograms as 232.34 ms and 336.89 ms for frequency spectrums which demonstrates the superiority of this feature over frequency spectrums.

In conclusion, We observed how different architectures would demonstrate different performances for anomaly classification within the bearing, and while architectures like EfficientNet and ResNet indicated astonishing results, NASNet was failing the experiments. By evaluating the classification performance using varied channels, different models, and different features we could justify the superiority of the multi-channel models in comparison with their peer single-channel architecture and also identify the EfficientNet using spectrograms as the highest performance model by the accuracy of 100%.

Table 5.3 shows a comparison of the results of the developed deep learning models with the state-of-the-art algorithms. Scholars have tried different methods and different features to evaluate and increase the classification performance. Tao *et al* (T. Lu et al., 2020), Khorram *et al* (Khorram et al.,

CWRU dataset		Single	Dual
AlexNet	Polars	0.550	0.983
	Spectrogram	0.968	0.977
	Spectrums	0.55	0.991
VGG16	Polars	0.677	0.974
	Spectrogram	0.723	0.966
	Spectrums	0.79	0.971
EfficientNet	Polars	0.724	0.983
	Spectrogram	0.838	1
	Spectrums	0.641	1
LeNet	Polars	0.635	0.967
	Spectrogram	0.889	0.997
	Spectrums	0.853	0.974
NASNet	Polars	0.641	0.596
	Spectrogram	0.711	0.361
	Spectrums	0.358	0.347
ResNet	Polars	0.682	0.977
	Spectrogram	0.683	1
	Spectrums	0.539	0.997

Table 16: Performance of various networks on the CWRU dataset

NASA dataset		Single	Dual
AlexNet	Polars	0.961	0.9930
	Spectrogram	0.662	0.993
	Spectrums	0.658	0.993
VGG16	Polars	0.9712	0.9895
	Spectrogram	0.766	0.986
	Spectrums	0.4912	0.993
EfficientNet	Polars	0.459	0.9965
	Spectrogram	0.905	1
	Spectrums	0.905	1
LeNet	Polars	0.9686	0.9930
	Spectrogram	0.324	0.993
	Spectrums	0.324	0.993
NASNet	Polars	0.8217	0.331
	Spectrogram	0.658	0.641
	Spectrums	0.656	0.331
ResNet	Polars	0.9285	0.9965
	Spectrogram	0.662	0.996
	Spectrums	0.662	0.996

Table 17: Performance of various networks on the NASA dataset

2021), Xiang *et al* (X. Li et al., 2019), and Guo *et al* (Guo et al., 2018) have tried 1 channel CNN models. The results ranged from 70% to 99.7%. Tao *et al* (T. Lu et al., 2020), by providing 99.7% accuracy using AlexNet and spectrogram, has demonstrated a significant result. However, their research lacks validation on a second dataset. Add to the point, that using one type of data may not provide extensive

insight into the data and may not be applicable for real-world scenarios. Wang *et al* (J. Wang et al., 2021), and Ramin*et al* (Rajabioun et al., 2023) have applied multi-sensor fusion methods. Wang *et al* (J. Wang et al., 2021) has leveraged 2D images of vibrational signals of CWRU in their experiments and achieved 99.92% accuracy. Ramin*et al* (Rajabioun et al., 2023) also by using raw signals within a self-generated dataset could achieve 98% accuracy. In comparison to our experiments, our proposed dual-channel method leverages different features derived from one specific sensor. Mao *et al* (Mao et al., 2018), and Sayanjit*et al* (S. S. Roy et al., 2020) have used machine learning methods within their experiments. They have leveraged ISVM and random forest classifier and achieved 98.3% and 97.9% in bearing fault classification. In our work, we have applied a mid-fusion architecture by simultaneously leveraging 1D and 2D channels. We achieved promising results, and could significantly prove the efficiency of the Dual-Channel EfficientNet using spectrograms and frequency spectrums by providing 100% accuracy on both datasets. This validation can make our model adjustable with real-world data and benefit industries by providing more accurate and robust classification performance. Also, applying all these different experiments can become a blueprint for other researchers for their further research in this domain.

Author	**Dataset**	**Feature**	**Model**	**Accuracy**
Wang *et al* (H. Wang et al., 2022)	CWRU	highest envelope spectral kurtosis	Multi-sensor 1D fused VGG16	100%
Shao *et al* (Shao, Jiang, Zhang, & Liang, 2017)	Electric Locomotive Bearing Dataset	vibrational data	CDBN	97.43%
Wang *et al* (P. Wang et al., 2017)	Gearbox data	Wavelets	DCNN	99.58%
Jiang *et al* (Jiang et al., 2019)	CWRU	Time series data	GAN	100%

Continued on next page

Table 18 – continued from previous page

Author	Dataset	Feature	Model	Accuracy
Sayanjit*et al* (S. S. Roy et al., 2020)	CWRU-NASA	Auto correlation on vibrational data	Random forest classifier	99.2%-97.9%
Ronak *et al* (Bhadra et al., 2018)	NASA	Statistical features	1D CNN - 2D CNN	93.3%-92.9%
Mao *et al* (Mao et al., 2018)	CWRU	Frequency spectrum	ISVM	98.33%
Roy *et al* (M. Roy et al., 2018)	NASA	Auro encoder	OSELM	100%
Purarj *et al* (Purarjomandlangrudi et al., 2014)	NASA	FFT spectrum - kutrosis - Non Guassianility score	Gaussian distribution	95%
Tao *et al* (T. Lu et al., 2020)	CWRU	Spectogram	AlexNet	99.7%
Wang *et al* (J. Wang et al., 2021)	CWRU - NASA	signal to image	Multi sensor fusion usning CNN	99.92% - 92.68%
Khorram *et al* (Khorram et al., 2021)	NASA	time series data	CNN-LSTM	97.13%
Enshaei *et al* (Enshaei & Naderkhani, 2019)	CWRU	raw vibrational data	deep deep bidirectional long short-term memory	100%

Continued on next page

Table 18 – continued from previous page

Author	Dataset	Feature	Model	Accuracy
Xiang *et al* (X. Li et al., 2019)	CWRU	deep transfer learning	DNN	70%
Guo *et al* (Guo et al., 2018)	CWRU	raw signal data	1D CNN	82%
Ramin*et al* (Rajabioun et al., 2023)	self generated data	raw signal	Multi sensor fusion	98%
Sayanjit*et al* (S. S. Roy et al., 2020)	CWRU - NASA	Auto correlation	random forest classifier	99.9% - 97.9%
Meng*et al* (S. S. Roy et al., 2020)	CWRU	Mapping to Greyscale image	improved convolutional neural network (CNN) incorporating a convolutional attention	97.13%
saadi*et al* (Nacer, Nadia, Abdelghani, & Mohamed, 2023)	CWRU	Greyscale image-reshape-fft	ANN and SVM	99.9%
Ruxue*et al* (Bai et al., 2021)	NASA	pyramidal dilatated convolution is designed to extract the features	ResNet-FCF	99.55%

Continued on next page

Table 18 – continued from previous page

Author	Dataset	Feature	Model	Accuracy
Yuhang*et al* (Chen, Xiao, & Li, 2022)	CWRU	Rw vibrational signal	1D CNN – trains at one speed test at other speed	99.29%
Zaming*et al* (Liu, Zhang, Meng, & Zhang, 2022)	CWRU	Wavelets	CNN for feature extraction sparrow search algorithm (SSA) optimized extreme learning machine for classification	97%
Our work (Dual-channel Efficient-Net)	**CWRU - NASA**	**Spectogram/ Frequency spectrum and statistical features**	**2D-1D CNN**	**100% - 100%**

6 Conclusion and Future Work

6.1 Conclusion

In this thesis, we provided a comprehensive study on bearing anomaly classification using varied deep-learning models including AlexNet, EfficientNet, LeNet, NASNet, ResNet, and VGG16. Then, in order to overcome the misclassification challenge among the different fault types, we incorporated statistical features into our models using mid-level fusion and developed dual-channel models that could significantly take advantage of varied features to increase the classification performance. In order to validate our hypothesis, we tried this method on both the NASA dataset and the CWRU dataset.

By conducting experiments based on different features, datasets, and models and also comparing their results, we could justify the superiority of the dual-channel models and prove how they can contribute to providing better solutions for anomaly classification within the bearing. In the single-channel approach, AlexNet, and VGG16 have demonstrated higher classification performance within both CWRU and and NASA datasets respectively. Using AlexNet, in the CWRU dataset, Spectograms have contributed to the highest accuracy of 96.8%, and in using VGG16 in the NASA dataset polars have been used to generate the highest classification accuracy of 97.2%. In a dual-channel approach, these two classification performances have been increased to 97.7%, and 99.3% respectively. However, EfficientNet using Spectogram has demonstrated a higher classification performance of 100% within both datasets. Based on our results, we believe that the industries that are providing predictive maintenance solutions can integrate our proposed method to increase their performance, and scholars can leverage this comprehensive comparative analysis to provide further solutions.

6.2 Future Work

Anomaly classification in industrial machinery plays a significant role in providing predictive maintenance. In order to delve deeper into the concept of this thesis, and provide decisive solutions that can enhance the performance of the rotating equipment, evaluating and deploying this method in real-time situations can be considered as a future work. Additionally, our evaluation was solely signal-based, and we have tried only signal-based features in our experiments. Therefore, incorporating new types of features such as temperature data and lubrication analysis can be useful in providing more comprehensive results and it can be considered as another future work.

References

Abbasi, T., Lim, K. H., & San Yam, K. (2019). Predictive maintenance of oil and gas equipment using recurrent neural network. In *Iop conference series: Materials science and engineering* (Vol. 495, p. 012067).

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1–74.

Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., & Consortium, P. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making*, *20*, 1–9.

Aung, Z., Mikhaylov, I. S., & Aung, Y. T. (2020). Artificial intelligence methods application in oil industry. In *2020 ieee conference of russian young researchers in electrical and electronic engineering (eiconrus)* (pp. 563–567).

Baesler, F. F., Moraga, M., & Ramis, F. J. (2002). Productivity improvement in the wood industry using simulation and artificial intelligence. In *Proceedings of the winter simulation conference* (Vol. 2, pp. 1095–1098).

Bagchi, S., & Mitra, S. K. (2012). *The nonuniform discrete fourier transform and its applications in signal processing* (Vol. 463). Springer Science & Business Media.

Bai, R., Xu, Q., Meng, Z., Cao, L., Xing, K., & Fan, F. (2021). Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation. *Measurement*, *184*, 109885.

Bazi, Y., Al Rahhal, M. M., Alhichri, H., & Alajlan, N. (2019). Simple yet effective fine-tuning of deep cnns using an auxiliary classification loss for remote sensing scene classification. *Remote Sensing*, *11*(24), 2908.

Bechhoefer, E., & Fang, A. (2012). Algorithms for embedded phm. In *2012 ieee conference on prognostics and health management* (pp. 1–8).

Bhadra, R., Dutta, S., Kedia, A., Gupta, S., Panigrahy, P. S., & Chattopadhyay, P. (2018). Applied machine learning for bearing fault prognostics. In *2018 ieee applied signal processing conference (aspcon)* (pp. 158–162).

Castells, F., Laguna, P., Sörnmo, L., Bollmann, A., & Roig, J. M. (2007). Principal component analysis in ecg signal processing. *EURASIP Journal on Advances in Signal Processing*, *2007*, 1–21.

Chen, Y., Xiao, L., & Li, Z. (2022). Partial domain fault diagnosis of bearings under cross-speed con-

ditions based on 1d-cnn. In *2022 global reliability and prognostics and health management (phm-yantai)* (pp. 1–8).

Compare, M., Baraldi, P., & Zio, E. (2019). Challenges to iot-enabled predictive maintenance for industry 4.0. *IEEE Internet of Things Journal*, *7*(5), 4585–4597.

Davari, N., Veloso, B., Costa, G. d. A., Pereira, P. M., Ribeiro, R. P., & Gama, J. (2021). A survey on data-driven predictive maintenance for the railway industry. *Sensors*, *21*(17), 5739.

Dey, D., & Jana, R. (2022). Bearing fault predictive maintenance using lstm. In *2022 3rd international conference on computing, analytics and networks (ican)* (pp. 1–6).

Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, *9*(2), 85–112.

Enshaei, N., & Naderkhani, F. (2019). Application of deep learning for fault diagnostic in induction machine's bearings. In *2019 ieee international conference on prognostics and health management (icphm)* (pp. 1–7).

Feldman, M. (2011). Hilbert transform in vibration analysis. *Mechanical systems and signal processing*, *25*(3), 735–802.

Goutham, V., Sameerunnisa, A., Babu, S., & Prakash, T. B. (2022). Brain tumor classification using efficientnet-b0 model. In *2022 2nd international conference on advance computing and innovative technologies in engineering (icacite)* (pp. 2503–2509).

Griffin, D., & Lim, J. (1984). Signal estimation from modified short-time fourier transform. *IEEE Transactions on acoustics, speech, and signal processing*, *32*(2), 236–243.

Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2018). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, *66*(9), 7316–7325.

Hakim, M., Omran, A. A. B., Inayat-Hussain, J. I., Ahmed, A. N., Abdellatef, H., Abdellatif, A., & Gheni, H. M. (2022). Bearing fault diagnosis using lightweight and robust one-dimensional convolution neural network in the frequency domain. *Sensors*, *22*(15), 5793.

Haque, M. A., Verma, A., Alex, J. S. R., & Venkatesan, N. (2020). Experimental evaluation of cnn architecture for speech recognition. In *First international conference on sustainable technologies for computational intelligence: Proceedings of ictsci 2019* (pp. 507–514).

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). Mlcm: Multi-label confusion matrix. *IEEE Access*, *10*, 19083–19095.

Javaid, M., Haleem, A., Singh, R. P., Rab, S., & Suman, R. (2021). Significance of sensors for industry 4.0: Roles, capabilities, and applications. *Sensors International*, *2*, 100110.

Jiang, W., Hong, Y., Zhou, B., He, X., & Cheng, C. (2019). A gan-based anomaly detection approach for imbalanced industrial time series. *IEEE Access*, *7*, 143608-143619. doi: 10.1109/ACCESS.2019.2944689

Jin, X., Sun, Y., Que, Z., Wang, Y., & Chow, T. W. (2016). Anomaly detection and fault prognosis for bearings. *IEEE Transactions on Instrumentation and Measurement*, *65*(9), 2046–2054.

Khorram, A., Khalooei, M., & Rezghi, M. (2021). End-to-end cnn+ lstm deep learning approach for bearing fault diagnosis. *Applied Intelligence*, *51*(2), 736–751.

Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., ... others (2023). What is machine learning, artificial neural networks and deep learning?—examples of practical applications in medicine. *Diagnostics*, *13*(15), 2582.

Latif, S., Zou, Z., Idrees, Z., & Ahmad, J. (2020). A novel attack detection scheme for the industrial internet of things using a lightweight random neural network. *IEEE access*, *8*, 89337–89350.

Li, X., Zhang, W., Ding, Q., & Li, X. (2019). Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE transactions on industrial informatics*, *16*(3), 1688–1697.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*.

Liu, X., Zhang, Z., Meng, F., & Zhang, Y. (2022). Fault diagnosis of wind turbine bearings based on cnn and ssa–elm. *Journal of Vibration Engineering & Technologies*, 1–17.

Lu, Q. (2012). The processing of the rolling bearing's fault signal based on wavelet analysis. In *2012 international conference on image analysis and signal processing* (pp. 1–5).

Lu, T., Yu, F., Han, B., & Wang, J. (2020). A generic intelligent bearing fault diagnosis system using convolutional neural networks with transfer learning. *IEEE Access*, *8*, 164807–164814.

Lyzenga, D. R. (2017). Polar fourier transform processing of marine radar signals. *Journal of Atmospheric and Oceanic Technology*, *34*(2), 347–354.

Mao, W., Tian, S., Liang, X., & He, J. (2018). Online bearing fault diagnosis using support vector machine and stacked auto-encoder. In *2018 ieee international conference on prognostics and health management (icphm)* (pp. 1–7).

Matzka, S. (2020). Explainable artificial intelligence for predictive maintenance applications. In *2020 third international conference on artificial intelligence for industries (ai4i)* (pp. 69–74).

Muthamizh Selvan, A., & Rajesh, R. (2012). Spectral histogram of oriented gradients (shogs) for tamil language male/female speaker classification. *International Journal of Speech Technology*, *15*, 259–264.

Nacer, S. M., Nadia, B., Abdelghani, R., & Mohamed, B. (2023). A novel method for bearing fault diag-

nosis based on bilstm neural networks. *The International Journal of Advanced Manufacturing Technology*, *125*(3-4), 1477–1492.

Narkhede, S. (2018). Understanding auc-roc curve. *Towards Data Science*, *26*(1), 220–227.

Nayana, B., & Geethanjali, P. (2017). Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal. *IEEE Sensors Journal*, *17*(17), 5618–5625.

Neupane, D., & Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, *8*, 93155–93178.

Nivesrangsan, P., & Jantarajirojkul, D. (2018). Bearing fault monitoring by comparison with main bearing frequency components using vibration signal. In *2018 5th international conference on business and industrial research (icbir)* (pp. 292–296).

Osadchiy, A., Kamenev, A., Saharov, V., & Chernyi, S. (2021). Signal processing algorithm based on discrete wavelet transform. *Designs*, *5*(3), 41.

Patel, S. (2020). A comprehensive analysis of convolutional neural network models. *International Journal of Advanced Science and Technology*, *29*(4), 771–777.

Purarjomandlangrudi, A., Ghapanchi, A. H., & Esmalifalak, M. (2014). A data mining approach for fault diagnosis: An application of anomaly detection algorithm. *Measurement*, *55*, 343–352.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration*, *289*(4-5), 1066–1090.

Radhika, K., Devika, K., Aswathi, T., Sreevidya, P., Sowmya, V., & Soman, K. (2020). Performance analysis of nasnet on unconstrained ear recognition. *Nature inspired computing for data science*, 57–82.

Rajabioun, R., Afshar, M., Mete, M., Atan, Ö., & Akin, B. (2023). Distributed bearing fault classification of induction motors using 2d deep-learning model. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*.

Roy, M., Bose, S. K., Kar, B., Gopalakrishnan, P. K., & Basu, A. (2018). A stacked autoencoder neural network based automated feature extraction method for anomaly detection in on-line condition monitoring. In *2018 ieee symposium series on computational intelligence (ssci)* (pp. 1501–1507).

Roy, S. S., Dey, S., & Chatterjee, S. (2020). Autocorrelation aided random forest classifier-based bearing fault detection framework. *IEEE Sensors Journal*, *20*(18), 10792–10800.

Sacerdoti, D., Strozzi, M., & Secchi, C. (2023). A comparison of signal analysis techniques for the diagnostics of the ims rolling element bearing dataset. *Applied Sciences*, *13*(10), 5977.

Samimifard, Z. D. (2023). *Binary anomaly classification using lstm.* (Assignment submitted for CS4020 at University of South Eastern Noway (USN))

Sapijaszko, G., & Mikhael, W. B. (2018). An overview of recent convolutional neural network algorithms for image recognition. In *2018 ieee 61st international midwest symposium on circuits and systems (mwscas)* (pp. 743–746).

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2019a). Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *Ieee Access*, *7*, 122644–122662.

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2019b). Low-speed bearing fault diagnosis based on arssae model using acoustic emission and vibration signals. *IEEE Access*, *7*, 46885–46897.

Shah, D. S., & Patel, V. N. (2014). A review of dynamic modeling and fault identifications methods for rolling element bearing. *Procedia Technology*, *14*, 447–456.

Shahid, S. M., Ko, S., & Kwon, S. (2022). Performance comparison of 1d and 2d convolutional neural networks for real-time classification of time series sensor data. In *2022 international conference on information networking (icoin)* (pp. 507–511).

Shao, H., Jiang, H., Wang, F., & Zhao, H. (2017). An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowledge-Based Systems*, *119*, 200–220.

Shao, H., Jiang, H., Zhang, H., & Liang, T. (2017). Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. *IEEE Transactions on Industrial Electronics*, *65*(3), 2727–2736.

Wang, H., Sun, W., He, L., & Zhou, J. (2022). Rolling bearing fault diagnosis using multi-sensor data fusion based on 1d-cnn model. *Entropy*, *24*(5), 573.

Wang, J., Wang, D., Wang, S., Li, W., & Song, K. (2021). Fault diagnosis of bearings based on multi-sensor information fusion and 2d convolutional neural network. *IEEE Access*, *9*, 23717–23725.

Wang, P., Yan, R., Gao, R. X., et al. (2017). Virtualization and deep recognition for system fault classification. *Journal of Manufacturing Systems*, *44*, 310–316.

Zeidan, F. Y., Paquette, D. J., et al. (1994). Application of high speed and high performance fluid film bearings in rotating machinery. In *Proceedings of the 23rd turbomachinery symposium*.

Zhang, B., Georgoulas, G., Orchard, M., Saxena, A., Brown, D., Vachtsevanos, G., & Liang, S. (2008). Rolling element bearing feature extraction and anomaly detection based on vibration monitoring. In *2008 16th mediterranean conference on control and automation* (pp. 1792–1797).

Zhang, B., Sconyers, C., Byington, C., Patrick, R., Orchard, M. E., & Vachtsevanos, G. (2010). A prob-

abilistic fault detection approach: Application to bearing fault detection. *Ieee transactions on industrial electronics*, *58*(5), 2011–2018.

Zhang, X., Zhao, B., & Lin, Y. (2021). Machine learning based bearing fault diagnosis using the case western reserve university data: a review. *IEEE Access*, *9*, 155598–155608.

Appendices

Appendix

The codes can be accessed through the following GitHub repository:

https://github.com/zahra88ir/MasterThesis