



UNIVERSITY OF 20 AUGUST 1955 OF SKIKDA

Faculty of technology

Department of petrochemical and process engineering

Thesis presented for the degree of Master in petrochemical industries

Option: Automation in petrochemical industries

Rotary machine's fault diagnosis and prognosis using artificial intelligence

by

Khencha Mohammed Said

Boudiar Achraf

Supervised by

Dr. Kerboua Adlen

June 2019

I. ACKNOWLEDGMENT

We humbly bow our head with utmost gratitude before the God Almighty who always showed me the path to go and without whom I could not have done any of these.

Completion of this project and thesis would not have been possible without the help of many people, to whom we are very thankful. First, we would like to express my sincere gratitude to our supervisor, Dr. Kerboua Adlen. His constant motivation, guidance and support helped us a great deal to achieve this feat.

We are also thankful to the faculty teachers of the department of petrochemical and process engineering for guiding and inspiring me in many ways.

We are deeply indebted to our parents who always had their belief in us and gave all their support for all the choices that we have made.

Khencha Mohammed Said

Boudiar Achraf

chapters

neural network basics

CNN

TCN

Rotary machine faults detection

Vibration based faults diagnosis.

Signal processing.

Experimental testing bed

Implementation

Device designing

Framework

Neural network with details

Acquisition interface

Results and discussion

II. Table of Contents

I. ACKNOWLEDGMENT	2
chapters	3
II. Table of Contents	4
III. List of Figures	5
IV. List of Tables	7
V. List of abbreviations	8
1. Introduction.....	9
Background and Motivation	10
Problem definition	10
Thesis organization	11
2. Neural Network Basics	12
2.1. Introduction.....	13
2.2. Convolutional Neural Network (CNN):	15
2.3. Recurrent Neural Network (RNN):.....	17
2.4. Proposed fault diagnosis approach	18
2.4.1. Convolution layer	19
2.4.2. Pooling layer	19
2.4.3. Batch Normalization (BatchNorm).....	20
2.4.4. ReLU (Rectified Linear Unit).....	20
2.4.5. Global Average Pooling.....	21
2.4.6. Fully connected layers	22
2.4.7. SoftMax Layer	23
2.4.8. Classification Layer	23
3. Vibration-Based Fault Diagnosis.....	24
3.1. Introduction.....	25
3.2. Common faults in rotary machines	25
3.2.1. Bearing Failures	26
a. Inner Race Fault.....	28
b. Outer Race Fault	28
c. Ball Fault:	28
3.2.2. Shafts	30
a. Bow:.....	30
b. Coupling Fault:	30
c. Foundation Fault:.....	30
3.2.3. Gears	30
a. Chipped Teeth:.....	30
b. Missing Teeth:	30
c. Root Cracks:	31
3.2.4. Imbalance	31
3.2.5. Misalignment	31
3.2.6. Looseness.....	32
3.2.7. Rubbing.....	32

3.3.	Signal Processing	33
	i. Time domain analysis:	33
	ii. Frequency domain analysis.....	34
	iii. Time Frequency domain analysis:.....	34
3.3.1.	Maximal overlap discrete wavelet transforms (modwpt)	34
3.3.2.	Continuous 1-D wavelet transforms (cwt).....	35
3.3.3.	Short-time Fourier transforms(stft).....	36
3.3.4.	Cross spectrogram using short-time Fourier transforms(xspectrogram) 37	
3.3.5.	Power spectrum(pspectrum)	38
3.3.6.	Visualize spectral kurtosis(kurtogram).....	39
3.3.7.	The instantaneous frequency (instfreq).....	40
3.3.8.	The Spectral kurtosis (Pkurtosis).....	41
3.4.	Experimental testing bed	42
3.4.1.	Simulated faults	44
	3.4.1.1. Bearing Faults	44
	3.4.1.2. Misalignment	44
	3.4.1.3. Imbalance.....	45
4.	Implementation	46
4.1.	Device Designing.....	47
4.1.1.	Arduino and sensor	50
	4.1.1.1. Arduino	50
	4.1.1.2. Role and programming	50
	4.1.1.3. Sensor connection	50
	4.1.1.4. Programming	52
	4.1.1.5. Send data.....	52
	4.1.1.6. Energy consumption	53
	4.1.1.7. Power system	53
	4.1.1.8. Anti-discharge protection	54
	4.1.1.9. Idle mode	54
4.2.	Framework	55
	4.2.1.1. Neural Network in details	55
	4.2.1.2. Proposed ResNet structure.....	55
	4.2.1.3. Acquisition interface.....	57
4.3.	Results and Discussion	60
	Bibliography	63

III. List of Figures

Figure 2.1.	An artificial neuron	14
Figure 2.2.	Regular neural network.	15
Figure 2.3.	Architecture of a typical CNN.....	17
Figure 2.4.	Structure of a typical RNN.	18
Figure 2.5.	ReLU function.....	21
Figure 3.1.	Stages of vibration-based fault diagnosis.....	25
Figure 3.2.	Basic parts of a rotating machinery	26

Figure 3.3. Bearings.....	27
Figure 3.4. Causes of bearing failures	28
Figure 3.5. Bearing faults	29
Figure 3.6. The vibration patterns of possible defects.....	29
Figure 3.7. (a)Chipped tooth (b)Missing tooth (c)Root crack	31
Figure 3.8. (a)angular misalignment(b)parallel misalignment	32
Figure 3.9. Mechanical looseness	32
Figure 3.10. Time and frequency domain representations of a signal	34
Figure 3.11. The modwpt.....	35
Figure 3.12. The cwt	36
Figure 3.13. The stft.....	37
Figure 3.14. The xspectrogram	38
Figure 3.15. The Pspectrum.....	39
Figure 3.16. The Kurtogram	40
Figure 3.17. The instfreq.....	41
Figure 3.18. The Pkurtosis.....	42
Figure 3.19. The PT 500	43
Figure 3.20. The simulated bearing faults	44
Figure 4.1. The 3-D printed box	47
Figure 4.2. The MPU-9250.....	48
Figure 4.3. The ESP-32.....	49
Figure 4.4. (a)Physical (b)electrical Connection	50
Figure 4.5. I2C-protocol	51
Figure 4.6. ESP-32 Access Point Mode.....	53
Figure 4.7. The device assembly	54
Figure 4.8. The ResNet structure	56
Figure 4.9. The interface of the application.....	57
Figure 4.10 Indication of successful connection	57
Figure 4.11. Signal acquired	58
Figure 4.12. Domain choosing.....	58
Figure 4.13. Method choosing	59
Figure 4.14. Domain and method have been chosen	59
Figure 4.15. Signal processed.....	59
Figure 4.16. The predicted status of the motor	60
Figure 4.17. Confusion matrix.....	61

IV. List of Tables

Table 3.1. Experiments details.....	45
Table 4.1. Parameters used to train the ResNet model	56
Table 4.2. Accuracies of signal processing techniques	61

V. List of abbreviations

ANN: Artificial neural network,

CNN: Convolutional neural network,

RNN: Recurrent neural network,

LSTM: Long short-term memory,

CWT: continuous wavelet transforms,

FT: Fourier transform,

1. Introduction

Background and Motivation

In the landscape of modern manufacturing, the increasing complexity and integration of engineering systems present challenges that transcend the simple replacement of components. Unanticipated faults in rotary machines can lead to consequences ranging from the replacement of inexpensive bearings to catastrophic accidents incurring significant losses. Conventional maintenance strategies, while valuable, may fall short in addressing the demands of today's intricate industrial environments.

Predictive maintenance emerges as an ideal solution, aiming to collect data, detect changes in equipment conditions, and unveil fault information. This proactive approach allows for timely maintenance, optimizing equipment lifespan without escalating the risk of failure. The key to successful predictive maintenance lies in assessing equipment health and revealing intricate details about current and potential faults through collected data.

However, implementing predictive maintenance encounters challenges, notably in accessing necessary industrial data, integrating, and leveraging this data for effective fault diagnosis and prognosis, and employing accurate data-driven models for fault prediction. Fortunately, the convergence of information and communication technologies (ICT) and artificial intelligence (AI), particularly deep learning, heralds a new era for predictive maintenance.

Problem definition

Traditional machine learning approaches for machinery fault diagnosis face limitations in handling massive real-time data due to intricate signal processing techniques. In the digital age, where large volumes of real-time data are collected, the need for a more efficient and automated approach becomes evident. This thesis addresses this challenge by proposing methods for fault diagnosis, classification, and prognosis of rotating machinery using a deep learning-based approach. This approach automates feature extraction from mechanical big data, eliminating the dependence on manual signal processing expertise.

This thesis proposes methods for fault diagnosis, classification and prognosis of rotating machinery using a deep learning-based approach that can automatically and effectively mine features from big mechanical data without the human expertise needed.

Thesis organization

This thesis is structured across three cohesive chapters. Chapter 1 establishes the thesis's motivation, introducing neural network fundamentals. Chapter 2 explores the intricacies of rotary machine fault diagnosis, encompassing vibration-based approaches, signal processing techniques, and an in-depth examination of studied faults. While in chapter 3 we detail the design process of the proposed fault diagnosis device, featuring the MPU-9250 accelerometer and ESP-32 microcontroller. This chapter further delves into the proposed neural network architecture, the signal processing steps and the acquisition interface, and presents tangible outcomes of fault diagnosis. It culminates in a comprehensive discussion, unraveling insights and potential avenues for refinement.

2. Neural Network Basics

2.1. Introduction

A neural network is a type of machine-learning algorithm that is modelled on the functioning of the brain. Back in 1943, two neuroscientists by the names of Warren McCulloch and Walter Pitts wrote a paper called “A Logical Calculus of the Ideas Imminent in Nervous Activity”. They expressed the functioning of neurons in the brain mathematically, fundamentally arriving at the “all or nothing principle” which expresses neuron activity in a very binary way – if the message from all the input neurons is sufficient, a neuron fires. If it is insufficient, it does not. This binary on/off behavior didn’t take long to make its way to computer scientists who saw obvious connections to the same “on/off” behavior of bits in the machine.

Frank Rosenblatt ran with this “mathematical brain” concept and introduced the Artificial Neural Network, the “Perceptron” in his paper in 1958. The evolution of artificial intelligence, machine learning, and deep learning has made so many people start asking questions about what exactly the process of machine learning is?

We found that data scientists, enthusiasts, and developers are very curious to learn how a neural network works for helping artificial intelligence to perform better. An artificial neural network is like a biological neural network in a brain. A biological neural network works as follows: information flows in, is processed by the neurons, and the results flow out. [1]

The basis of the neuron is to react to previously learned patterns. When we are creating the same kind of replication in terms of technology and computer science, we call it an artificial neural network. Just like the biological neuron, information flows in, is processed by an artificial neural network, and results flow out.

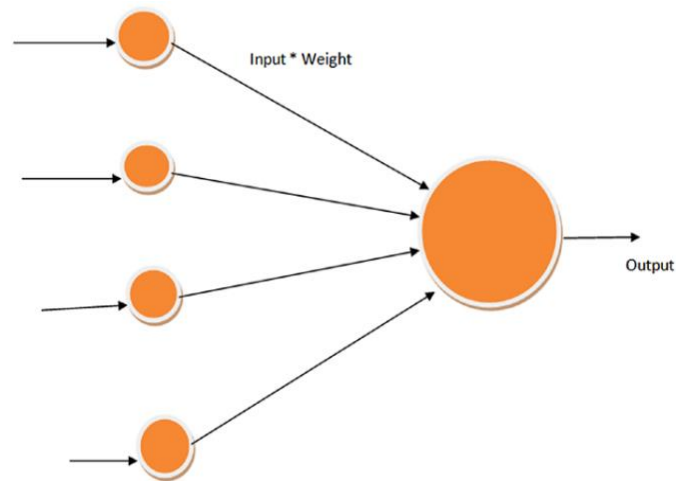


Figure 2.1. An artificial neuron

A neural network consists of interconnected nodes, or artificial neurons, organized in layers. The key components include:

- **Input Layer:** The input layer is the initial layer of a neural network that receives external data. Each node in this layer represents a feature or input. It serves as the entry point for information into the neural network.
- **Output Layer:** The output layer is the final layer that produces the network's result or prediction. Each node in this layer corresponds to an output or class. It provides the ultimate response or classification based on the network's processing.
- **Weights:** Weights are parameters associated with connections between neurons. They determine the strength of signals, influencing the importance of inputs. Weights control the impact of each input on the network's output during training.
- **Biases:** Biases are additional parameters in neurons that provide an offset, adjusting the output of the neuron. They contribute to the flexibility of the network. Biases allow the model to learn and adapt, improving the overall performance.
- **Activation function:** Activation functions introduce non-linearity to the network. They determine the output of a neuron, enabling the network to learn complex relationships. Activation functions help the neural network capture intricate patterns and relationships in the data.

Artificial neural networks are generally a chain of nodes associated with each other via the link from which they start interacting accordingly. Neurons perform operations and carry that result. [2]

A regular neural network looks like this:

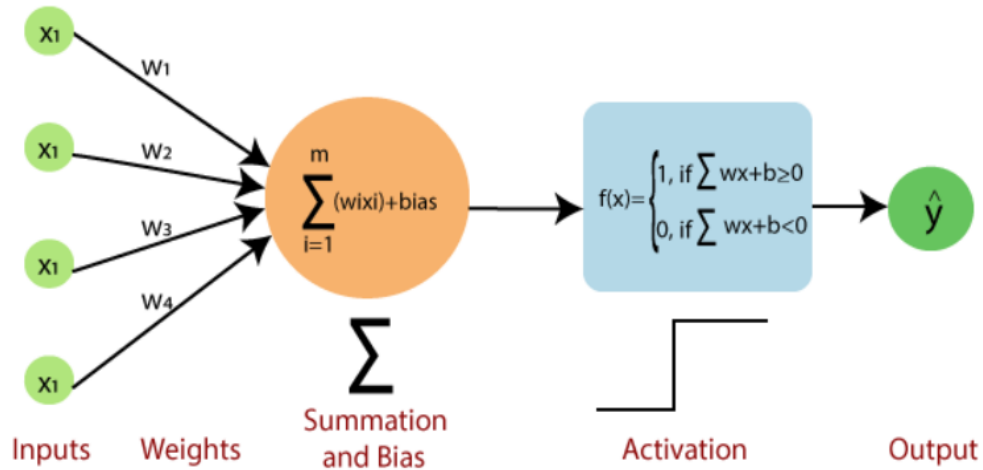


Figure 2.2. Regular neural network.

In the realm of fault diagnosis for rotating machinery, a variety of deep learning models, particularly neural networks, have been explored for their efficacy.

To understand the neural network models used in our study, we will first provide an overview of their structures and functionalities. This involves exploring various architectures, such as recurrent neural network (RNN), Convolutional Neural Network (CNN).

These models play distinct roles in processing vibration signals and sequential data, enhancing the accuracy of fault diagnosis. As part of this process, we will provide descriptions of each model and their applications in the context of rotating machinery fault diagnosis. [3]

2.2. Convolutional Neural Network (CNN):

A convolutional neural network, or CNN, is a deep learning neural network designed for processing structured arrays of data such as images. Convolutional neural networks are widely used in computer vision and have become the state of the art for

many visual applications such as image classification and have also found success in natural language processing for text classification. Convolutional neural networks are very good at picking up on patterns in the input image, such as lines, gradients, circles, or even eyes and faces. It is this property that makes convolutional neural networks so powerful for computer vision. [4]

Due to the ability to generalize the local and global features, CNNs turn out to be the most popular deep learning methods. CNNs are exceptionally successful in extracting features from input data and using them to make a trustworthy prediction. A basic CNN structure mainly has an input layer, convolution layer, pooling layer, and fully connected layer. [5]

The input data could be either two-dimensional or one-dimensional such as time–frequency spectrum or time series data, respectively. The convolution layer uses a set of weights and convolutes at each layer to form the layer-wise features, which are called a feature map.

The output of the convolutional layer is calculated as:

$$Y_n = f(X * W_n + b_n) \quad (1)$$

Where $*$ represents an operator of the convolution, n denotes the number of convolution filters, W_n is the weight matrix, b_n is the filter kernel bias vector.

Following the convolution, the model parameters are reduced by subsampling, named as pooling process. After the pooling layer, multiple fully connected layers are used to convert the matrix to a row or a column. Finally, a classification or regression layer is added to get the predictions or results. [5]

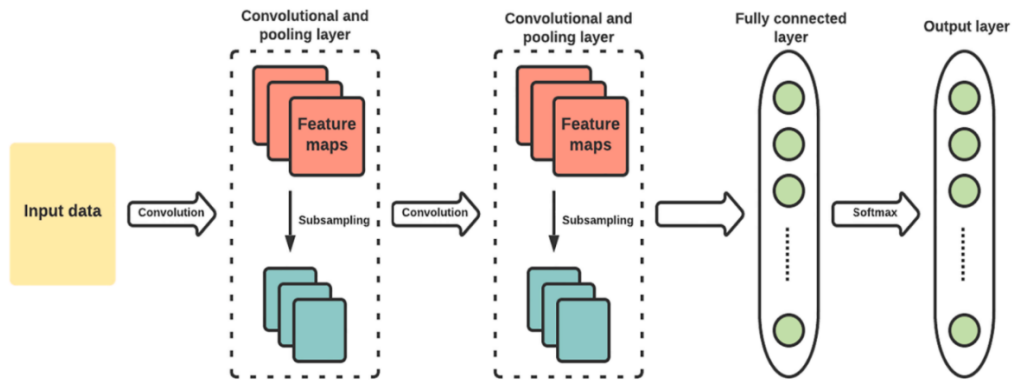


Figure 2.3. Architecture of a typical CNN.

2.3. Recurrent Neural Network (RNN):

A Recurrent Neural Network (RNN) is a type of artificial neural network designed for processing sequential data by incorporating feedback loops that allow information persistence across different time steps. RNNs are particularly effective in tasks where context and temporal dependencies are crucial, such as natural language processing, speech recognition, and time series analysis. One of the key features of RNNs is their ability to capture and remember information from earlier inputs, making them suitable for modelling dynamic patterns in sequential data. [6]

In the context of fault detection, RNNs have shown significant promise. They can be applied to sensor fault detection and mechanical fault diagnosis, providing a valuable tool for identifying abnormalities in various systems. RNNs have been utilized in the fault diagnosis of rotating machinery, demonstrating their effectiveness in identifying fault types based on one-dimensional data. Additionally, RNNs have been applied to condition monitoring in wind turbines, presenting a fault detection framework for analysing and modelling data to enhance the maintenance of these systems. [6]

As shown in Figure 2.4, the structure of the RNN constitutes a deep network with one layer per time step and shares the parameters across the layers. The concept of parameter sharing is a useful way to capture the relationship between one input item and its neighboring context. This makes the RNNs very successful over the traditional NNs and CNNs.

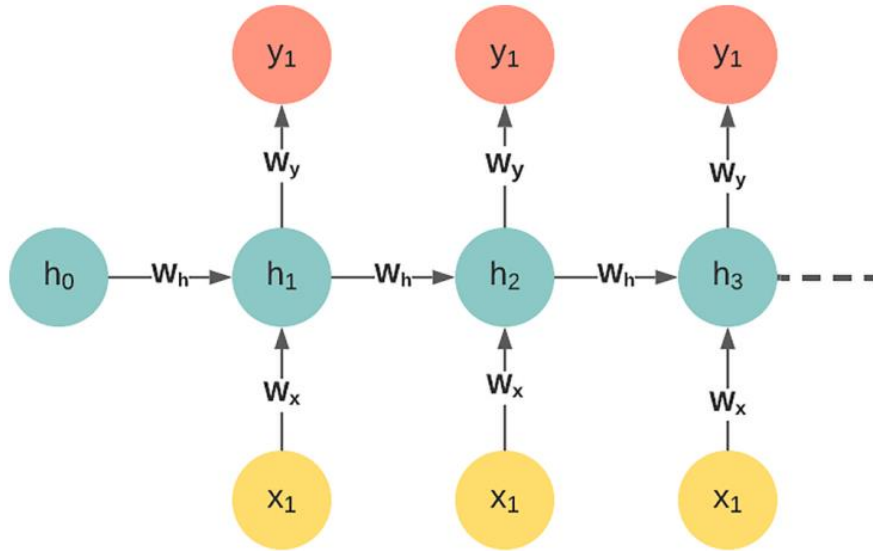


Figure 2.4. Structure of a typical RNN.

The recurrent nature of RNNs allows them to maintain a memory of past inputs, enabling the model to consider the entire input sequence when making predictions or identifying patterns. This makes RNNs well-suited for tasks that involve dependencies over time, offering advantages in capturing long-range dependencies and improving the understanding of sequential data. [6]

2.4. Proposed fault diagnosis approach

To detect faults, we train a ResNet neural network model (residual neural network) which is a deep learning model introduced in 2015 by researchers at Microsoft Research Asia. The model was designed to address the problem of vanishing gradients in deep neural networks, which can make it difficult to train models with many layers. ResNet uses a technique called residual learning, which involves adding shortcut connections between layers to allow information to bypass certain layers and flow directly to deeper layers. This helps to prevent the vanishing gradient problem and allows ResNet to be trained effectively even with hundreds of layers. [7]

ResNet has been used in a variety of applications, including image classification, object detection, and semantic segmentation. [8] It consists of multiple layers. The exact number of layers can vary depending on the specific implementation of ResNet, but the original ResNet model proposed by researchers at Microsoft Research in 2015 had 152

layers. The ResNet model is composed of several residual blocks, which are made up of convolutional layers, batch normalization layers, and activation functions such as ReLU. The residual blocks are connected by skip connections, which allow information to bypass certain layers and flow directly to deeper layers. This helps to prevent the vanishing gradient problem and allows ResNet to be trained effectively even with hundreds of layers. [9]

2.4.1. Convolution layer

The discrete convolution between two functions f and g is defined as:

$$(f * g)(x) = \sum_t f(t)g(x + t) \quad (2)$$

For 2-dimensional signals such as images, we consider the 2D-convolutions:

$$(K * I)(i, j) = \sum_{m, n} K(m, n)I(i + n, j + m). \quad (3)$$

K is a convolution kernel applied to a 2D signal (or image). The principle of 2D convolution is to drag a convolution kernel on the image.

At each position, we get the convolution between the kernel and the part of the image that is currently treated. Then, the kernel moves by a number s of pixels, s is called the stride. When the stride is small, we get redundant information. Sometimes, we also add a zero padding, which is a margin of size p containing zero values around the image to control the size of the output. [10]

2.4.2. Pooling layer

Pooling is a form of nonlinear down-sampling, which reduces the amount of calculation by reducing network parameters and can control overfitting to a certain extent. Currently, a pooling layer is added after the convolutional layer. Maximum pooling is to divide the input layer into different regions with non-overlapping rectangular boxes. The maximum number of rectangular boxes is taken as the output layer. The transformation function of maximum pooling is expressed as:

$$P_i^{l+1}(j) = \max_{(j-1)V+1 \leq n \leq jV} \{q_i^l(n)\} \quad (5)$$

where $q_i^l(n)$ represents the value of the n -th neuron in the i -th eigenvector of the l -th layer, $n \in [(j - 1)V + 1 \leq n \leq jV]$, V is the width of the pooling area, $P_i^{l+1}(j)$ represents the corresponding value of neurons in the $(l + 1)$ -th layer. [11]

2.4.3. Batch Normalization (BatchNorm)

Batch normalization is a technique used to improve the training speed and stability of neural networks. It normalizes the activations of each layer by adjusting and scaling them to have zero mean and unit variance. This helps alleviate issues like vanishing or exploding gradients during training. The formula for batch normalization applied to a mini-batch of size m and a feature map with n dimensions (assuming zero-centered data) is as follows:

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (6)$$

Where x_i is the input to the normalization layer, μ is the mean of the mini batch, σ^2 is the variance of the mini batch, ϵ is a small constant added for numerical stability.

Batch normalization also learns two additional parameters, γ and β , which scale and shift the normalized values, allowing the network to learn the optimal normalization for each layer. [12]

2.4.4. ReLU (Rectified Linear Unit)

ReLU is an activation function commonly used in neural networks to introduce non-linearity. It replaces all negative values in the input with zero, while leaving positive values unchanged. The ReLU function can be mathematically expressed as:

$$ReLU(x) = \max(0, x) \quad (7)$$

The rectified linear units come with three significant advantages in convolutional neural networks compared to the traditional logistic or hyperbolic tangent activation functions:

- Rectified linear units propagate the gradient efficiently and therefore reduce the likelihood of a vanishing gradient problem that is common in deep neural architectures.

- Rectified linear units' threshold negative values to zero, and therefore solve the cancellation problem as well as result in a much sparser activation volume at its output. The sparsity is useful for multiple reasons but mainly provides robustness to small changes in input such as noise.
- Rectified linear units consist of only simple operations in terms of computation (mainly comparisons) and therefore are much more efficient to implement in convolutional neural networks.

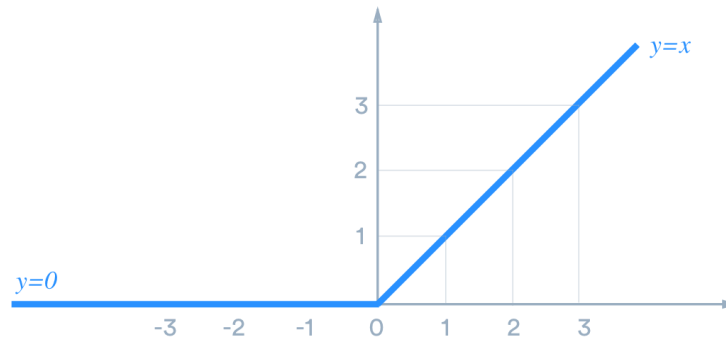


Figure 2.5. ReLU function.

As a result of its advantages and performance, most of the recent architectures of convolutional neural networks utilize only rectified linear unit layers (or its derivatives such as noisy or leaky ReLUs) as their non-linearity layers instead of traditional non-linearity and rectification layers. [13]

2.4.5. Global Average Pooling

Global average pooling is a technique often used in convolutional neural networks for reducing the spatial dimensions of feature maps to a single value per feature map. It computes the average of each feature map across its entire spatial extent. This operation helps reduce the number of parameters in the network and encourages spatially invariant features. The formula for global average pooling applied to a feature map with dimensions $H \times W$ is: [14]

$$\text{Global Avg Pooling}(x) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_{i,j} \quad (8)$$

2.4.6. Fully connected layers

The fully connected layers in a convolutional network are practically a multilayer perceptron (generally a two- or three-layer MLP) that aims to map the $m^{(l-1)}_1 \times m^{(l-1)}_2 \times m^{(l-1)}_3$ activation volume from the combination of previous different layers into a class probability distribution. Thus, the output layer of the multilayer perceptron will have $m^{(l-i)}_1$ outputs, i.e. output neurons where i denotes the number of layers in the multilayer perceptron.

The key difference from a standard multilayer perceptron is the input layer where instead of a vector, an activation volume is taken as the input. As a result, the fully connected layer is defined as:

If $l-1$ is a fully connected layer

$$y_i^{(l)} = f(z_i^{(l)}) \quad (9)$$

$$\text{with } z_i^{(l)} = \sum_{j=1}^{m_i^{(l-1)}} w_i^{(l)} y_j^{(l-1)} \quad (10)$$

otherwise.

$$y_i^{(l)} = f(z_i^{(l)}) \quad (11)$$

$$\text{with } z_i^{(l)} = \sum_{j=1}^{m_1^{(l-1)}} \sum_{r=1}^{m_2^{(l-1)}} \sum_{s=1}^{m_3^{(l-1)}} w_{i,j,r,s}^{(l)} (y_i^{(l-1)})_{r,s} \quad (12)$$

The goal of the complete fully connected structure is to tune the weight parameters $W^{(l)}_{i,j}$ or $W^{(l)}_{i,j,r,s}$ to create a stochastic likelihood representation of each class based on the activation maps generated by the concatenation of convolutional, non-linearity, rectification, and pooling layers. Individual fully connected layers operate identically to the layers of the multilayer perceptron with the only exception being the input layer. [15]

It is noteworthy that the function f once again represents non-linearity, however, in a fully connected structure the non-linearity is built within the neurons and is not a separate layer.

2.4.7. SoftMax Layer

SoftMax is an activation function often used in the output layer of a neural network for classification tasks. It converts the raw scores or logits produced by the network into probabilities that sum up to 1, representing the likelihood of each class. The SoftMax function is defined as

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (13)$$

Where z_i is the raw score for class i , and N is the total number of classes. [16]

2.4.8. Classification Layer

The classification layer is the final layer of a neural network used for classification tasks. It applies the SoftMax activation function to the network's output logits to obtain class probabilities. The predicted class is often the one with the highest probability. [17]

3. Vibration-Based Fault Diagnosis

3.1. Introduction

Machines consist of moving parts that generate sound and vibration. Each part has a specific vibration signal determined by the machine's state and construction. Changes in the vibration signature indicate incipient defects, allowing for detection and correction before failure.

Fault diagnosis and vibration control track industry mechanical components' performance using measured data, simulations, and heuristic experience. This proactive approach describes the current and expected future performance for critical events. Advanced intelligent fault diagnosis methods, with abundant measurement and simulation data, significantly enhance industry mechanical systems. [18]

Vibration-based fault diagnosis is a cutting-edge discipline crucial for assessing and optimizing mechanical systems' performance. It involves meticulous tracking of industrial mechanical components through reliable vibration signal measurements. By employing sophisticated methodologies, fault diagnosis detects, analyses, and controls deviations in mechanical systems, ensuring seamless operation. [19]

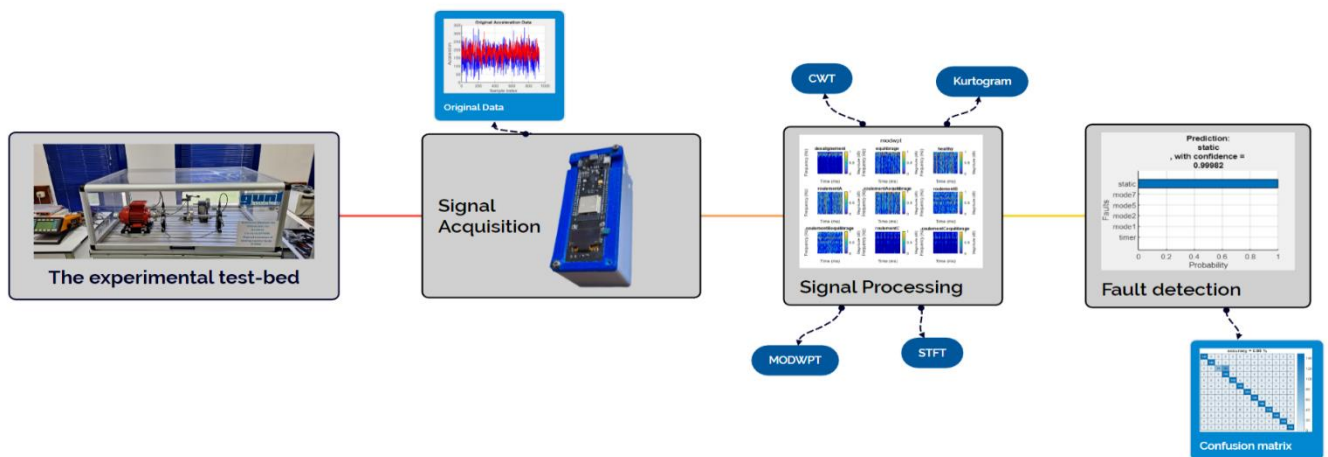


Figure 3.1. Stages of vibration-based fault diagnosis

3.2. Common faults in rotary machines

The faults in rotating machinery are a significant concern in various industrial applications, as they can lead to downtime, increased maintenance costs, and safety hazards. These faults manifest themselves at different frequencies in relation to the rotor spin frequency, making the identification of characteristic fault frequencies a crucial aspect of fault diagnosis.

Rotor unbalance, misalignment, bearing faults, and electrical faults are among the common issues encountered in rotating machinery. Differentiating these faults, irrespective of foundation flexibility and machine speed, is highly desirable for effective maintenance and fault diagnosis.

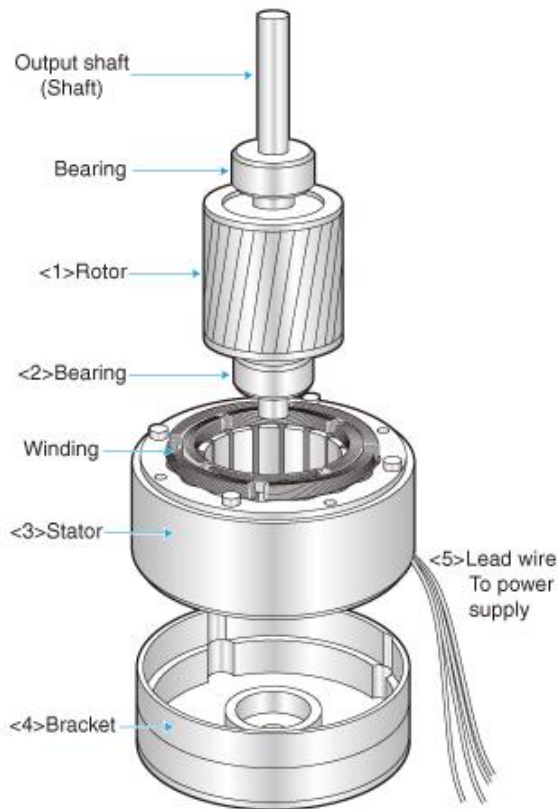


Figure 3.2. Basic parts of a rotating machinery

The most common faults in rotating machinery encompass a range of issues that can significantly impact the performance and reliability of the equipment. These faults include:

3.2.1. Bearing Failures

Bearings are essential components in rotating machinery, designed to reduce friction between moving parts and support the load. They play a crucial role in the optimal running of rotating machinery, and their failure can lead to significant downtime and maintenance costs. The type of bearing used in rotating equipment depends on factors such as the application, load capacity, speed, and operating conditions. [20]

There are various types of bearings used in different applications based on factors such as load, speed, and environmental conditions. Some common types of bearings used in rotating equipment include ball bearings, sleeve bearings, spherical roller bearings, and needle bearings. [20] Each type of bearing has its own unique characteristics and maintenance requirements.



Figure 3.3. Bearings

A bearing fault refers to any damage or wear and tear on any part of a bearing, which can lead to issues such as abrasion, pressure damage, fluting, and corrosion. [3b] Bearing faults are among the most common issues encountered in rotating machinery and can be caused by factors such as improper lubrication, contamination, and operational stress. [20]

Here's a chart that better displays the statistics behind common causes of bearing failures.

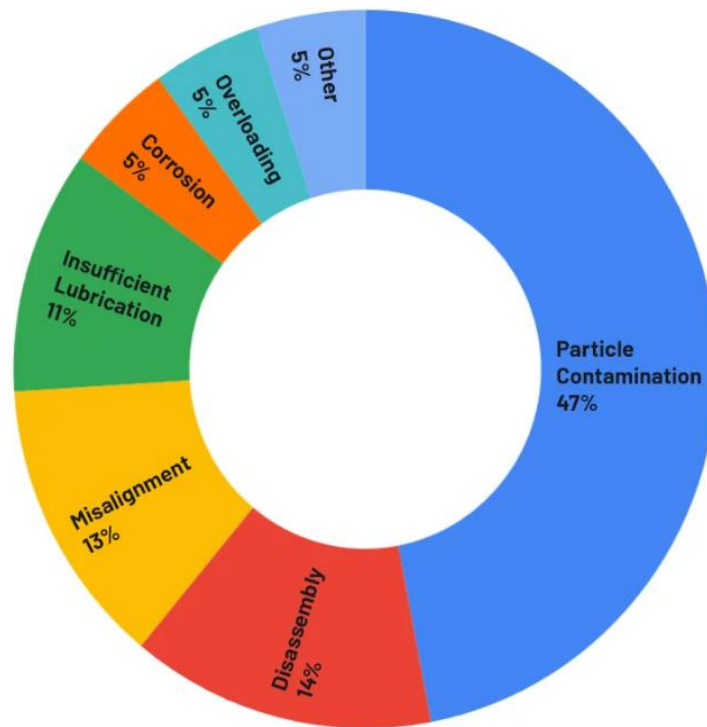


Figure 3.4. Causes of bearing failures

Moreover, the following faults are commonly associated with bearings in rotating machinery:

- a. **Inner Race Fault:** This type of fault occurs in the inner race of the bearing. In vibration analysis, it is identified by the presence of several harmonic peaks of the inner race failing frequency in the frequency spectrum. This fault is usually characterized by specific vibration patterns and can be detected through techniques such as spectral identification.
- b. **Outer Race Fault:** An outer race fault occurs in the outer race of the bearing. In vibration analysis, it is characterized by the presence of harmonic peaks of the outer race failing frequency in the frequency spectrum. Similar to inner race faults, this fault exhibits specific vibration patterns that can be identified through spectral analysis.
- c. **Ball Fault:** This type of fault pertains to the rolling elements, such as balls or rollers, within the bearing. In vibration analysis, ball faults are identified by the presence of harmonics of the rolling element in the frequency spectrum. The specific vibration patterns associated with ball faults can be used for detection and diagnosis. [21]

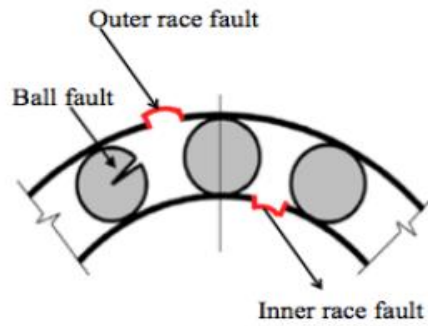


Figure 3.5. Bearing faults

The ball bearings themselves act as a source of vibration, even if there are no defects present and they are perfectly aligned and adjusted. A defect on one of the elements of a ball bearing can cause the vibration level to increase. There are several types of defects that can occur on a ball bearing, such as cracks or pits on rotating surface or rolling elements, distributed defects such as roughness or misaligned races.

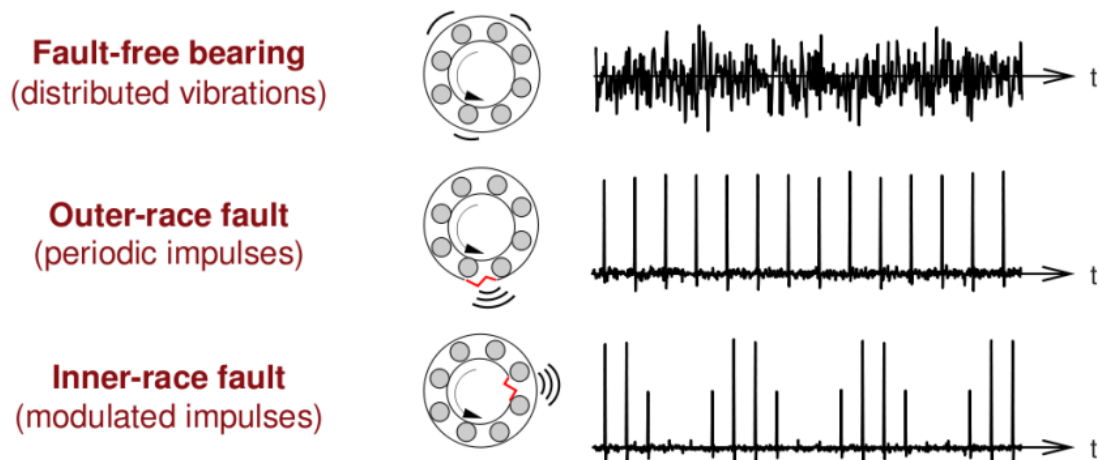


Figure 3.6. The vibration patterns of possible defects

Those distributed or localized defects form the vibration pattern that can be detected by a transducer and then analyzed and processed with the algorithm, which can enable the condition monitoring system to detect even the occurrence of a failure before it damages the machine or interrupt the production. [22]

3.2.2. Shafts

A shaft fault in a rotary machine refers to any issue that affects the rotating shaft, which can lead to excessive vibration and potential failure of the machinery. Bent shaft is the most common fault in rotating machinery, and it generates excessive vibration depending on the amount and location of the bend. [23] Other types of shaft faults include:

- a. **Bow:** A temporary shaft fault caused by the rotor gravitational sag, which can be corrected by adjusting the rotor position. [24]
- b. **Coupling Fault:** A fault in the shaft coupling, which can lead to vibration and potential failure of the machinery.
- c. **Foundation Fault:** A fault in the foundation supporting the rotary machine, which can cause vibration and potential failure of the machinery. [25]

Shaft faults can be diagnosed using various techniques, such as vibration analysis, order analysis, and machine learning algorithms. These techniques help identify the type and location of the fault, enabling proactive maintenance and minimizing the risk of unplanned downtime and costly repairs. By addressing shaft faults, maintenance teams can enhance the reliability and performance of the rotary machinery. [25]

3.2.3. Gears

Gear faults refer to abnormalities or defects that occur within gear mechanisms, compromising their functionality and reliability. These faults can arise due to various factors such as improper meshing, wear over time, or material fatigue. Three common types of gear faults include:

- a. **Chipped Teeth:** This fault occurs when one or more teeth of the gear exhibit chips or breaks, leading to irregular motion and potential performance issues.
- b. **Missing Teeth:** When teeth are entirely absent from the gear, it results in missing tooth faults, which can cause erratic operation and reduced transmission efficiency.

- c. **Root Cracks:** Root cracks develop at the base or root of the gear tooth, weakening the structural integrity of the gear and posing a significant risk of catastrophic failure.



Figure 3.7. (a)Chipped tooth (b)Missing tooth (c)Root crack

These gear faults can manifest as visible damage to the gear teeth and may lead to variations in vibration signals, which are often monitored using accelerometers for early detection and mitigation. [26]

3.2.4. Imbalance

Imbalance occurs when the mass distribution around a machine's rotating axis is not uniform. This can lead to excessive vibration and potential damage to the machine and its components. It is often identified by high vibration amplitudes at the running speed of the machine and its harmonics. [27]

3.2.5. Misalignment

Misalignment occurs when the shaft centerlines of the driver and driven machine(s) are not properly aligned. This can result from improper installation or wear over time. Misalignment can cause vibration, increased bearing loads, and potential damage to the machinery. It is commonly detected through vibration analysis and inspection of coupling and foundation conditions. There are two basic types of misalignments: [28]

- Parallel misalignment.
- Angular misalignment.

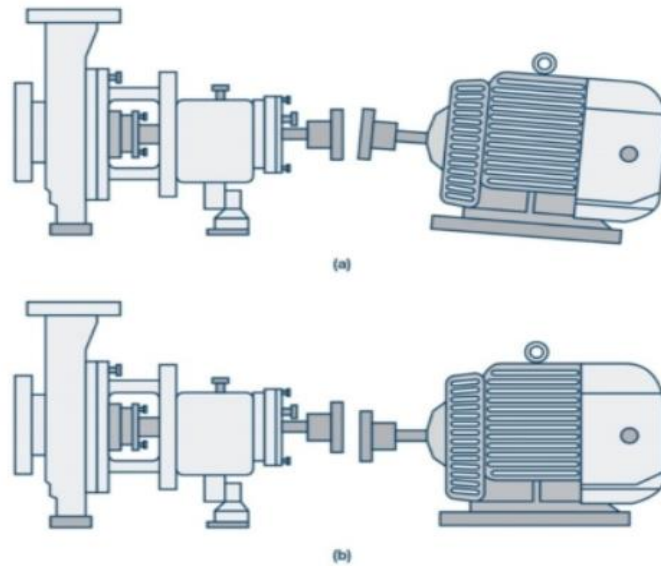


Figure 3.8. (a)angular misalignment(b)parallel misalignment

3.2.6. Looseness

Looseness refers to excessive movement or play in machine components, such as bearings, couplings, or foundation. It can lead to vibration, noise, and wear in the machine. Looseness is often identified through vibration analysis, visual inspection, and monitoring of machine conditions.

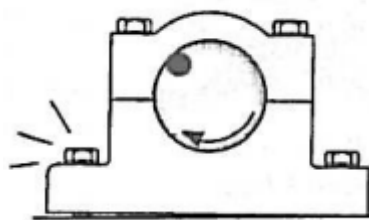


Figure 3.9. Mechanical looseness

3.2.7. Rubbing

Rubbing occurs when two machine components come into contact due to factors such as misalignment, bent shaft, or bearing defects. This can lead to increased

vibration, heat generation, and potential damage to the components. Rubbing is typically detected through vibration analysis and inspection of machine components. [29]

In addition to the mechanical faults discussed above, the integration of advanced signal processing techniques plays a pivotal role in enhancing fault diagnosis and condition monitoring in rotating machinery.

3.3. Signal Processing

Signal processing methods are essential tools in fault detection and diagnosis, particularly in the realm of electromechanical systems. Advanced techniques, such as Discrete Wavelet and Wavelet Packet Transforms, are employed to analyze and process vibration signals, enabling the detection and classification of occurring faults. This focus on signal processing is integral in enhancing the accuracy and efficiency of fault detection methodologies. [30]

These methods contribute to the extraction of valuable information from vibration signals, aiding in the identification of faults and potential failures. The integration of signal processing in fault diagnosis not only enhances the understanding of system health but also contributes to preventive maintenance strategies. [31] Signal processing methods, coupled with machine learning techniques, introduce intelligent fault diagnosis approaches. These methods focus on extracting fault features from signals, further contributing to the development of effective fault detection strategies. The combination of signal processing and machine learning showcases a synergistic approach to fault diagnosis, ensuring the reliability and efficiency of the diagnostic process. [32]

The condition monitoring system involves the signal processing techniques. There are different types of faults and processing these signals is crucial. The selection of the appropriate technique depends on the nature of the captured signal. These techniques include:

- i. ***Time domain analysis:*** Time domain analysis involves the analysis of physical signals or time series of data, with respect to time. A time-domain graph is used for visualizing the change in a signal with respect to time. Time domain analysis adopts statistical analysis concept.

- ii. **Frequency domain analysis:** Frequency domain analysis involves the analysis of physical signals or time series of data, with respect to frequency. It shows the number of signals lying within the given frequency band over a range of frequencies.
- iii. **Time Frequency domain analysis:** It comprises the techniques used in both the time and frequency domains simultaneously. This analysis involves the study of two-dimensional signals. [33]

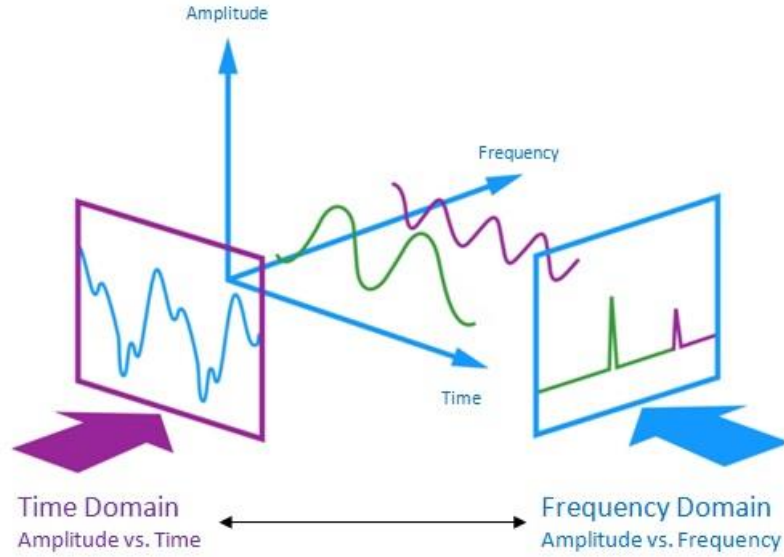


Figure 3.10. Time and frequency domain representations of a signal

The used techniques in our work:

3.3.1. Maximal overlap discrete wavelet transforms (*modwpt*)

MODWPT is a wavelet-based time-frequency analysis method. It decomposes a signal into wavelet packets, providing insights into both time and frequency domains.

Given j , n , t are the decomposition level, oscillation index, and time index, respectively, the MODWPT is given by:

$$W(j, n, t) = \sum_l u(n, l) * W(j - 1, \lfloor n/2 \rfloor, t - 2^{(j-1)} * l \text{ mod } N)$$

The variable L is the length of the filters, where g and h are the scaling filter and wavelet filter, respectively.

By definition

$$W(0,0,t) = X(t)$$

Were X being the original time series. [34]

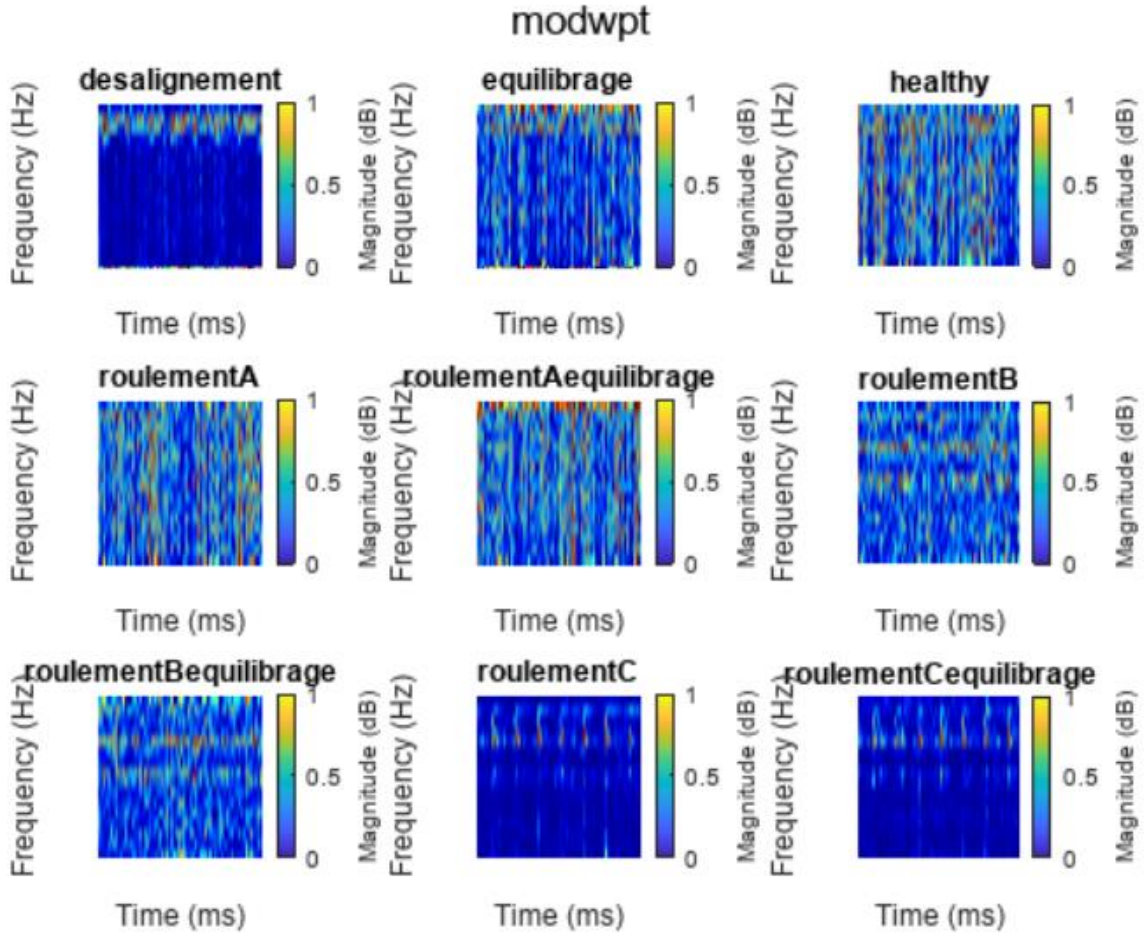


Figure 3.11. The modwpt

3.3.2. Continuous 1-D wavelet transforms (cwt)

The CWT of a signal $x(t)$ can be performed through a convolution operation between the signal $x(t)$ and complex conjugate of a family of wavelets.

which is expressed as

$$cwt(s, \tau) = \frac{1}{\sqrt{s}} \int x(t) \psi^*\left(\frac{t - \tau}{s}\right) dt$$

Where s and τ denotes the scale and translation parameters, respectively. $\psi^*(\blacksquare)$ is the complex conjugate of the scaled and shifted wavelet function $\psi(\blacksquare)$. [35]

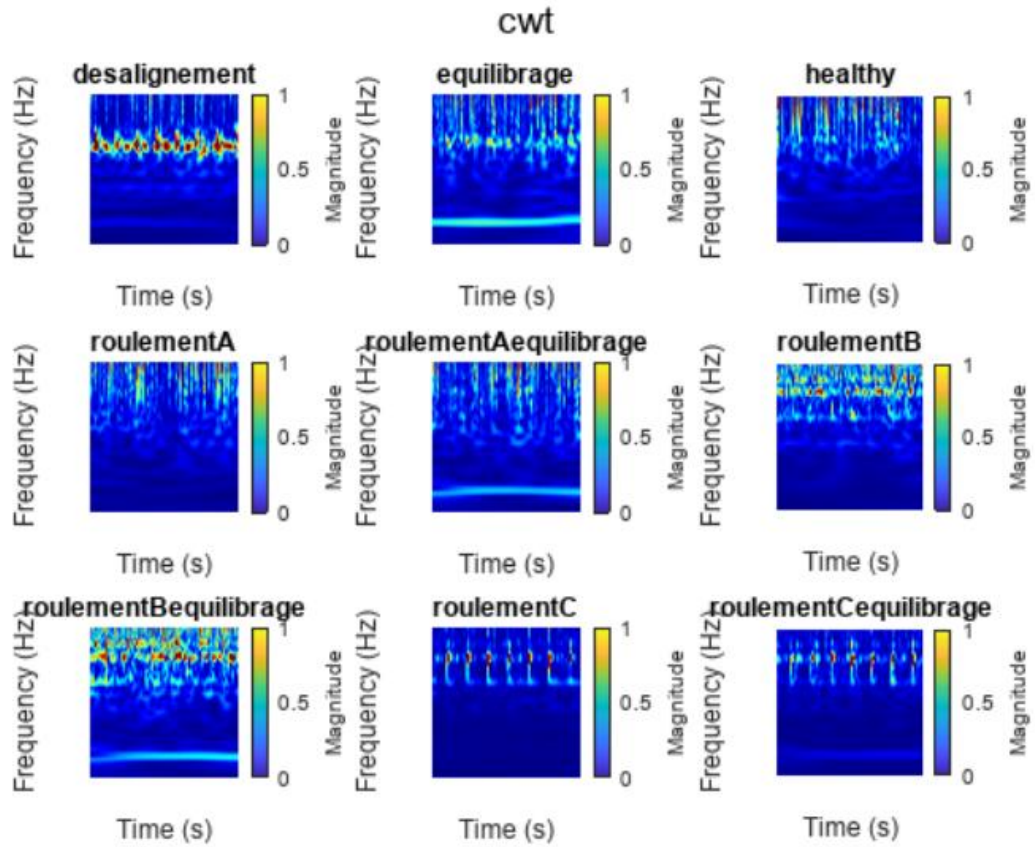


Figure 3.12. The cwt

3.3.3. Short-time Fourier transforms(stft)

The STFT is a time-frequency analysis method that represents how the frequency content of a signal changes over time. It is useful for analyzing non-stationary signals.

It is given by:

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x(m)w(m - n)e^{-j\omega m}$$

Where $x(m)$ is the input signal, $w(m)$ is the window function, n is the time index, ω is the frequency index, and j is the imaginary unit. [36]

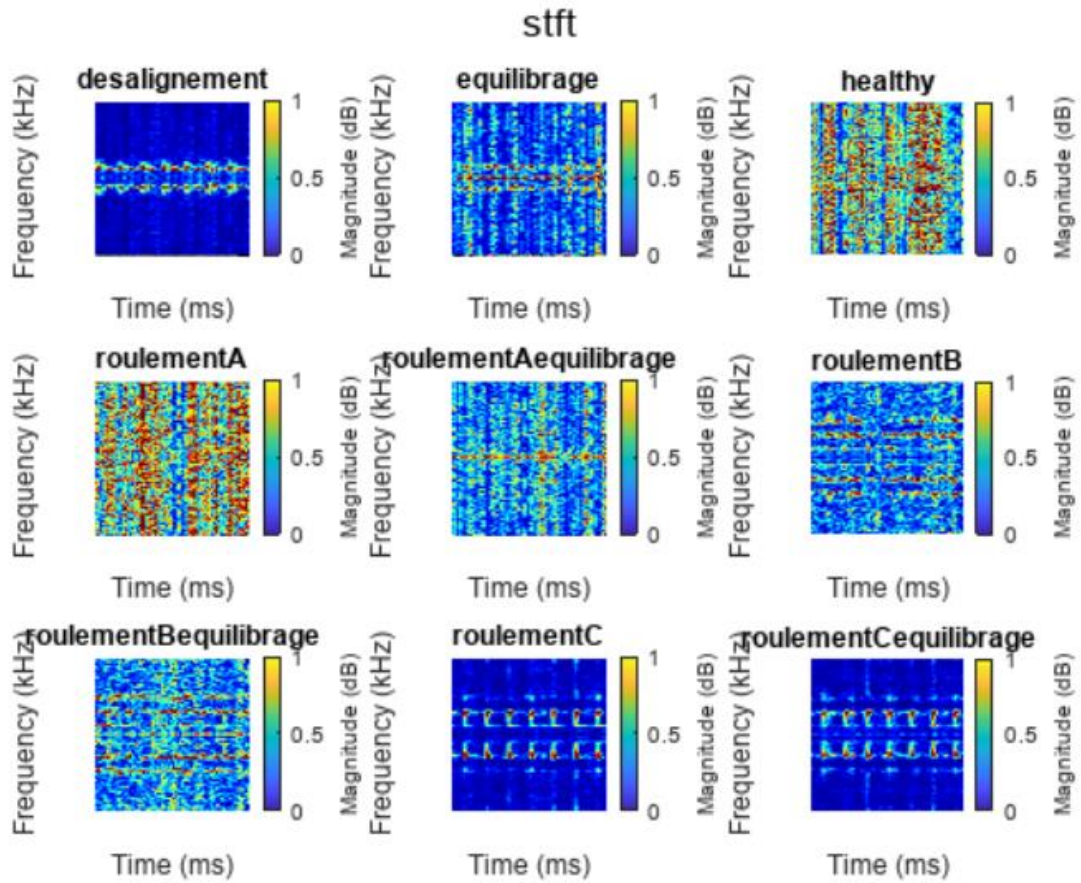


Figure 3.13. The stft

3.3.4. Cross spectrogram using short-time Fourier transforms(xspectrogram)

The xspectrogram is a time-frequency analysis method that is used to analyse the relationship between two signals. It is a powerful tool for detecting and analysing the frequency content of non-stationary signals. It is given by:

$$C_{xy}(n, \omega) = \sum_{m=-\infty}^{\infty} x(m)y(m-n)w(m-n)e^{-j\omega m}$$

Where $x(m)$ and $y(m)$ are the input signals, $w(m-n)$ is the window function, n is the time index, ω is the frequency index, and j is the imaginary unit. [37]

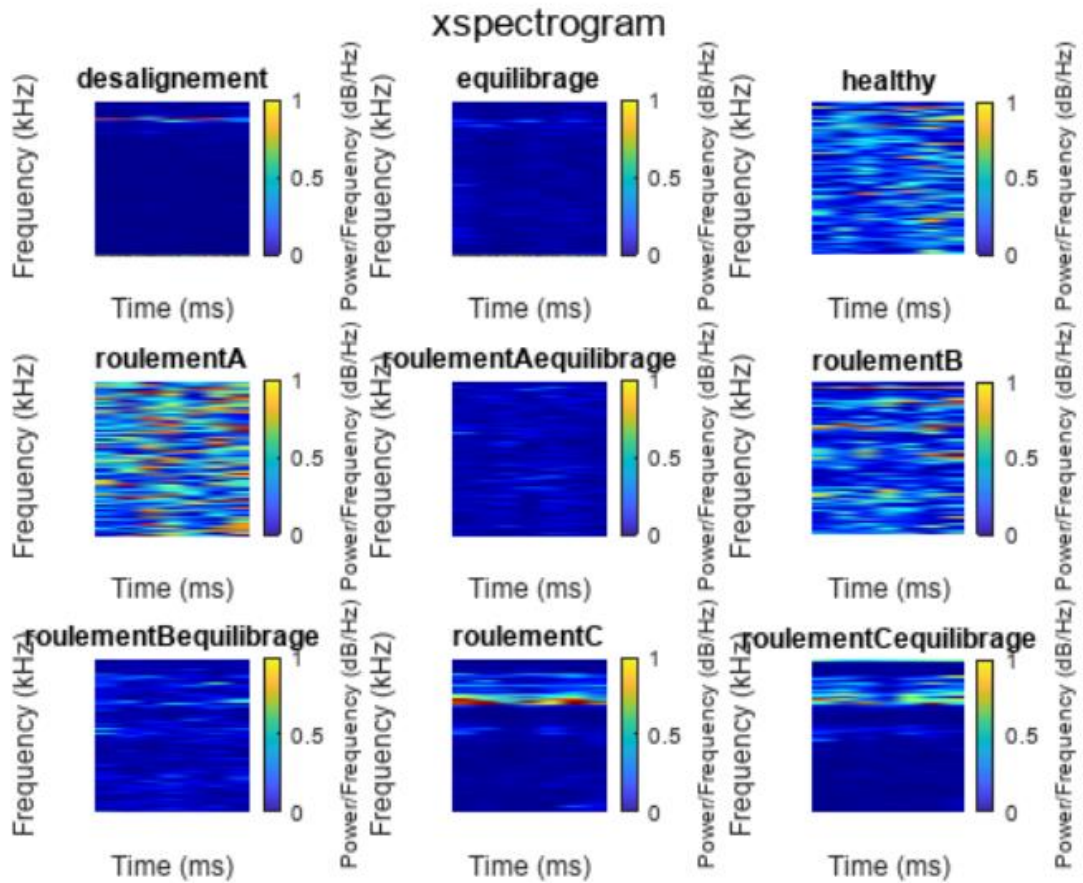


Figure 3.14. The xspectrogram

3.3.5. Power spectrum(*pspectrum*)

The *pspectrum* function in MATLAB is a powerful tool for computing the power spectrum of a signal, which can be used to analyse signals in the time domain, frequency domain, or time-frequency domain.

The equation for the power spectrum is given by:

$$P(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 e^{-j\omega t} dt$$

where $x(t)$ is the input signal, ω is the frequency index, and j is the imaginary unit. [38]

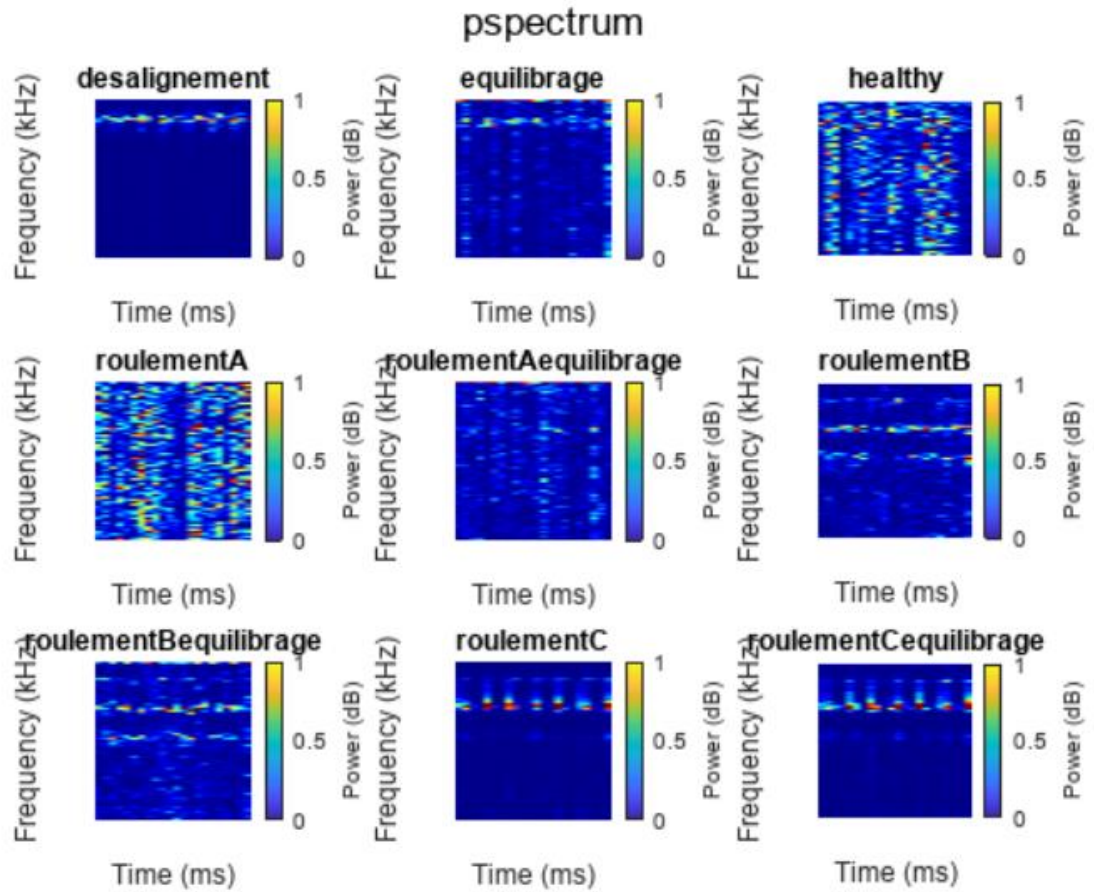


Figure 3.15. The Pspectrum

3.3.6. Visualize spectral kurtosis(kurtogram)

The kurtogram is a signal processing technique that is used for fault diagnosis of rotating machinery. It is a powerful tool for detecting faults in bearings, gears, and other rotating machinery. The kurtogram is based on the spectral kurtosis, which is a measure of the non-Gaussianity of a signal. The kurtogram is obtained by calculating the spectral kurtosis of a signal at different frequency bands and plotting the results as a function of frequency. The kurtogram can be used to identify the frequency band that contains the maximum amount of fault information, which can then be used for further analysis.

The equation for the spectral kurtosis is given by:

$$K(\omega) = \frac{\sum_{i=1}^N f_i^4 |X_i(\omega)|^2}{(\sum_{i=1}^N f_i^2 |X_i(\omega)|^2)^2} - 2$$

where $X_i(\omega)$ is the Fourier transform of the i frequency band, f_i is the center frequency of the i frequency band, N is the number of frequency bands, and ω is the frequency index. [39]

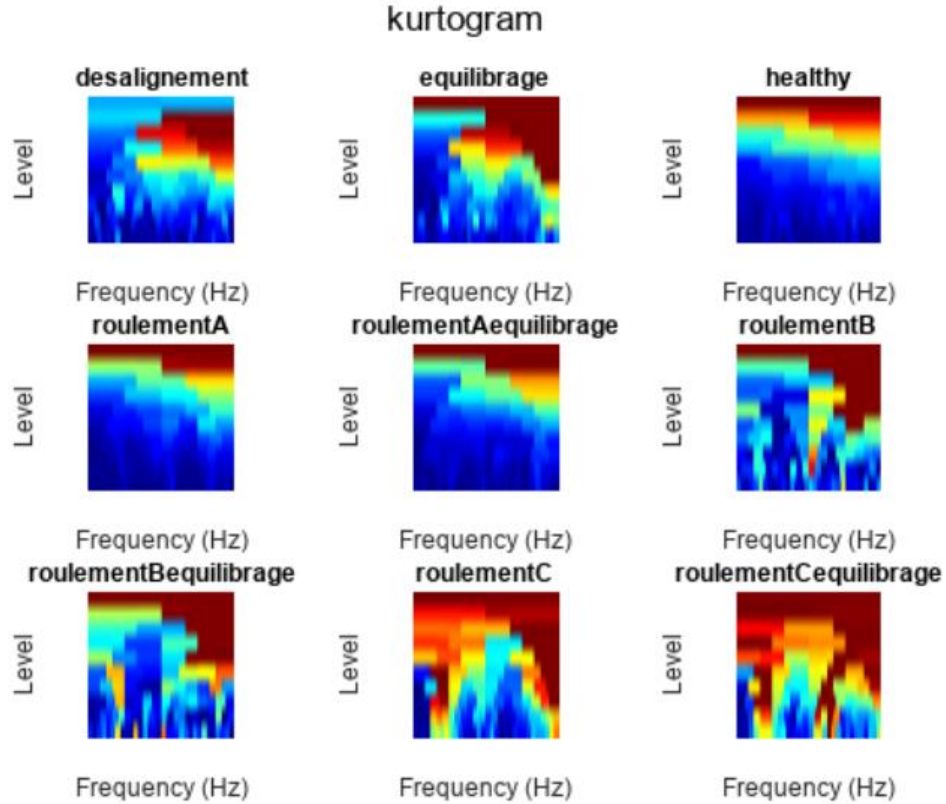


Figure 3.16. The Kurtogram

3.3.7. The instantaneous frequency (*instfreq*)

Instantaneous frequency refers to the frequency content of a signal at a specific point in time. This function is particularly useful for analyzing signals with time-varying frequency components, such as non-stationary signals.

The equation to estimate the instantaneous frequency $f(t)$ of a signal $x(t)$ sampled at a rate f_s is:

$$f(t) = \frac{1}{2\pi} \frac{d}{dt} \phi(t)$$

Where $f(t)$ is the instantaneous frequency at time t , $x(t)$ is the signal, f_s is the sample rate, $\phi(t)$ is the instantaneous phase of the signal. The instantaneous phase $\phi(t)$ can be

computed using various time-frequency analysis techniques such as the Short-Time Fourier Transform (STFT) or the Hilbert Transform. [40]

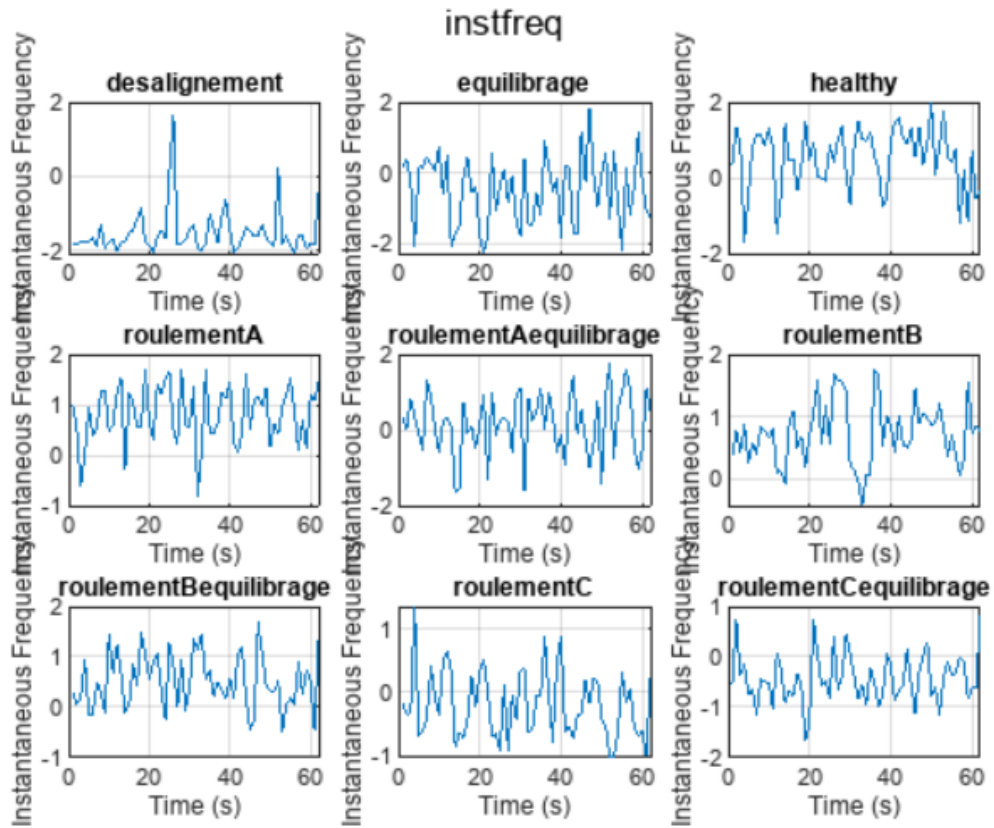


Figure 3.17. The instfreq

3.3.8. The Spectral kurtosis (Pkurtosis)

Pkurtosis is a measure used in signal processing to quantify the non-Gaussianity and sharpness of spectral peaks in the frequency domain. It assesses how much a signal's frequency spectrum deviates from a Gaussian distribution, often indicating the presence of transient events or non-Gaussian noise. The equation for calculating Pkurtosis is typically based on the kurtosis of the signal's power spectrum.

The equation for Pkurtosis can be represented as:

$$P_k(f) = \frac{\frac{1}{N} \sum_{n=0}^{N-1} |X(n, f)|^4}{\left(\frac{1}{N} \sum_{n=0}^{N-1} |X(n, f)|^2\right)^2} - 2$$

where $P_k(f)$ is the spectral kurtosis at frequency f , N is the length of the signal, $X(n, f)$ is the Fourier transform of the signal at time n and frequency f . [41]

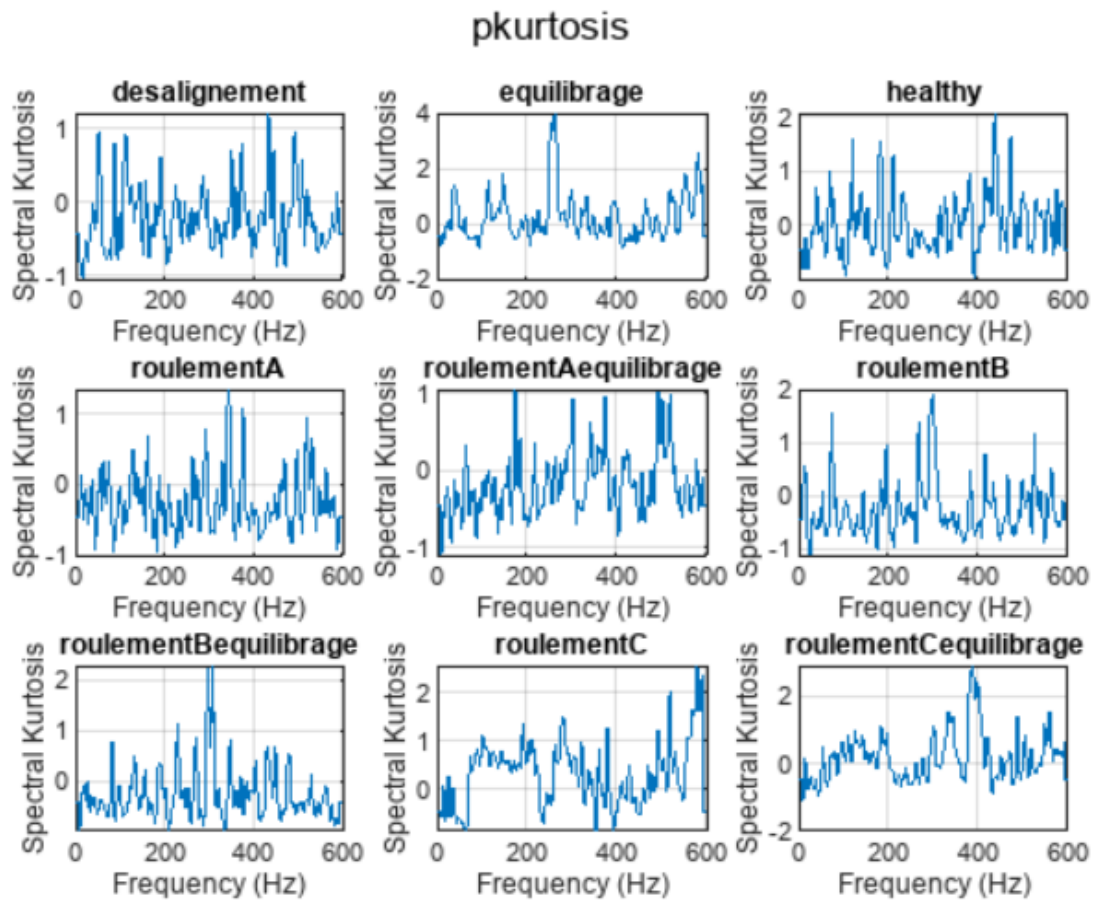


Figure 3.18. The Pkurtosis

3.4. Experimental testing bed

We conducted our experiments using the experimental testing bed at the **IAP SKIKDA PSE** mechanical engineering laboratory located in Filfila SKIKDA, Algeria. The institute is associated with the **Algerian Petroleum Institute (IAP)**, which is a national graduate institute of petroleum engineering located in Boumerdès, Algeria. The Physical Simulation of Rotating Machine Faults (PT 500) is a modular training system that is used to simulate, measure, and evaluate vibration signals of typical malfunctions and damage in rotating machinery. The PT 500 system provides a flexible and modular learning platform offering an introduction to the complex and demanding subject of machine diagnosis. The target group is students in mechanical

engineering/machine dynamics, but it also includes other professional people as part of on-the-job training and development in maintenance and servicing.

The PT 500 system can be used to simulate a wide range of faults in rotating machinery, including unbalance, misalignment, bearing faults, gear faults, and other types of faults. The system can also be used to evaluate the effectiveness of different signal processing techniques for fault diagnosis, such as envelope analysis, wavelet transform, Hilbert transform, and acoustic emission.

The PT 500 system plays an important role in fault detection and predictive maintenance of rotating machinery. By simulating different types of faults, the system can be used to train students and professionals in the detection and diagnosis of faults in rotating machinery. The system can also be used to evaluate the effectiveness of different signal processing techniques for fault diagnosis, which can help in the development of new techniques for fault detection and predictive maintenance.

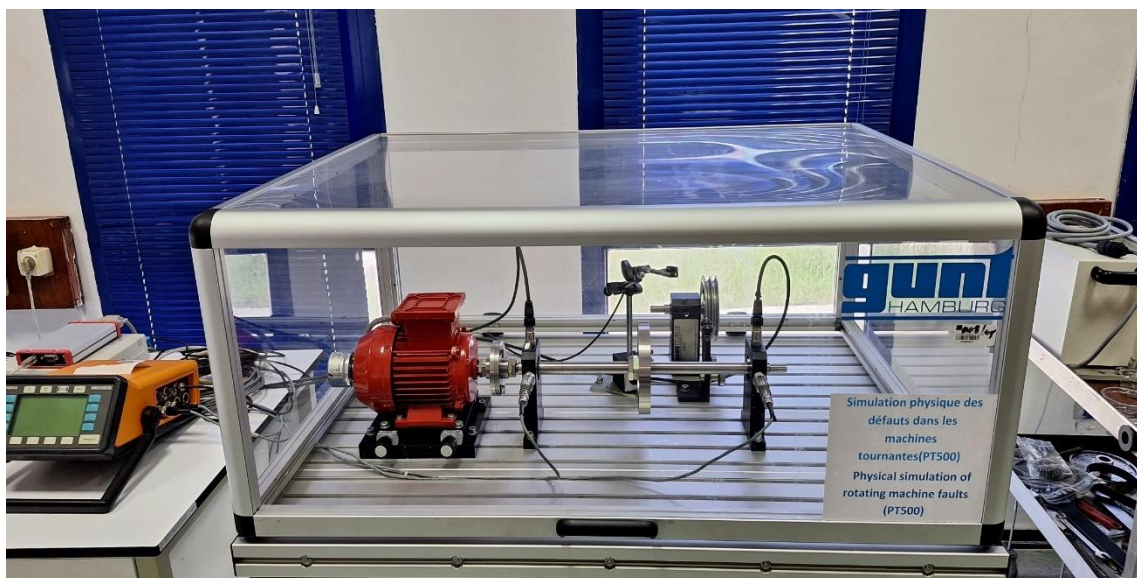


Figure 3.19. The PT 500

The PT 500 was used to simulate the faults, and a database was collected to use in our research. The acquisition of the data was done by two accelerometers. The accelerometer A reads the horizontal accelerations, while the accelerometer B reads the vertical accelerations. The sample rate used was 8kHz.

3.4.1. Simulated faults

3.4.1.1. Bearing Faults

In-depth analysis of bearing faults involved simulating various types, each representing distinct damage scenarios. This included outer and inner race damage, damaged retainer, a combination of these faults, and a bearing with high usage. These simulations were performed across multiple speeds (1500, 1800, 2400, 3000 RPM), providing a comprehensive dataset for studying the impact of bearing faults on vibration signals.

The simulated type of bearing faults:

- Type A: Undamaged bearing.
- Type B: Bearing with outer race damaged.
- Type C: Bearing with inner race damaged.
- Type D: Bearing with damaged roller body.
- Type E: Bearing with damaged roller body, outer and inner race.
- Type F: Heavily worn bearing.

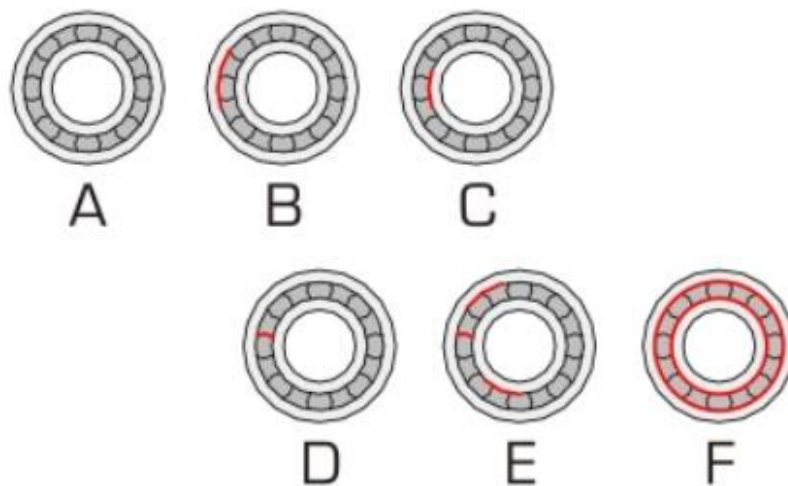


Figure 3.20. The simulated bearing faults

3.4.1.2. Misalignment

The testing bed allowed the simulation of misalignment, a common issue in rotating machinery. This fault's effects on vibration patterns were examined meticulously to enhance our understanding of misalignment-related dynamics.

3.4.1.3. Imbalance

A prevalent fault leading to vibration issues, was simulated at different magnitudes (0g, 13.6g, 29.5g) across varying speeds. This systematic approach facilitated a detailed investigation into the relationship between imbalance severity, rotational speed, and resultant vibration characteristics.

Table 3.1. Experiments details

Motor condition		The speed of motor (rpm)
Single fault (separated)	Misalignment	1500, 1800, 2400, 3000
	Imbalance 13.6g	1500, 1800, 2400, 3000
	Imbalance 29.5g	1500, 1800, 2400, 3000
	Bearing type B	1500, 1800, 2400, 3000
	Bearing type C	1500, 1800, 2400, 3000
	Bearing type D	1500, 1800, 2400, 3000
	Bearing type F	1500, 1800, 2400, 3000
Combined faults (cumulated)	Bearing type E	1500, 1800, 2400, 3000
	Bearing type B + Imbalance 13.6g	1500, 1800, 2400, 3000
	Bearing type B + Imbalance 29.5g	1500, 1800, 2400, 3000
	Bearing type C + Imbalance 13.6g	1500, 1800, 2400, 3000
	Bearing type C + Imbalance 29.5g	1500, 1800, 2400, 3000
	Bearing type D + Imbalance 13.6g	1500, 1800, 2400, 3000
	Bearing type D + Imbalance 29.5g	1500, 1800, 2400, 3000
Healthy	Bearing type A	1500, 1800, 2400, 3000

In conclusion, the PT500 Experimental Testing Bed serves as an invaluable resource for researchers and practitioners, offering a sophisticated environment for replicating and analysing complex fault scenarios in rotating machinery.

4. Implementation

4.1. Device Designing

Device design plays a pivotal role in the effectiveness of fault diagnosis systems for rotary machines. In this section, we delve into the intricacies of our device design process and the thoughtful decisions guiding its construction.

The device consists of an MPU-9250 accelerometer that reads the accelerations from the rotary machine and sends it to an ESP-32 microcontroller. The ESP-32 saves the accelerations collected by the MPU-9250 and sends them to the PC to perform the signal processing and predictions. The ESP-32 is powered by a battery and is connected to the PC via Wi-Fi. The assembly of the device is enclosed in a 3D-printed box.

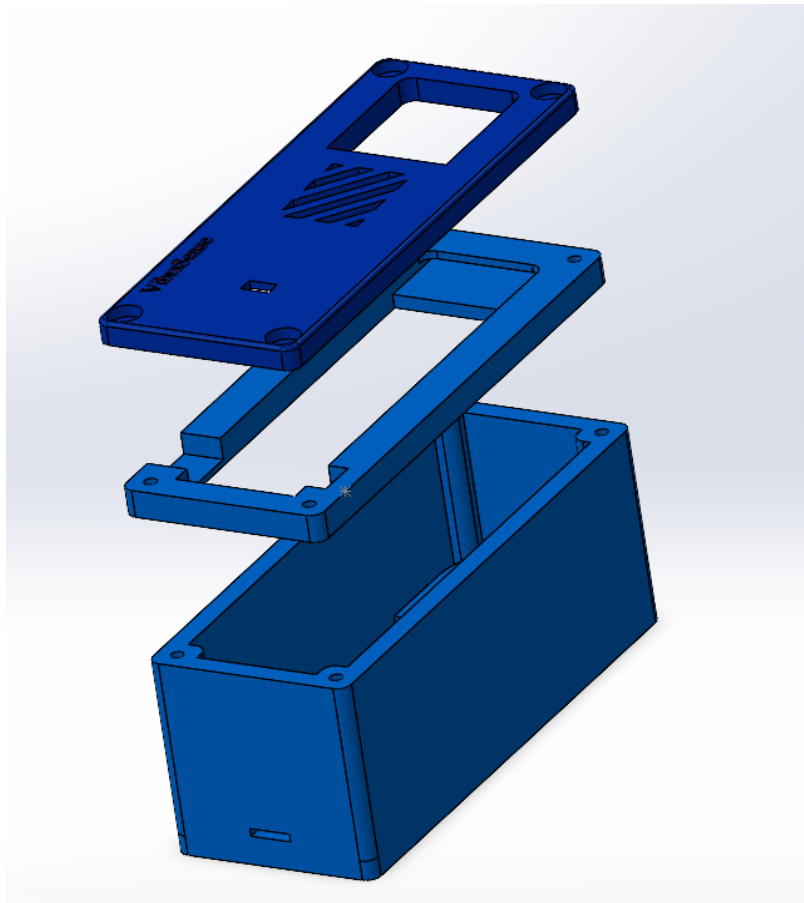


Figure 4.1. The 3-D printed box

We used the MPU-9250 accelerometer because it is a reliable and accurate sensor that can measure the vibrations of the rotary machine. The MPU-9250 The MPU-9250 is the world's smallest 9-axis Motion Tracking device and incorporates the

latest InvenSense design innovations, enabling dramatically reduced chip size and power consumption, while at the same time improving performance and cost. [42]

Technical data sheet of the MPU-9250:

- Manufacturer: InvenSense.
- Model: MPU-9250.
- Dimensions : 3mm x 3mm x 1mm.
- Type: Accelerometer, gyroscope, magnetometer.
- Key Features: Low power, high precision, integrated gyroscope and magnetometer.
- Measurement Range: $\pm 2g$, $\pm 4g$, $\pm 8g$, $\pm 16g$.
- Resolution: 16 bits.
- Output Data Rate (ODR): Adjustable, up to 1 kHz.
- Communication Protocol: I2C, SPI.

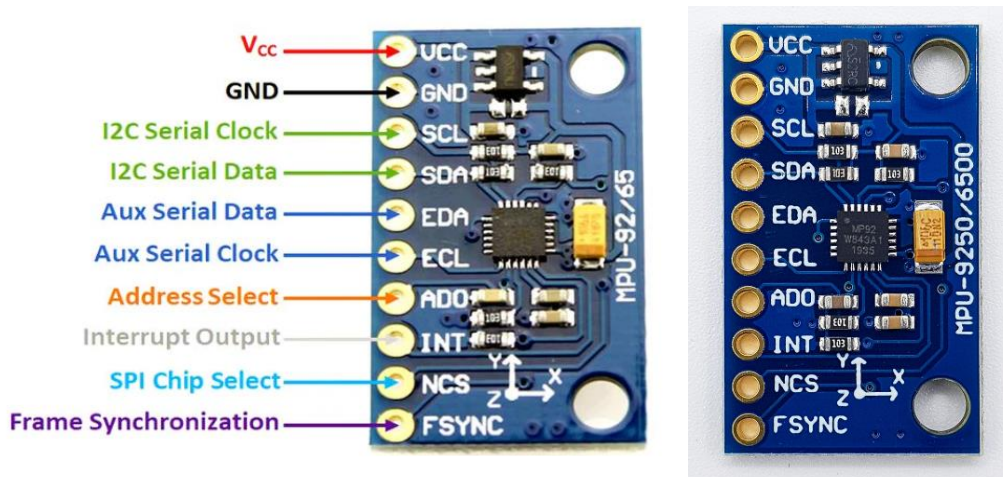


Figure 4.2. The MPU-9250

We also chose to use the ESP-32 microcontroller because it has a powerful processor that can perform the signal processing required for fault diagnosis.

The ESP32 is a low-cost, low-power system on a chip microcontroller with integrated Wi-Fi and dual-mode Bluetooth. It is designed for mobile, wearable electronics, and Internet-of-Things (IoT) applications. The ESP-32 is capable of functioning reliably in industrial environments, with an operating temperature ranging from -40°C to $+125^{\circ}\text{C}$ [3]. It is powered by advanced calibration circuitries, which can dynamically remove external circuit imperfections and adapt to changes in external

conditions 3. The ESP-32 is created and developed by Espressif Systems, a Chinese company based in Shanghai, and is manufactured by TSMC using their 40 nm. [43]

Technical data sheet of the ESP-32:

- Manufacturer: Espressif Systems.
- Model: ESP-32.
- Type: Microcontroller.
- Processor: Dual-core Tensilica LX6.
- Clock Frequency: Up to 240 MHz
- Memory: 520 KB SRAM, 4 MB Flash.
- Wireless Connectivity: Wi-Fi, Bluetooth.
- Integrated Peripherals: GPIO, I2C, I2S, UART, etc.
- AI Processing Capability: Yes.
- Power Supply: 3.3V.
- Communication Interfaces: Wi-Fi for wireless communication.

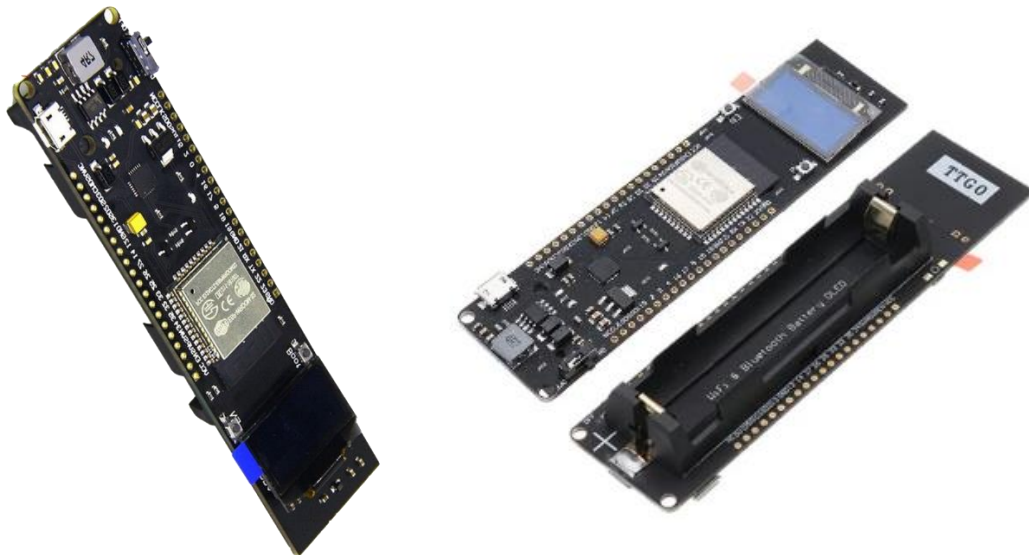


Figure 4.3. The ESP-32

During the design process, we faced several challenges. One of the challenges was to ensure that the device was mounted properly on the rotary machine and that all the vibrations of the rotary machine were detected properly by the device. We overcame

this challenge by placing a magnet in the bottom of the 3D-printed box to mount it properly.

4.1.1. Arduino and sensor

4.1.1.1. Arduino

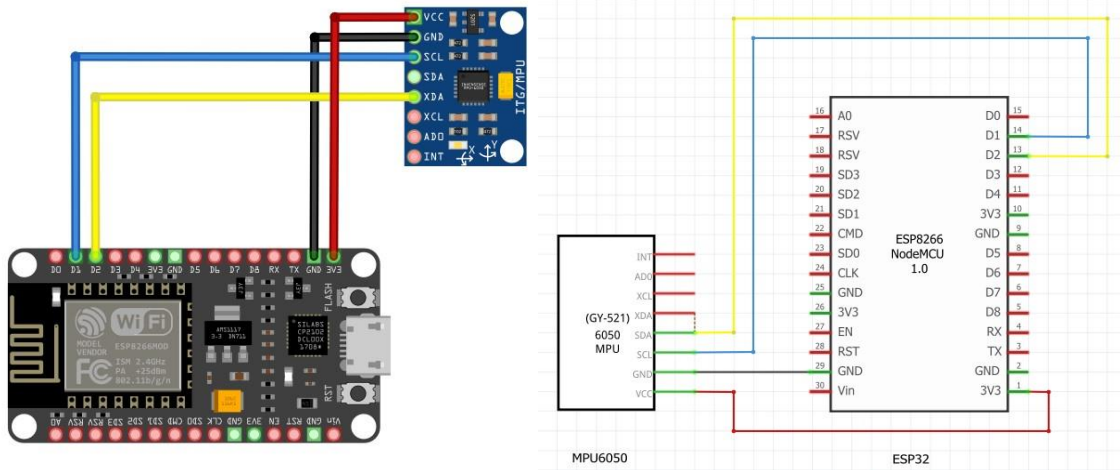


Figure 4.4. (a)Physical (b)electrical Connection

4.1.1.2. Role and programming

The Arduino is an open-source development board used to turn ideas into existing electronic devices. In VibraSens, Arduino plays an essential role in analyzing vibration data and making decisions. Arduino is programmed using the C++ programming language and can be connected to various components.

4.1.1.3. Sensor connection

To achieve efficient communication between the Arduino and the MPU9250 accelerometer sensor, an I2C interface is used. This connection is essential for transferring data between the two devices. This allows us to use the Wire.h library for programming, which provides simplified I2C module control functions.

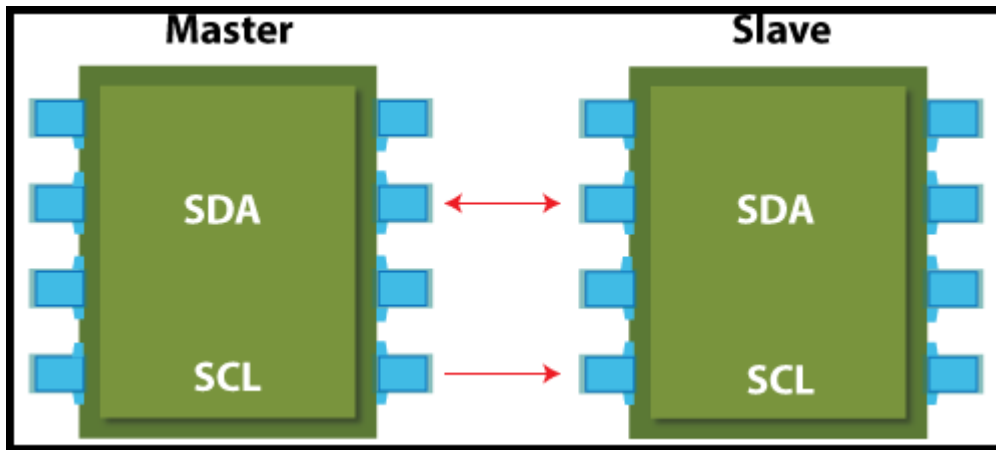


Figure 4.5. I2C-protocol

Through the code, we set the accelerometer sensor address to 0x68, and read the accelerometer data continuously. The mini code below shows how this can be achieved:

```
#include <Wire.h>

#define MPU9250_ADDRESS 0x68

#define ACC_FULL_SCALE_16_G 0x18

// Read acceleration data via I2C interface

void I2Cread(uint8_t Address, uint8_t Register, uint8_t Nbytes, uint8_t*
Data)
{
    Wire.beginTransmission(Address);

    Wire.write(Register);
}
```

4.1.1.4. Programming

a. Use of libraries

The Wire.h library is used by the developer to control the I2C module and read sensor data, while the WiFi.h library is used to configure the ESP32 as an access point to communicate with the computer.

```
#include <WiFi.h>
#include <WiFiClient.h>
#include <WiFiAP.h>
```

4.1.1.5. Send data

It reads acceleration data from the sensor and sends it to the computer connected via the Wi-Fi interface.

```
WiFiServer server(80);
WiFiClient client;
void setup() {
    Wire.begin();
    Serial.begin(115200);
    WiFi.softAP("ESP32_AP", "012345678");
    server.begin();
}
void loop() {
    if (client.connected()) {
    }
    for (int i = 0; i < bufferSize; ++i) {
        String data = String(azBuffer[i]-16664) + "," +
String(ayBuffer[i]+147);
```

```
    client.println(data);
  }
} else {
  Serial.println("No client connected - try again");
  client = server.available();
}
}
```

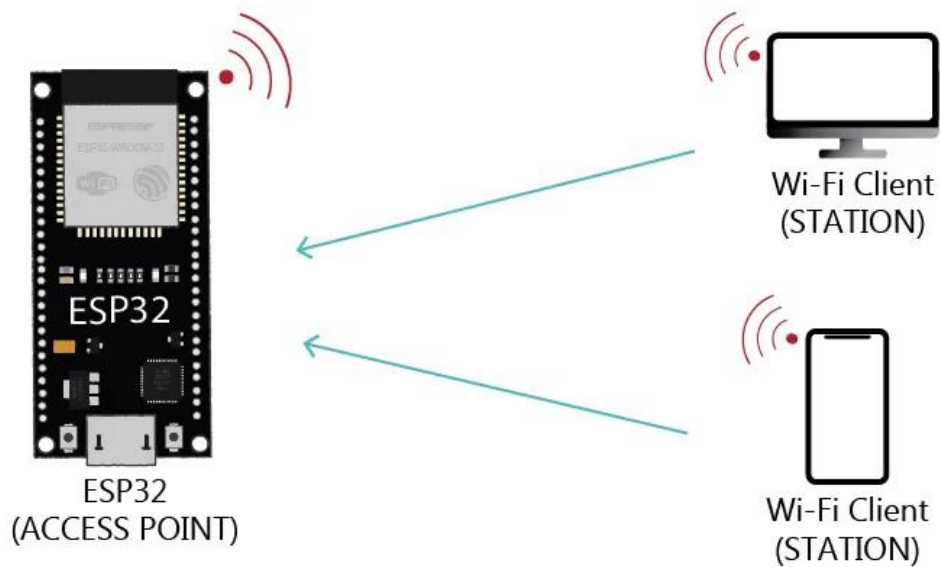


Figure 4.6. ESP-32 Access Point Mode

4.1.1.6. Energy consumption

The power consumption of VibraSens depends on the operating period and frequency of data reading. The project included a tight design architecture to maximize battery power and improve energy efficiency, contributing to improved battery life and sustained performance.

4.1.1.7. Power system

It includes an 18650-charging system and takes advantage of a 3000mAH 18650 LG battery to provide up to 17 hours of playback. It benefits from a built-in charging system and intelligent power control.

4.1.1.8. Anti-discharge protection

Provide anti-discharge protection to ensure the battery is not damaged by over-discharge and extend battery life.

4.1.1.9. Idle mode

These measures are part of a smart strategy to achieve efficient performance with low power consumption, sustainably enhancing the VibraSens experience.

The results of the design process were successful. The MPU-9250 accelerometer was able to accurately measure the vibrations of the rotary machine, and the built deep learning model was able to predict the type of fault if there was a fault in the machine.

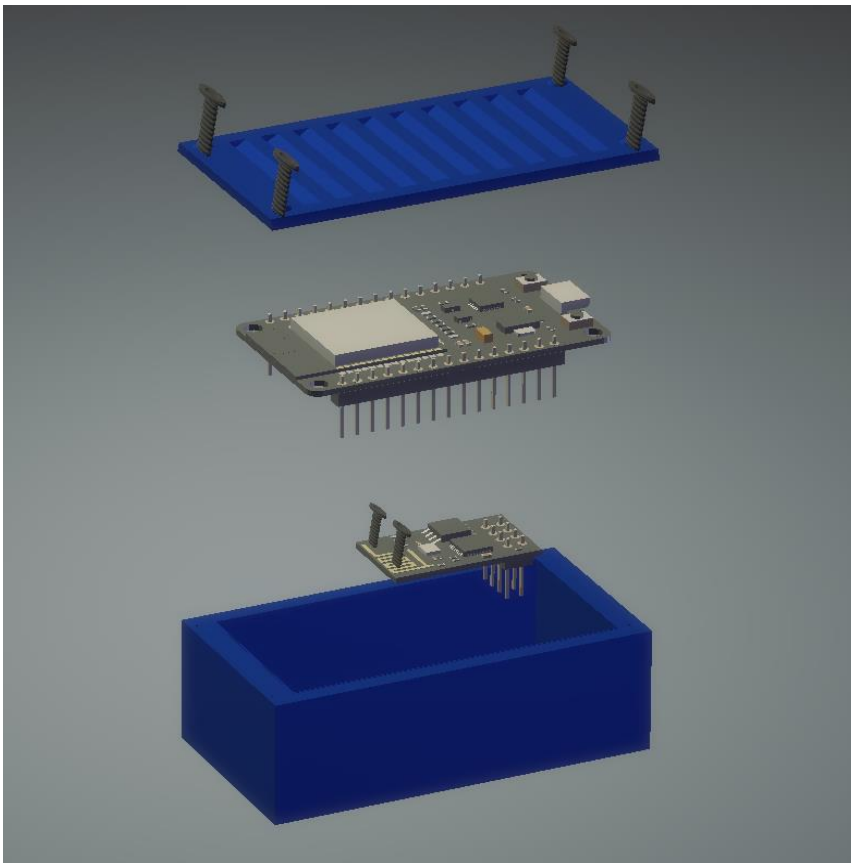


Figure 4.7. The device assembly

In conclusion, the device design is a critical component of the fault diagnosis system for rotary machines. The design process involved several challenges, but we

were able to overcome them and produce a successful design. The device design is an essential step towards ensuring the efficient and effective operation of rotary machines.

4.2. Framework

4.2.1.1. Neural Network in details

Neural networks are a type of machine learning algorithm that are inspired by the structure and function of the human brain. They consist of layers of interconnected nodes that process and learn from data. Neural networks can be used for various tasks, such as classification, regression, clustering, generation, etc. There are many types and variations of neural networks, such as convolutional neural networks, recurrent neural networks, generative adversarial networks, etc. Each type has its own advantages and disadvantages and is suitable for different kinds of problems and data.

In this section, we describe the neural network architecture used in our implementation for rotary machine fault detection.

4.2.1.2. Proposed ResNet structure

Once the raw signals have been converted to images, ResNet can be effectively trained to classify these images. In our research, we have tailored a ResNet-based model specifically for addressing image classification tasks in fault diagnosis. The ResNet architecture is renowned for its versatility and adaptability, allowing for the creation of custom variants with varying numbers of layers, block configurations, and skip connections.

Our custom ResNet-like structure adheres closely to the fundamental principles of ResNet, integrating crucial elements such as residual blocks, batch normalization, and skip connections. However, we have further optimized the architecture by fine-tuning layer configurations and arrangements to suit the specific requirements of fault diagnosis applications.

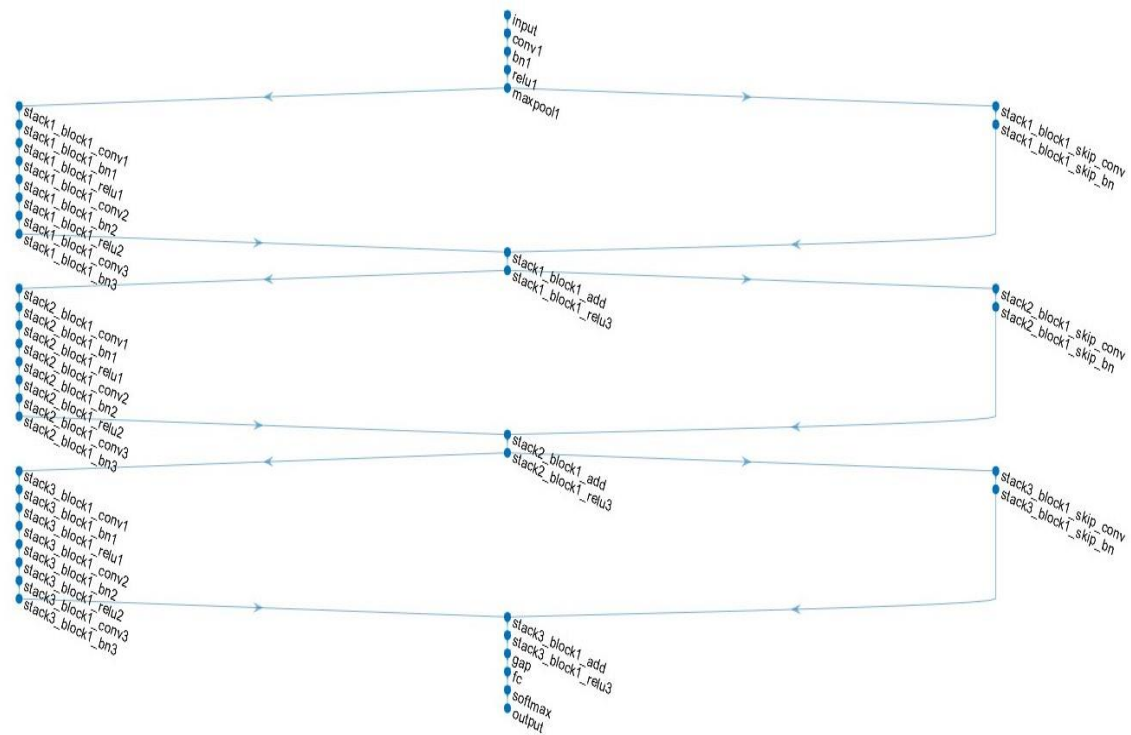


Figure 4.8. The ResNet structure

The model begins with an input layer for 64x64x3 images and consists of several convolutional layers, batch normalization layers, and skip connections, organized into different blocks and stages. The model ends with a global average pooling layer, a fully connected layer, a SoftMax activation, and a classification output layer. The architecture we've provided aligns with the typical structure of a ResNet model, which is characterized by residual blocks and skip connections to enable the training of very deep neural networks. The use of residual blocks and skip connections is a key feature of ResNet, allowing the network to learn the identity function more quickly and prevent the degradation problem as the network depth increases.

We chose these parameters to train our ResNet model.

Table 4.1. Parameters used to train the ResNet model

Parameters	Value
Input sequence size	3x64x64 RGB images
Epochs	20
Mini batch size	64
Shuffle the data	Every epoch
Optimization algorithm	Adam

4.2.1.3. Acquisition interface

VibraSens is a rotary machine fault detection application designed to acquire acceleration from the rotary machine and detect faults using a 1-D Convolutional Neural Network (CNN). The application is user-friendly and intuitive, allowing users to easily set acquisition parameters, plot the accelerations, and predict the type of fault or the status of the rotary machine.

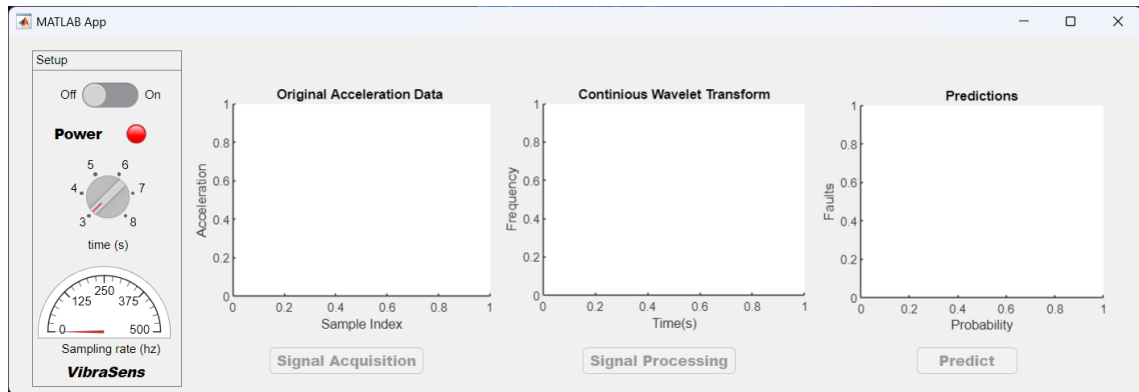


Figure 4.9. The interface of the application

The application consists of a ‘Setup’ panel on the left that allows users to turn the application on or off using the power switch.

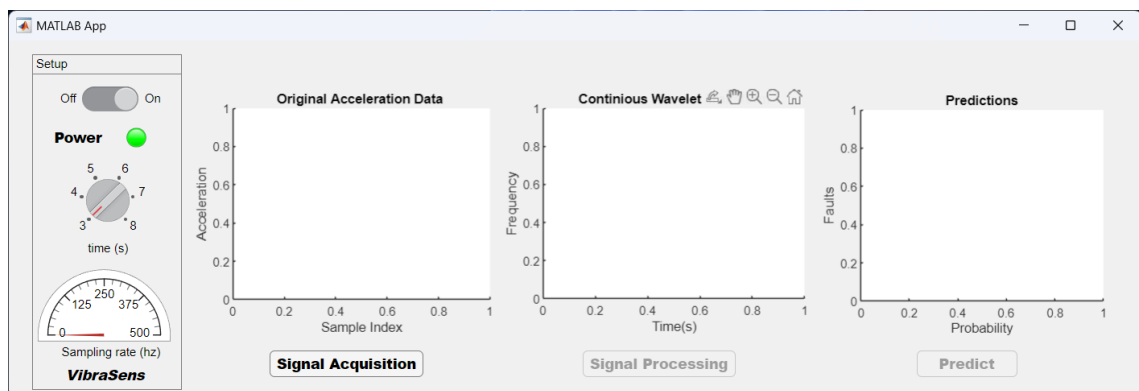


Figure 4.10 Indication of successful connection

Once the application is turned on, a lamp will change color to green to indicate that the application is working properly and the connection between the application and the device is good. The ‘Signal Acquisition’ button will be enabled, and the user can set the acquisition time using the time knob. After the user has set the acquisition

parameters, they can click the ‘Signal Acquisition’ button to read the accelerations of the rotary machine.

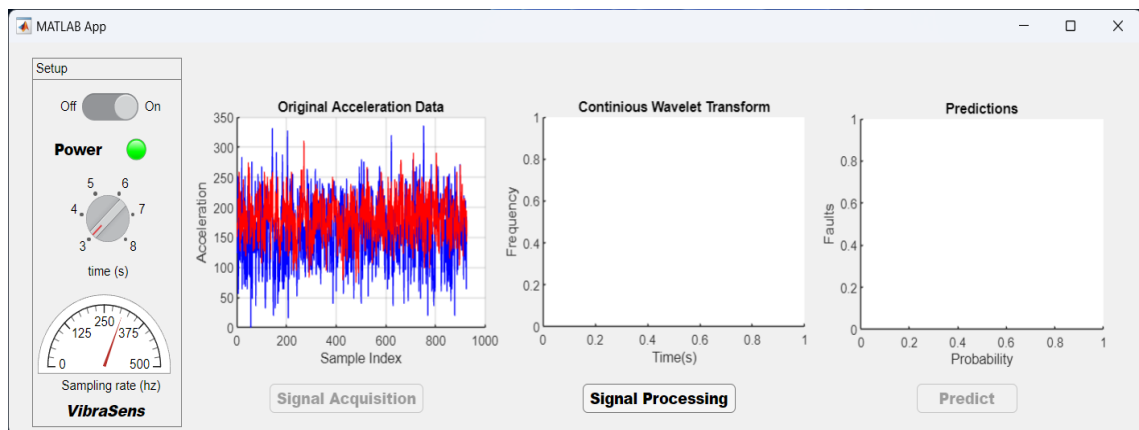


Figure 4.11. Signal acquired

Once the accelerations are plotted, the ‘Signal Processing’ button will be enabled so the user can click the ‘Signal Processing’ button to open a window that allows them to choose the domain and method they want to use for signal processing.

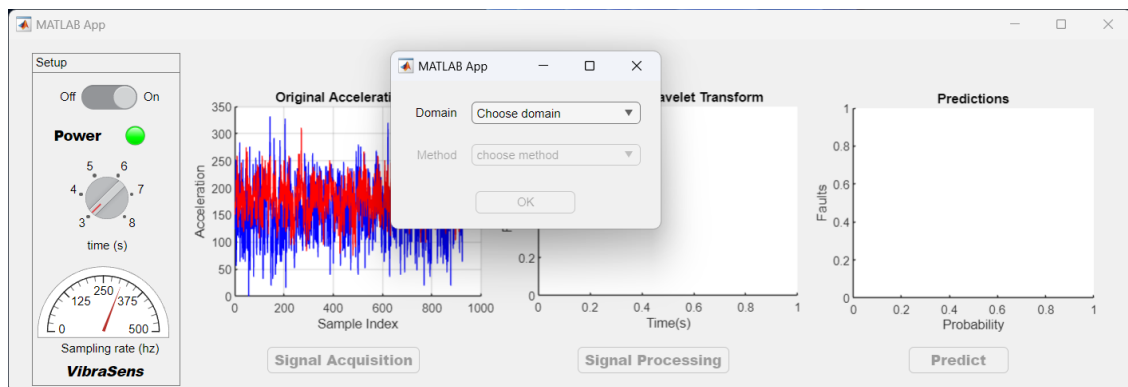


Figure 4.12. Domain choosing

Once the user chooses the domain the methods will appear according to the domain that the user has chosen.

The three options available are time, frequency, and time-frequency.

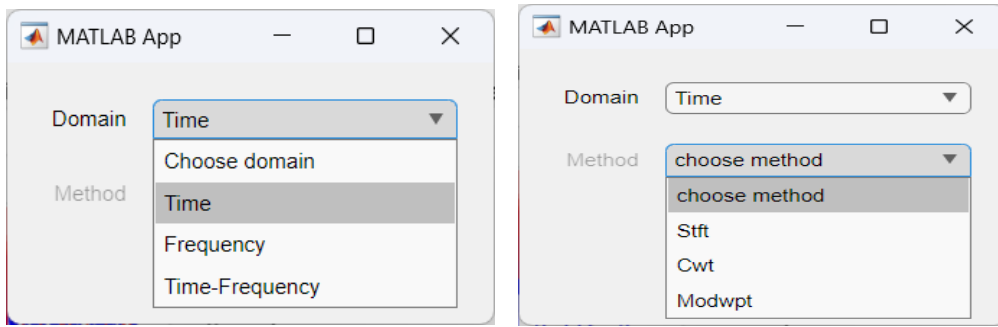


Figure 4.13. Method choosing

Once the user has chosen the domain and method, they can push the ‘OK’ button to close the window.

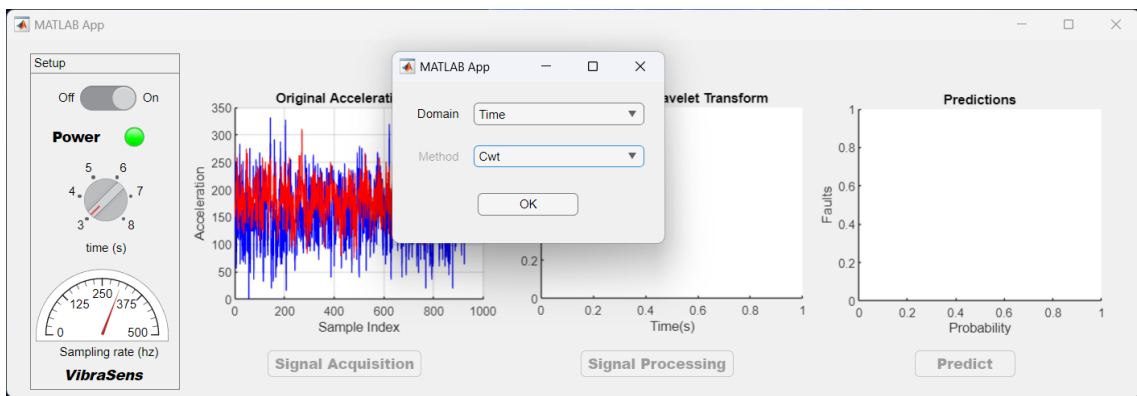


Figure 4.14. Domain and method have been chosen

The application will then plot the image that corresponds to the method the user has chosen and enable the ‘Predict’ button. The ‘Signal Processing’ is still enabled in case the user wants to use other methods.

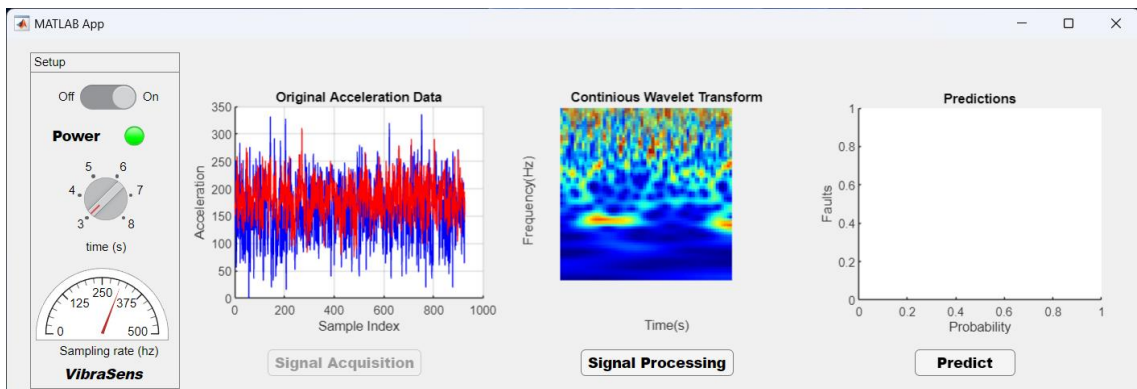


Figure 4.15. Signal processed.

Finally, the user can push the ‘Predict’ button to predict the type of fault or the status of the rotary machine.

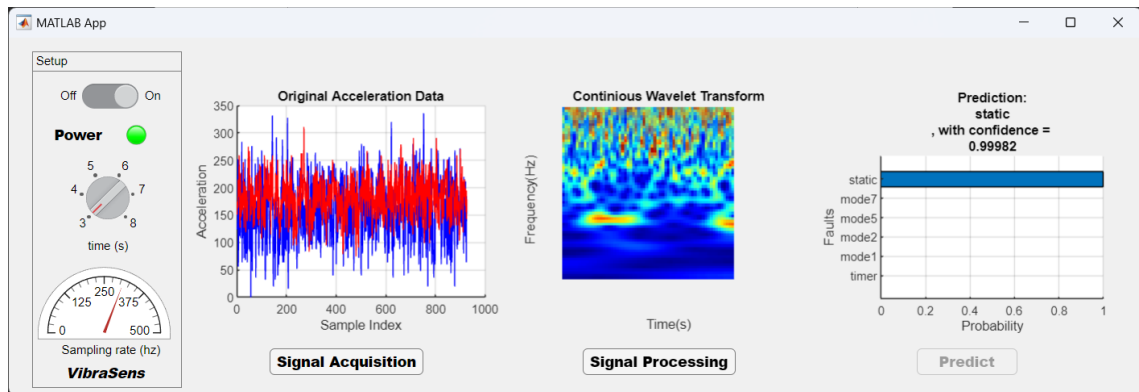


Figure 4.16. The predicted status of the motor

4.3. Results and Discussion

In our study, we conducted an extensive investigation into fault diagnosis techniques for rotating machinery, aiming to enhance diagnostic accuracy and reliability. Utilizing signal processing methodologies, including `modwpt`, `stft`, `cwt`, `xspectrogram`, `pspectrum`, `pkurtosis`, `instfreq`, and `kurtogram`, we meticulously analyzed machinery performance to detect potential faults. By leveraging these techniques, we transformed raw signals into image representations, facilitating the application of deep learning models for fault classification.

The integration of ResNet architecture played a pivotal role in our approach, allowing for the development of a tailored model capable of effectively discerning fault patterns from processed images. We adhered closely to ResNet principles while optimizing the architecture to suit the specific requirements of fault diagnosis applications. This involved fine-tuning parameters such as layer configurations and skip connections, resulting in a robust and adaptable diagnostic framework.

Our research yielded promising results, with the developed ResNet model achieving an impressive accuracy of approximately 98.01%. Furthermore, sensitivity analysis revealed the significant impact of parameters such as window size and stride on diagnostic performance, underscoring the importance of meticulous optimization.

However, despite the overall high accuracy, we noted several misclassified cases represented in the confusion matrix.

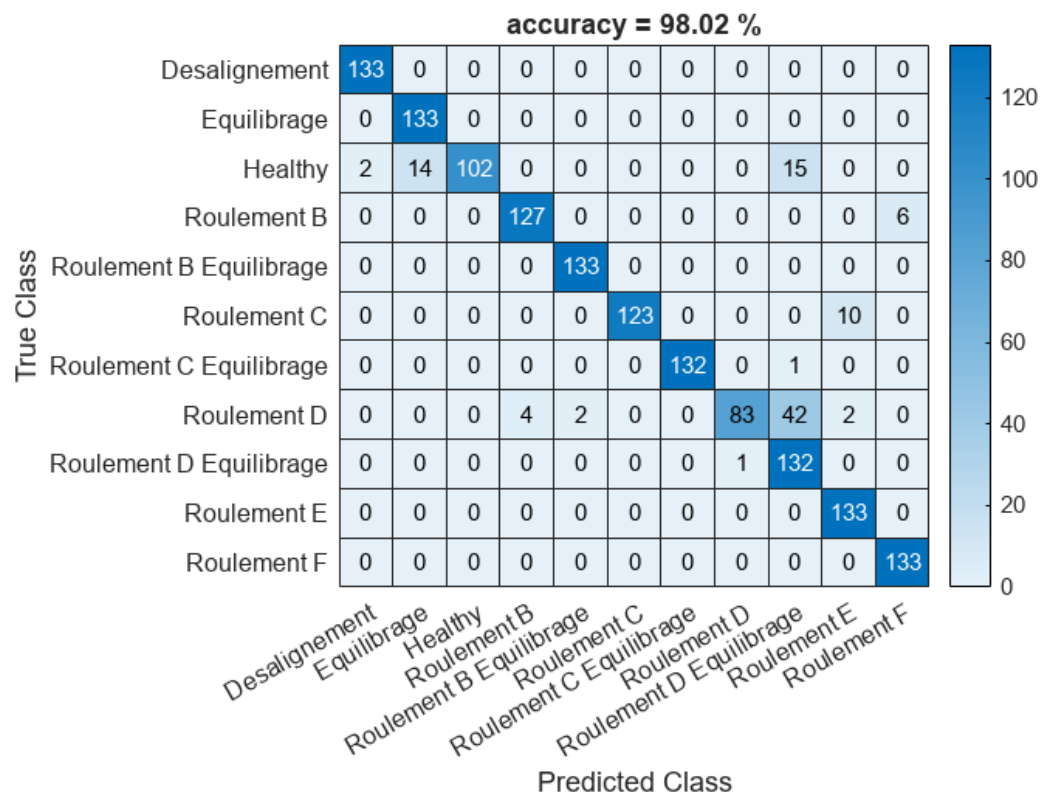


Figure 4.17. Confusion matrix

Table 4.2 presents the accuracies achieved by various signal processing techniques utilized in our study. Each technique was evaluated based on its ability to enhance fault diagnosis accuracy in rotating machinery.

Table 4.2. Accuracies of signal processing techniques

Signal processing technique	Accuracy (%)
Modwpt	92.48
Cwt	98.01
Stft	93.91
Pspectrum	94.39
Xspectrogram	93.23
Kurtogram	71.01

Among the signal processing techniques utilized, the Continuous Wavelet Transform (CWT) exhibited the highest accuracy at 98.01%, followed closely by the Power Spectrum (Pspectrum) at 94.39%. Nonetheless, the Kurtogram technique

demonstrated a lower accuracy of 71.01%. One potential reason for misclassifications could be the limited feature representation offered by certain signal processing techniques. For instance, the Kurtogram technique, despite its widespread use, may not adequately capture all relevant features of the data, leading to incomplete representations that hinder accurate classification.

In conclusion, our research demonstrates the effectiveness of utilizing ResNet models for fault diagnosis in rotating machinery. We achieved a commendable accuracy of approximately 98.01% through meticulous optimization of model parameters. Sensitivity analysis highlighted the critical influence of window size and stride on diagnostic performance, emphasizing the need for careful parameter tuning. While the ResNet model showcased high overall accuracy, the analysis of misclassified cases revealed potential limitations in certain signal processing techniques, such as the Kurtogram, which exhibited a lower accuracy of 71.01%. Future research could focus on refining feature representations to address these limitations and further enhance diagnostic accuracy. Overall, our study contributes valuable insights into the application of deep learning models and signal processing techniques for fault diagnosis, paving the way for more reliable and efficient machinery maintenance strategies.

Bibliography

- [1] Jaspreet, "Towards Data Science," 13 Aug 2016. [Online]. Available: <https://towardsdatascience.com/a-concise-history-of-neural-networks-2070655d3fec>.
- [2] Y. Adan, "Medium," 29 September 2018. [Online]. Available: <https://medium.com/swlh/do-neural-networks-really-work-like-neurons-667859dbfb4f>.
- [3] S. Y. Y. C. T. G. Jingli Yang, "A Fault Diagnosis Method of Rotating Machinery Based on One-Dimensional, Self-Normalizing Convolutional Neural Networks," *Sensors*, vol. 20, no. 14, 2020.
- [4] T. Wood, "Deep Ai," [Online]. Available: <https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network>.
- [5] M. F. R. H. X. T.-L. B. T. Yuxin Wen, "Recent advances and trends of predictive maintenance from data-driven," *Elsevier*, no. 11, 2021.
- [6] P. B. L. B. T. Yue Cui, "A fault detection framework using recurrent neural networks for condition monitoring of wind turbines," *Wiley*, vol. 24, no. 11, pp. 1249-1262, 2021.
- [7] "Wikipedia," Wikipedia , 11 July 2023. [Online]. Available: https://en.wikipedia.org/wiki/Residual_neural_network .
- [8] "Keras," Keras, [Online]. Available: <https://keras.io/api/applications/resnet/>.

- [9] "Geeks for Geeks," Geeks for Geeks, 2008. [Online]. Available: <https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/> . [Accessed 2008].
- [10] "Wikistat," [Online]. Available: <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-hdstat-rnn-deep-learning.pdf>.
- [11] C. W. J. Z. S. Z. Wanlu Jiang, " Application of Deep Learning in Fault Diagnosis of," *Processes*, vol. 9, no. 19, p. 22, 2021.
- [12] C. S. Sergey Ioffe, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," p. 15, 2015.
- [13] Wiki.tum.de., "Layers of a Convolutional Neural Network - Convolutional Neural Networks for Image and Video Processing," TUM Wiki., [Online]. Available: <https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network> . [Accessed 10 May 2019].
- [14] V. V. S. I. J. S. Z. W. Christian Szegedy, "Rethinking the Inception Architecture for Computer Vision.," in *Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [15] M. documentation., "Multiscale variance of maximal overlap discrete wavelet transform," MATLAB modwtvar., [Online]. Available: <https://www.mathworks.com/help/wavelet/ref/modwtvar.html>. [Accessed 10 May 2019].
- [16] A. G. B. Richard S. Sutton, "Reinforcement Learning: An Introduction.," *MIT Press.*, 2018.

- [17] C. M. Bishop, "Pattern Recognition and Machine Learning," *Springer*, 2006.
- [18] M. K. K. S. K. Sujesh Kumar, "Vibration based Fault Diagnosis Techniques for Rotating Mechanical Components: Review Paper," *IOP Science*, vol. 376, no. 012109, p. 7, 2018.
- [19] J. L. M. X. Xiaohua Song, "Advanced Vibration-Based Fault Diagnosis and Vibration Control Methods.," *Sensors*, vol. 23, no. 7704, p. 6, 2023.
- [20] B. Minner, "acoem," ACOEM USA, 26 January 2021. [Online]. Available: <https://acoem.us/blog/other-topics/6-common-faults-of-rotating-machinery/>. [Accessed 26 January 2021].
- [21] O. O. B. M. G. R. R. M. Saxena, "Bearing Fault Monitoring Using CWT Based Vibration Signature," *Science Direct*, vol. 144, pp. 234-241, 2016.
- [22] A. K. Ovacikli, "Condition Monitoring of Ball Bearings Using Vibration Analysis," p. 54, 2010.
- [23] S. P. M. D. I. Lalwani, "Fault diagnosis of bent shaft in rotor bearing system," *Mechanical Science and Technology*, vol. 31, pp. 1-4, 2017.
- [24] A. W. L. M. I. F. S. Edwards, "Fault Diagnosis of Rotating Machinery," *The Shock and Vibration Digest*, vol. 30, no. 1, pp. 4-13, 1998.
- [25] K. Adamsab, "Machine learning algorithms for rotating machinery bearing fault diagnostics," *materialstoday:Proceeding*, vol. 44, no. 6, pp. 4931-4933, 2021.

- [26] S. H. Y. W. Chaofan Hu, "A classification method to detect faults in a rotating machinery based on kernelled support tensor machine and multilinear principal component analysis," *Applied Intelligence*, vol. 51, no. 2, pp. 1-13, 2021.
- [27] B. Minner, "acoem," ACOEM USA, 1983. [Online]. Available: <https://acoem.us/blog/other-topics/6-common-faults-of-rotating-machinery/>. [Accessed 26 January 2018].
- [28] D. I. L. S. P. Mogal, "Fault diagnosis of bent shaft in rotor bearing system," *Mechanical Science and Technology*, vol. 31, pp. 1-4, 2017.
- [29] K. S. P. Mahesh B. JOSHI, "FAULT DIAGNOSIS OF HIGH-SPEED ROTATING MACHINES USING MATLAB," *DIAGNOSTYKA*, vol. 24, no. 2, pp. 1-8, 2023.
- [30] G. S. Zoltán Germán-Salló, "Signal processing methods in fault detection in manufacturing systems," *Procedia Manufacturing*, vol. 22, pp. 613-620, 2018.
- [31] S. U. Akhand Rai, "A Review on Signal Processing Techniques Utilized in the Fault Diagnosis of Rolling Element Bearings," *Tribology international*, pp. 1-53, 2015.
- [32] S. M. V. D. Y. N. S. S. D. Viplav Barai, "Bearing fault diagnosis using signal processing and machine learning techniques: A review," *Materials Science and Engineering*, pp. 1-11, 2022.
- [33] M. L. K. K. S. K. Sujesh Kumar, "Vibration based Fault Diagnosis Techniques for Rotating Mechanical Components: Review Paper," *Materials Science and Engineering*, pp. 1-7, 2018.

- [34] MATLAB, "MathWorks," 2016. [Online]. Available: <https://www.mathworks.com/help/wavelet/ref/modwptdetails.html>. [Accessed 2023].
- [35] R. X. G. X. C. Ruqiang Yan, " Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, vol. 96, pp. 1-15, 2013.
- [36] Wikipedia, "Wikipedia," Wikimedia Foundation, Inc., [Online]. Available: https://en.wikipedia.org/wiki/Short-time_Fourier_transform. [Accessed 16 December 2023].
- [37] MATLAB, "MATLAB documentation," The MathWorks, Inc., 1994. [Online]. Available: <https://www.mathworks.com/help/signal/ref/spectrogram.html>.
- [38] Y. Y. T. F. Y. S. X. Z. Yiyang Liu, "Research on Rotating Machinery Fault Diagnosis Method Based on Energy Spectrum Matrix and Adaptive Convolutional Neural Network," *Processes*, vol. 9, no. 1, pp. 1-25, 2020.
- [39] C. L. D. P. JIAYU CHEN, "Fault Diagnosis of Rotating Machinery: A Review and Bibliometric Analysis," *Digital Object Identifier*, pp. 1-19, 2020.
- [40] M. Documentations, "Estimate instantaneous frequency," MathWorks, Inc., 1994. [Online]. Available: <https://www.mathworks.com/help/signal/ref/instfreq.html>. [Accessed 2023].
- [41] M. documentations, "Spectral kurtosis from signal or spectrogram," Mathworks, Inc., 1984. [Online]. Available: <https://www.mathworks.com/help/signal/ref/pkurtosis.html>. [Accessed 2022].

[42] T. InvenSense, "MPU-9250 Datasheet," 20 June 2016. [Online]. Available: <https://invensense.tdk.com/download-pdf/mpu-9250-datasheet/>. [Accessed 2022].

[43] E. Systems, "The internet of things with ESP-32," [Online]. Available: <http://esp32.net/>. [Accessed 2022].