زانکۆی پۆلیتەکنیکی هەولێر
ERBIL POLYTECHNIC UNIVERSITY

Gear Fault Detection Based on Time and Non-Time Series Feature Representation Using Machine Learning

A Doctoral Dissertation
Submitted to the Council of Erbil Technical Engineering College, at Erbil Polytechnic University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Information Systems Engineering.

By

Zrar Khald Abdul

B.Sc. Computer Engineering (2009)

M.Sc. Control System Engineering (2012)

Supervised by

Asst. Prof. Dr Abdulbasit Kamil Faeq

Erbil, Kurdistan

October 2023

DECLARATION

I declare that the PhD Dissertation entitled: **Gear Fault Detection Based on Time and Non-Time Series Feature Representation Using Machine Learning**, is my own original work, and hereby I certify that unless stated, all work contained within this dissertation is my own independent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgment is made in the text.

Signature:

Student Name: Zrar Khald Abdul

Date: / /2023

CERTIFICATION OF PROOFREADING

This is to certify that this dissertation entitled "**Gear Fault Detection Based on Time and Non-Time Series Feature Representation Using Machine Learning"** written by the postgraduate student (**Zrar Khald Abdul**) has been proofread and checked for grammatical, punctuation, and spelling mistakes. Therefore, after making all the required corrections by the student for further improvement, I confirm that this last copy of the dissertation is ready for submission.

Signature:

Name: Chalak Ali Mohammed Ameen

Lecturer, Applied linguistics.

Phone No. 07717590272

Email address: chalak.ali@charmouniversity.org

Date: / /2023

SUPERVISOR CERTIFICATION

This dissertation has been written under my supervision and has been submitted for the award of the Doctor of Philosophy in Information Systems Engineering with my approval as supervisor.

Signature:

Name: Assistance Prof. Dr. Abdulbasit Kamil Fayaq
(Supervisor)
Date: / /2023

I confirm that all requirements have been fulfilled.

Signature:

Name:
Head of the Department of Information Systems Engineering

Date: / /2023

I confirm that all requirements have been fulfilled.

Postgraduate Office

Signature:

Name:

Date: / /2023

EXAMINING COMMITTEE CERTIFICATION

We certify that we have read this Dissertation: "**Gear Fault Detection Based on Time and Non-Time Series Feature Representation Using Machine Learning**" and as an examining committee examined the student (**Zrar Khald Abdul**) in its content and what related to it. We approve that it meets the standards of a dissertation for the degree of Doctor of Philosophy in Information Systems Engineering.

Signature:
Name: Assist. Prof. Abbas Mohamad Ali
(Member)
Date: / /2023

Signature:
Name: Assist. Prof. Kayhan Zrar Ghafoor
(Member)
Date: / /2023

Signature:
Name: Assist. Prof. Alaa Muheddin Abdulrahman
(Member)
Date: / /2023

Signature:
Name: Asst. Prof. Shahab Wahhab Kareem
(Member)
Date: / /2023

Signature:
Name: Asst. Prof. Dr Abdulbasit Kamil Faeq
(Supervisor-Member)
Date: / /2023

Signature:
Name: Prof. Shavan Kamal Askar
(Chairman)
Date: / /2023

Signature
Name: Prof. Dr. Ayad Zaki Saber AGHA
Dean of the College of Erbil Technical Engineering
Date: / /2023

ACKNOWLEDGMENTS

First of all, I would like to thank ALLAH almighty, the most merciful and compassionate, for His support, help, and generosity.

I am deeply indebted to my supervisor's Assistant Prof. Dr. Abdulbasit Kamil Faeq. I would like to express my sincere gratitude to him for his continuous support, his patience, motivation, and immense knowledge during my Ph.D. study and writing this dissertation.

To my family, thank you for all that you did. Your support means the world to me. Finally, I would like to thank Charmo University and Erbil Polytechnic University for their support during my PhD study.

ABSTRACT

The automatic fault detection in rotating machinery has emerged as crucial factors for ensuring the high reliability of modern industrial systems. Therefore, developing automatic fault detection is a vital challenge in modern industry. This dissertation intends to develop an automatic model based on machine learning to detect gear faults. During the development of an automatic model for gear fault detection, certain issues are found that need to be addressed to establish a reliable monitoring system. Firstly, the literature has not yet explored the potential of representing vibration signals, despite their time-series nature, using both non-time-series based and/or time-series-based features. Secondly, the vibration signal may have different channels based on the type of accelerometer sensor. A lack of studies is noticed to show the impact of the representation of these channels on the performance of the fault detection system. And thirdly, the fault diagnosis process for rotating machinery becomes challenging due to the non-stationary and non-linear characteristics that commonly arise from varying operating conditions. Various conditions make it difficult for traditional linear to effectively capture the underlying fault patterns.

To address these problems, some investigations are required to be studied including investigating various feature representations for fault detection, investigating the nature of fault detection in terms of time series or non-time series analysis, investigating various forms of fusion models using traditional concatenation or adopting multi reservoir to model with multi-channels, and investigating time-consuming of model.

Regarding the utilized features, Mel Frequency Cepstral Coefficients (MFCCs) and Gamma tone Cepstral Coefficients (GTCC) have been adopted as a feature and various forms of these features have been used to feed two main different models including the time series model and non-time series

model. For the time series model, Long Short-Term Memory (LSTM) and Echo State Network (ESN) have been adopted to classify the gear faults. The high performance of LSTM is achieved for gear fault classifications despite its being time-consuming during the training phase. To avoid time consumption, the ESN, which consumes less time as some of their layer's weight values are non-trainable and selected randomly. Further investigation has been studied by adopting multi channels reservoir which has led to achieving reliable gear fault detection.

Regarding the non-time series model, Support Vector Machine is fed by two different forms of the feature representation, the first of which is the statistical form called in this dissertation (stat-SVM). The problem with the statistical form is that it may lead to loss of some important features related to gear fault and it may lead to degradation in gear fault detection. To address this problem, we use a concatenation of the frames of both features (MFCC and GTCC) to be fed to the SVM (concat-SVM). Consequently, a high-performance rate of gear fault detection is achieved. Further, investigation is adopted by optimizing the features hyperparameters as both features were originally designed to extract features from speech signals. For this purpose, Grey Wolf Optimization (GWO) and Fitness Dependent Optimizer (FDO) have been utilized to optimize three hyperparameters of both GTCC and MFCC. The performance of optimizing the hyperparameters of MFCC has not shown any improvements. Oppositely, improvement in GTCC performance by the same optimization process is observed and validated.

All the proposed models have been evaluated by two public datasets namely, Prognostic Health Monitoring 2009 (PHM09) and Drivetrain Dynamic Simulator (DDS). Based on the result, non-time series model demonstrated superior performance compared to time series models within a margin of 2 to 15% accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
ANN	Artificial Neural Network
BiGRU	Bidirectional Gated Recurrent Units
BPNN	Back-Propagation Neural Network
CNN	Convolutional Neural Network
CNN-B	Convolutional Neural Network Bayesian
concat-SVM	Concatenating- Support Vector Machine
DCT	Discrete Cosine Transform
DDS	Drivetrain Dynamic Simulator
DFT	Discrete Fourier Transformer
DT	Decision Tree
DWT	Discrete Wavelet Transform
EI	Expected Improvement
ESN	Echo State Network
FDO	Fitness Dependent Optimizer
FFT	Fast Fourier Transform
FNN	Fuzzy Neural Network
fw	Fitness Weight
GMF	Gear Mesh Frequency
GP	Gaussian Process
GRU	Gated Recurrent Unit
GTCC	Gamma tone Cepstral Coefficients
GWO	Grey Wolf Optimization
HTF	Hunting Tooth Frequency
IGWO	Improved Grey Wolf Optimization
IWO	Invasive Weed Optimization

Abbreviation	Meaning
k-NN	k-Nearest Neighbor
LBP	Local Binary Pattern
LFGRU	Local Feature-Based Gated Recurrent Unit
LR	Linear Regression
LSTM	Long Short-Term Memory
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multi-Layer Perceptron
MP	Matching Pursuit
PCA	Principal Component Analysis
PHM09	Prognostic Health Monitoring 2009
PI	Probability of Improvement
PSO	Particle Swarm Optimization
PSVM	Proximal Support Vector Machines
RBF	Radial Basis Function
RNN	Recurrent Neural Network
SAE-DNN	Stacked Autoencoder-Deep Neural Network
SSAE	Stacked Sparse Auto Encoder
Stat-SVM	Statistical- Support Vector Machine
SVM	Support Vector Machine
t-SNE	t-stochastic neighbor embedding
UCB	Upper Confidence Bound
VMD	Variational Modal Decomposition

CHAPTER ONE

1 INTRODUCTION

1.1 Overview

Rotating machinery is a crucial component and one of the most vital parts of mechanical equipment. It's functioning heavily depends on rotation to serve a specific purpose, and it finds widespread use in mechanical transmission applications such as aircraft engines, wind turbine generator systems, gas turbine engines, pumps, and gearbox systems. However, due to inevitable malfunction and equipment downtime during operation, diagnosing faults is of immense importance for ensuring the reliability and safety of rotating machinery (Tang et al., 2020a). Therefore, the reliability of machines or tools is gaining importance in the industry because of the need to decrease the possible loss of production whenever the machine experiences an abnormal situation during the working load. There are plenty of useful market assessment tools in the industry, which are utilized as well by analysts to understand the competitive dynamics of an industry (Candanedo *et al.*, 2018). A health monitoring system is one of the well-known assessment tools that has been used to monitor a dynamic system and predict a failure in the early stages (Henriquez *et al.*, 2014). This monitoring process is mostly conducted using various channels such as vibration (Wang *et al.*, 2019), temperature, lubrication, and acoustic signal (Qu *et al.*, 2013).

The successful implementation of a condition monitoring system can provide valuable and reliable information for maintenance programs, resulting in significant cost benefits for industries. By reducing unnecessary scheduled maintenance and minimizing unplanned downtime, the machinery can operate at as high as possible of its capacity. One of the most used and popular monitoring techniques for rotating machinery is vibration signature analysis,

which is considered an important predictive tool in most maintenance programs. Vibration signals contain crucial information about the mechanical condition of the various parts involved, as well as reflecting the overall system assembly's performance. The vibration signals are highly sensitive to any abnormality in the moving components, such as a bent rotating shaft or a damaged gear tooth, and are directly related to the periodic movement of the machine components (Edwards *et al.*, 1998).

An automatic condition monitoring system can be adopted by using machine learning algorithms to automatically monitor the condition of the equipment or the system and detect anomalies or faults that could indicate impending failure. The role of machine learning in fault detection is to enable the automated detection of faults or anomalies in a system or process, based on data from sensors or other sources. By using machine learning algorithms to analyze this data, it is possible to identify patterns and deviations that may indicate the presence of a fault or anomaly (Kateris *et al.*, 2014).

There is a variety of machine learning techniques that can be used for fault detection, including supervised learning, unsupervised learning, and reinforcement learning. Supervised learning algorithms can be used to train a model on a labeled dataset of normal and faulty behavior, and then to use that model to detect anomalies in real-time data. Unsupervised learning algorithms can be used to identify patterns in data that deviate from normal behavior, without requiring labeled data. Reinforcement learning can be used to detect faults by learning to optimize a reward function that encourages normal behavior and penalizes deviations. One of the advantages of using machine learning for fault detection is that it can be used to detect faults that may be difficult to detect using traditional rule-based methods. Machine learning can also adapt to changing conditions and learn from new data, making it a useful

tool for fault detection in dynamic or complex systems (Chahal and Gulia, 2019).

As machine learning advances, deep learning techniques have been increasingly applied to gear fault detection. Deep learning models can be trained using raw signals and/or hand-crafted features. In studies utilizing raw signals, researchers have directly fed vibration raw signals into Convolutional Neural Network (CNN) models, eliminating the need for prior knowledge of application-specific features (Yang *et al.*, 2019)-(Jing *et al.*, 2017). On the other hand, Shao et al. (Shao *et al.*, 2019) employed time-frequency distributions to generate images through wavelet transforms, converting raw input into image data. The authors (Wang *et al.*, 2017) introduced local feature-based gated recurrent unit networks, which are trained using both handcrafted and automatically learned features.

1.2 Problem Statement

The problem statements are listed below:

- Despite the time-series nature of the vibration signal, it can be represented in both non-time series-based features and/or time-series based features. However, this characteristic has not been investigated in the literature. Additionally, representation of the feature in the non-time series form for fault detection can also be problematic and need a suitable model design.
- Vibration signals for gear fault detection may have various channels based on the accelerometer sensor dimensions, representation of these channels can impact the performance of gear fault detection systems.

Designing models for such a situation is another problem that has been dealt with in this dissertation.

- Rotating machinery often exhibits non-stationary and non-linear characteristics due to varying operating conditions, load changes, and wearing and tearing over time. These characteristics complicate the fault diagnosis process, as traditional linear and stationary methods may not accurately capture the underlying fault patterns.

1.3 Challenges in Gear Fault Diagnosis

As with any automatic detection process, gear fault diagnosis faces various challenges especially when adopting vibration signal and utilizing machine learning tools, which require data collection, feature extraction and selection in addition to classification model selection. The main challenge in this dissertation that have been faced is lack of data. Collecting special datasets, requires access to proper laboratories, or collaborating with industry partners, and research institutions to pool data resources and create larger datasets, which are not available and applicable here. Therefore, this lack of data became a challenge during working on this dissertation. To overcome this challenge, we use existing data.

1.4 Research questions

This dissertation aims at addressing the problems stated in the previous section. During the investigation made in this work, some essential research questions have been raised, following are the research questions of this dissertation:

- How the time step sample dependencies of the vibration signal are important in fault detection?
- Is there any possibility to reduce the training time at the training phase for fault detection in the era of deep learning?
- Is the representation of various channels of the vibration signal affecting the performance of the model? What is the ideal way to deal with the channel representation?
- Regarding the available hand-crafted features which are originally extracted for speech applications such as (MFCC and GTCC), are the default parameters suitable for vibration signal-based applications?

1.5 The objective of this Work

- To Investigate and analyze various feature representations for fault detection, which gives insight into the time series-based features and the non-time series-based feature representation impact on fault detection task.
- To design a model that can overcome the challenge of high computation time in the training phase to minimize the need for resources such as the non-trained weight ESN model, which can perform comparably to the trained weight Long Short-Term Memory (LSTM).
- To design, implement and evaluate various forms of fusion models using traditional concatenation or adopting multi reservoir to model multi-channel gear fault vibration signal.
- To evaluate the proposed models using two main mechanical datasets that include various gears and gearboxes.

The connections between problem statement, research question and objectives are illustrated in Figure 1.1.

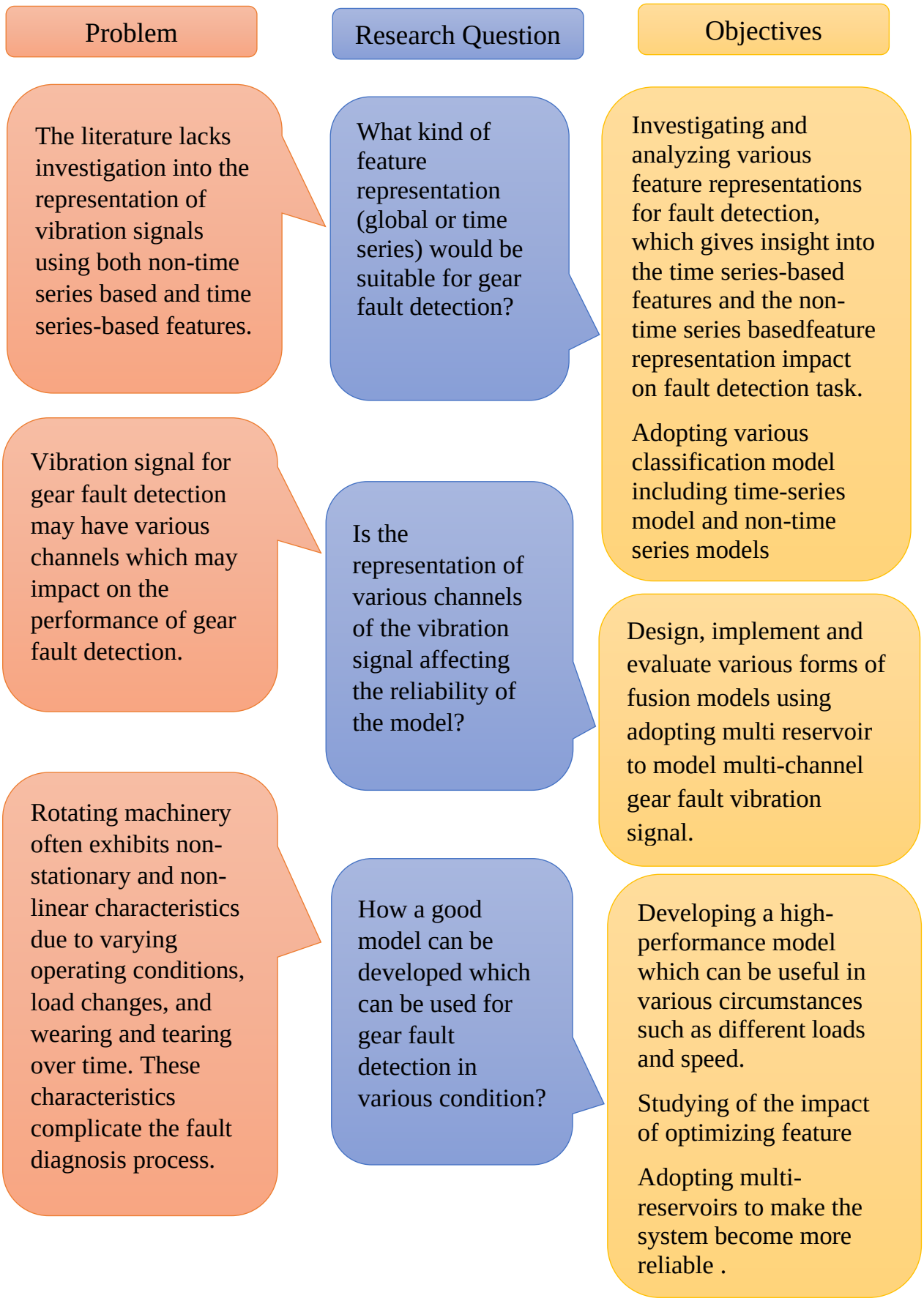

Figure 1.1: connection between problem, research question and objective.

1.6 Scope of the Work

In this dissertation, the vibration signal is adopted to identify faults in gear. Additionally, as the dissertation eventually tries to develop a model for automatic fault detection, the usefulness of machine learning algorithms has been utilized for this purpose. The performance of cepstrum features (Mel Frequency Cepstral Coefficients (MFCC) and Gamma tone cepstrum coefficients (GTCC) and some representation of them have been investigated to train the adopted and proposed classifiers. Regarding the utilized tools that were used for this dissertation, all adopted machine learning models were implemented on MATLAB except ESN model was implemented on Python. The experiments were conducted on a personal computer with Core (TM) i7-7500U CPU @ 2.70GHz 2.90 GHz.

1.7 Dissertation Organizations

The dissertation map is organized into five chapters. The first chapter gives an overview of the dissertation topic, problem statements, dissertation objectives, and dissertation organization.

The second chapter presents the literature review regarding fault detection and the theoretical background of algorithms. Based on the literature, condition monitoring systems can be categorized based on whether they analyze time series data or non-time series data. Brief literature for both of them is presented.

The third chapter is dedicated to gear fault detection based on the time series model. Two of the well-known machine learning algorithms (LSTM and ESN) are explained and two versions of ESN including one reservoir and three reservoirs are implemented.

In the fourth chapter, gear fault detection based on the non-time series feature is presented. Two types of feature representations are fed to Support Vector Machine (SVM) including statistical representation of MFCC and GTCC and concatenated both features. We named stat-SVM and concat-SVM for the statistic model and concatenated feature model.

Finally, the conclusion about time series and non-time series analysis is presented in chapter five. We also provided a reason for each of the obtained results. The suggested future studies are given at the end of the conclusion section.

CHAPTER TWO

2 BACKGROUND AND LITERATURE REVIEW

2.1 Rotating Machinery

Rotating machinery refers to mechanical systems that consist of various rotating components that convert energy, typically in the form of rotational motion, into useful work. These systems play a crucial role in numerous industries, including power generation, manufacturing, transportation, and process industries. Nowadays, plenty of rotating machinery systems are used such as turbines (extract energy from fluids)(Qiao and Lu, 2015), Pumps (move fluids)(Muralidharan and Sugumaran, 2013), Compressors (increase the pressure of a fluid), Electric motors (convert electrical energy into mechanical energy)(Choi *et al.*, 2020), and Gearboxes (transmit and convert rotational energy between different components of a machine). Changing energy is becoming increasingly important as the world continues to move towards renewable sources of energy. Therefore, rotating machinery is essential for modern industries, and effective maintenance and fault diagnosis is crucial to ensure their safe, efficient, and reliable operation (Liu *et al.*, 2018).

2.1.1 Gearbox

A gearbox is a mechanical apparatus that utilizes a set of gears to transfer power from a rotating power source, like an engine or motor, to another machine or mechanical component, enabling adjustment of the rotational motion's speed, torque, and direction. Three key components found within a gearbox are shafts, bearings, and gears (Abdul *et al.*, 2016).

A. Shaft

A shaft refers to a cylindrical component that is utilized to transmit power and motion between the rotating parts of a machine. Typically made of durable materials like steel, they are designed to be rigid enough to withstand various forces like bending, torsion, and other stresses that may arise during operation (see Figure 2.1). The principal function of a shaft is to transfer rotational motion and torque from a power source, like an engine or motor, to a driven machine or tool. For instance, an electric motor can utilize a shaft to transmit power to a conveyor belt, a pump, or a machine tool. Moreover, shafts can be utilized to support rotating components and create a sturdy connection between two or more machines. In such cases, the shaft serves as a stable base for the components to rotate around, maintaining the alignment and stability of the machines (Watson *et al.*, 2007).

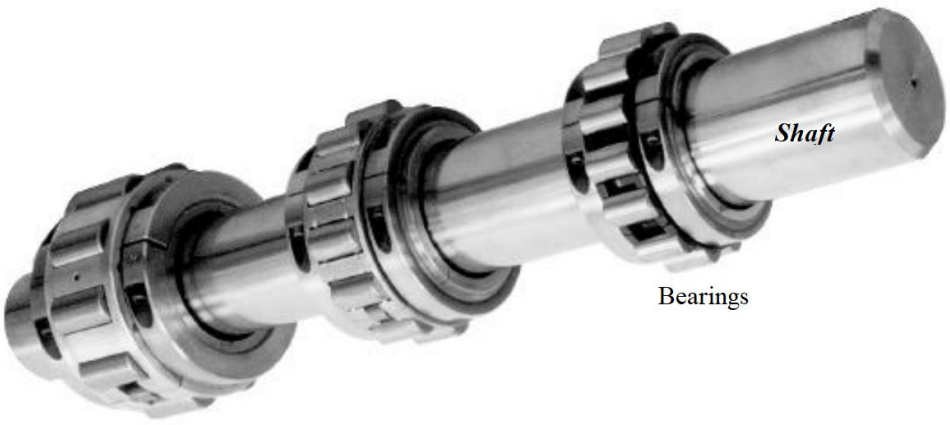

Figure 2.1: Structure of Shaft(Clive Jennings, 2020)

B. Bearings

Bearings are integral components that guide and support rotating shafts, which create a seamless and efficient connection between the shaft and the surrounding machinery. By minimizing friction between the moving parts, bearings prevent wear and tear, promote smoother operation, and improve the performance of machines. The principal purpose of bearings is to reduce friction between the rotating shaft and the housing. Typically, a bearing comprises two primary components: an outer race or housing and an inner race or sleeve, with a series of balls, rollers, or needles placed between them (see Figure 2.2). As the shaft rotates, the balls or rollers move along the raceways, evenly distributing the load and decreasing friction across the bearing (Klein *et al.*, 2011).

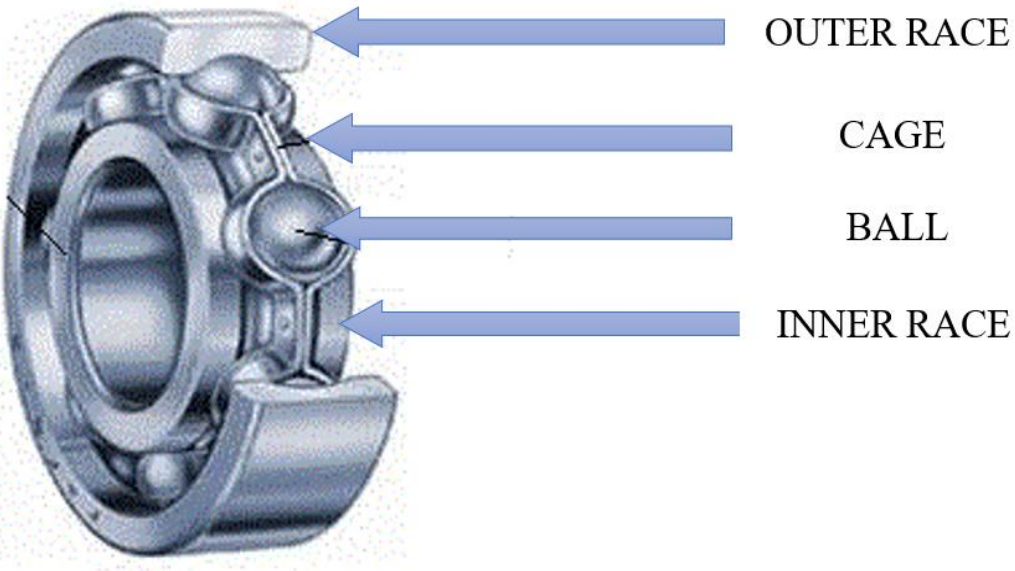

Figure 2.2: Structure of bearing(Brkovic *et al.*, 2017)

C. Gears

Transmitting power from one shaft to another can be done using gears through engaging gear teeth. When the teeth of the gears mesh with each other, motion, torque, and force are transmitted from one shaft to another. The speed of the shaft can be reduced or magnified by using two gears with different sizes or different numbers of teeth that they have. Nowadays, there are plenty of gears have been developed such as spur, helical, spiral bevel, bevel, internal, and worms which are used in the various tools and gearboxes (Smith, 2003). Figure 2.3 illustrates some of the gear types.

Figure 2.3: Some types of gears(Mec.Edu teams, 2020)

All machinery components, including gears, generate a mechanical force during normal condition operation which is named dynamic force. The force is created due to the mesh of the gear teeth, and this is the main source of the

vibration and noise in gears. The dynamic force leads to the construction of both broadband and impulsive noise, even gears run under normal situations. There are other sources of vibration and noise in gear namely, geometry factors, variations of speed and load, and expulsion of fluid. Pressure angle, gear pitch, contact ratio, alignment, tooth face width, and tooth surface finish, all are examples of geometry factors. A noise and shock wave might be generated due to the expulsion of air and lubricant when the teeth interact with each other (Norton and Nelson, 1990).

A. Characteristics of Vibration Frequency in Gears

During the condition monitoring of gears, multiple frequencies in the gear's vibration need to be taken into account. These frequencies include the Gear Mesh Frequency (GMF), sidebands of the gear mesh frequency, Hunting Tooth Frequency (HTF), assembly phase frequency, gear natural frequencies, and ghost or phantom frequencies. GMF refers to the frequency at which the teeth of two meshing gears come into contact or engage with each other. It is a measure of the rate at which the teeth make contact and separate as the gears rotate. Gear mesh frequency sidebands are additional frequencies that appear around the primary gear mesh frequency in a gear system. These frequencies are very useful in the diagnosis of the gear as they are indicators of the bad condition of the gears. HTF is related to the contacting teeth of the gears. The maximum vibration will observe if there are damages in both contacting teeth (Smith, 2003).

B. Defects in Gears

The tooth of gears is contacted at the same time with their respective pinions in the gearbox which consequences in the sliding of each tooth on the other thereby generating vibrations. Gear faults can arise due to a variety of reasons, which can be broadly categorized into manufacturing defects, material issues, operational factors, and maintenance-related problems. Here is a list of common reasons that contribute to gear faults (Sharma and Parey, 2016; Wilk *et al.*, 2008). Figure 2.4 illustrates some of the gear faults.

- Manufacturing defects: Imperfections during the gear manufacturing process, such as inaccuracies in gear tooth geometry, surface roughness, or heat treatment, can lead to weaknesses in the gear structure, making it more susceptible to faults.
- Material issues: The use of inappropriate or low-quality materials for gear production can cause localized weaknesses, such as inclusions, voids, or micro-cracks, which can increase the likelihood of gear faults, such as tooth breakage or surface damage.
- Improper gear design: Inadequate gear design can lead to excessive stress on gear teeth, resulting in tooth bending fatigue, breakage, or other faults. Factors such as gear size, tooth geometry, and load capacity should be carefully considered during the design phase to minimize the risk of gear faults.
- Operational factors: Excessive load, high speeds, or sudden changes in load can cause increased stress on gear teeth, leading to accelerated wear, pitting, scuffing, or even tooth breakage. Ensuring that the gear system is operated within its specified limits can help prevent gear faults.

- Misalignment: Incorrect alignment of gear axes can result in uneven load distribution and increased stress on gear teeth. This can lead to accelerated tooth wear, pitting, eventually causing gear faults.
- Backlash issues: Excessive or insufficient backlash (the clearance between mating gear teeth) can cause increased tooth stress, noise, and vibration, leading to premature gear wear and failure.
- Temperature fluctuations: Extreme temperature fluctuations can cause thermal expansion and contraction of gear components, leading to misalignment, increased stress on gear teeth, and potential gear faults.
- Inadequate lubrication: Insufficient or improper lubrication can cause increased friction, wear, and heat generation in the gear system. This can lead to various gear faults, such as scuffing, pitting, and tooth breakage.
- Contamination: The presence of abrasive particles, dirt, or debris in the lubricant can accelerate gear tooth wear and increase the likelihood of surface damage, such as pitting or spalling.
- Poor maintenance: Lack of regular inspection, maintenance, and replacement of worn or damaged components can exacerbate existing gear faults or cause new ones to develop.

Figure 2.4: gear fault types(Jiao *et al.*, 2018).

2.2 Condition Monitoring

Condition monitoring refers to the process of continuously monitoring the health and performance of a system or equipment, such as machinery or vehicles, to detect any signs of deterioration, wear, and tear, or impending failures. Predictive maintenance is becoming more and more interesting and even necessary for many industries to be cost-effective. The goal of condition monitoring is to prevent unplanned downtime, reduce maintenance costs, and improve the overall performance the system. The demand for higher efficiency and lower environmental impact of transmissions in the automotive and gearbox industries is on the rise, creating a need for advanced technical solutions that can fulfill these requirements (Muralidharan and Sugumaran, 2013; Salameh *et al.*, 2018; Zhu *et al.*, 2014). In this context, condition monitoring is a crucial aspect of the transmission life cycle, saving valuable resources and time. With the recent advancement in technology, condition monitoring has shifted from a reactive to a proactive approach, using machine learning techniques to predict and detect minor faults before they evolve into significant issues.

Many conventional approaches have been investigated to monitor rotation machinery by studying some important variables which can be categorized into four main variables namely, Vibration, Acoustic, Lubrication, and temperatures (Isermann, 1993). The most effective technique among these is vibration analysis as the vibration signal carries most of the information that is related to the condition of the rotating machinery (Edwards *et al.*, 1998).

2.2.1 Acoustic Analysis

Acoustic analysis is a field that deals with studying the properties of sound. Its application in industries involves the evaluation of sound emissions from machinery, equipment, and processes. The information obtained from sound analysis is critical in detecting potential problems, diagnosing issues, and optimizing performance. One of the primary benefits of acoustic analysis in the industrial field is predictive maintenance. It enables engineers to identify underlying issues with machines by analyzing the sound they produce, which can prevent significant problems before they occur. Predictive maintenance reduces unexpected downtimes, cuts maintenance costs, and ultimately improves the efficiency of the industrial process. In addition, acoustic analysis is also valuable in quality control. Engineers can assess the sound output of products, ensuring that they meet quality standards and are devoid of faults. Overall, acoustic analysis is a crucial aspect of the industrial field that improves performance, reduces expenses, and ensures that products conform to set standards (Wu *et al.*, 2019; Zhang *et al.*, 2020).

2.2.2 Lubrication Analysis

Lubrication analysis is a technique used to evaluate the condition and performance of lubricants in machinery and equipment. The process involves analyzing oil samples from machines and examining the physical and chemical properties of the lubricant. By analyzing the oil sample, engineers and technicians can identify potential problems and make recommendations for maintenance or lubricant replacement. Lubrication analysis is a critical aspect of predictive maintenance, which involves the detection of potential issues before they cause significant damage to machinery. By monitoring the

condition of the lubricant, engineers can identify issues such as contamination, wear, and degradation of the lubricant, which can lead to poor performance and equipment failure. Lubrication analysis can also help to optimize machinery performance and reduce maintenance costs. By identifying potential issues early, maintenance can be scheduled at convenient times, which reduces downtime and maintenance expenses. Additionally, the analysis can help identify the ideal lubrication requirements for specific machines, leading to better performance and longevity. In conclusion, lubrication analysis is a valuable tool in the field of predictive maintenance, which helps to reduce maintenance costs, optimize machinery performance, and increase the lifespan of the equipment (Belkacemi *et al.*, 2020).

2.2.3 Temperature Analysis

The assessment of temperature fluctuations in different industrial processes, equipment, and systems is known as temperature analysis. This process involves tracking temperature changes over time by measuring and analyzing temperature readings. The aim is to identify possible issues, optimize performance, and enhance efficiency. In predictive maintenance, temperature analysis plays a crucial role by enabling engineers and technicians to monitor equipment and systems for temperature changes that could be an indication of impending issues. For instance, if there is a sudden rise in temperature in a bearing, it could imply that the bearing is overheating due to friction or insufficient lubrication, which could lead to failure if not addressed promptly. Temperature analysis is also valuable in optimizing performance and reducing energy consumption. By analyzing temperature data, engineers can identify areas for adjustment in processes, equipment, and systems to operate at optimal

temperatures. Consequently, this leads to improved efficiency and energy cost reduction (Karabacak *et al.*, 2022).

2.2.4 Vibration Analysis

Among the array of techniques used for condition monitoring, vibration analysis stands out as a powerful and non-intrusive tool. Vibration signals serve as sensitive indicators of machinery health, capable of detecting early signs of faults and deviations from normal operation because the vibration signal carries most of the information that is related to the condition of the rotating machinery (Edwards *et al.*, 1998).

By continuously monitoring vibration patterns, maintenance teams can proactively address potential issues before they escalate into severe problems, leading to improved performance of the system and reduced downtime. The maximum vibration will be observed if there are damages in both contacting teeth and the gear frequencies are mostly located in the low frequency (Abdul *et al.*, 2020). The vibration signal analysis approach can be conducted in different domains including time, frequency, and time-frequency which are useful in terms of fault detection (Abdul and Talabani, 2022).

2.3 Gear Condition Monitoring Based on Vibration Signal.

Gear fault diagnostics refers to the process of identifying a gear's faults or conditions based on observable symptoms. Fault detection requires the skill to recognize a machine's condition from its symptoms, much like medical diagnosis. In the context of gearboxes, vibration is often seen as a symptom of their condition. Despite the complex structure of the vibration generated by

gearboxes, it can provide valuable information for analysis. Condition monitoring in early works of gear fault diagnosis was conducted using the visual inspection of the vibration signal and it is limited to those faults that even observe from the stationary signals.

To overcome this limitation, statistical analysis has been used for gear fault detection. However, there is no guarantee to identify all kinds of gear faults by statistical analysis as the pattern of some faults cannot be observed by statistical property. Nowadays, Machine learning algorithms play a vital role to detect the condition of gears. Therefore, they have been adopted for detecting unknown patterns of faults based on the extracted features (Wilk *et al.*, 2008).

2.3.1 Statistical Analysis

Statistical analysis is a method of data analysis that involves the use of statistical techniques and tools to analyze and interpret data. Statistical analysis based on the features, also known as analysis, based on non-time series features, is also an important aspect of gear fault detection, as they provide a summary of the overall characteristics of the gear's vibration signals. However, there are two main drawbacks of using non-time series features for fault detection, firstly non-time series features do not provide information about the dynamics and evolution of the gear's condition over time. Secondly, non-time series features might not be suitable to certain types of noise and interference(Al-Talabani, 2015).

2.3.2 Fault Detection-based Machine Learning.

Machine learning has brought up many significant improvements in various applications such as vehicle production systems (Luo and Wang, 2018), fault detection in industrial systems (Chen, Li, *et al.*, 2019), robotics (Zeng *et al.*, 2021), medical applications (Zeng *et al.*, 2018) and speech analysis (Gaikwad *et al.*, 2010). Machine learning-based fault detection can be defined as a model for detecting faults in systems and machines by utilizing machine learning algorithms to identify patterns in recorded data. The algorithms used in machine learning-based fault detection are trained to recognize patterns in data that indicate a fault or malfunction. Using this information, the system can either alert an operator of the fault or take corrective action to repair the system. Machine learning-based fault detection can be used in a variety of applications including manufacturing, automotive, aerospace, and medical systems (Kumar and Hati, 2020).

2.4 Gear Fault Detection Based on Machine Learning.

Machine learning algorithms can be used for gear fault detection by recognizing patterns in data that indicate a fault or malfunction. There are some pre-requirements to develop a gear fault detection system based on Machine Learning including Feature extraction and classification model. Both sections are explained below.

2.4.1 Feature Extraction

Feature extraction is a challenging topic in any machine-learning system. Many feature extraction techniques have been developed to extract effective fault information in rotating machinery. Due to the different transmission characteristics of the machine, the vibration features of rotary machines are unique (Aherwar, 2012). In the gear fault frequency domain-based features, various peaks indicate different types of defects. For gear problems, particular significance is attached to the Fast Fourier Transform (FFT) spectrum defect frequencies (Decker and Lewicki, 2003). FFT spectra produces peaks at the specified fault frequencies. These peaks define the stated defects. Harmonics must be also considered to work out whether the identified frequencies have been produced from the indicated fault (Sharma and Parey, 2016). If a peak shows up at the gear fundamental frequency and another peak shows up at two times the fundamental frequency, it is a powerful indicator that the fault is real. Another good indicator that the specified fault has occurred is that no peak shows up at the gear fundamental frequency. However, peaks do present at two, three, and four times the gear fundamental frequency (Cerrada *et al.*, 2017).

In FFT, comparing the amplitude of faulty and non-faulty signals is a well-established approach to detecting fault severity in gears, where a higher amplitude than normal indicates a problem (Dhamande and Chaudhari, 2016). This is exactly why the amplitude of various frequencies is considered an effective feature for fault detection in rotation machinery (Stamboliska *et al.*, 2015). To extract these kinds of information, bandpass filters are a useful adopted technique.

To track harmonic frequencies and extract all related information to the fault indicator, two feature extraction methods (MFCC and GTCC) are adopted. MFCC uses a filter bank that highlights the low frequencies (where gear

frequencies lie) over the higher frequencies. However, the accelerometer signal has a self-noise and MFCC is frequently reported to be sensitive to noise. Therefore, we have proposed the use of the GTCC feature which is reported to be more robust with noisy signals and overcome MFCC in various speech applications (Revathi *et al.*, 2014)(Dimitriadis *et al.*, 2011). There are two beneficial steps in the extraction of both proposed features, which are splitting the signals into some frames (windowing signal) and applying a sequential band-pass filter to each frame signal. Windowing ensures the stationary of the signal in each window where the extraction of spectral features is more accurate. Both feature extractions (MFCC and GTCC) are explained in the subsection below.

I. Mel-Frequency Cepstral Coefficients

MFCC is one of the reliable feature extraction techniques that has been used a lot for audio analysis (Tiwari, 2010). Recently, it has been used for fault detection and diagnostics and it shows an ability to extract defect information (Andrew F. Geib, Chung Chieh Kuo, Martin Gawecki, EnShuo Tsau, Je Won Kang, 2014; Zhang *et al.*, 2018). The common property of this technique is robustness in extracting linear and nonlinear properties from audio or vibration signals. However, it has a low level of noise immunity. Four steps are required to proceed MFCC technique after windowing the signal as listed below:

A. Discrete Fourier Transformer (DFT) is computed for each signal frame, and it can be computed by the below equation (2.1)

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{2\pi jnK}{N}} \qquad k = 1,2,3 \dots N-1 \qquad (2.1)$$

where $x(n)$ is the discrete signal and N is the length of the signal.

B. Applying Mel filter bank on the obtained power spectrum. Usually, the filter bank is constructed based on 40 triangular filters and the transfer function filters can be computed by equation (2.2) and shown in Figure 2.5.

$$H_n(L) = \begin{cases} 0 & L < f(n-1). \\ \dfrac{L - f(n-1)}{f(n) - f(n-1)} & f(n-1) \le L < f(n). \\ 1 & L = f(n). \\ \dfrac{f(n+1) - L}{f(n+1) - f(n)} & f(n) < L \le f(n+1). \\ 0 & L > f(n+1). \end{cases} \quad (2.2)$$

where, $f(n)$ is the midpoint frequency of the triangular filter and $\sum_n^{N-1} H_n(L) = 1$. The mathematical expression relating the Mel scale to the response frequency or vice versa can be computed by equations (2.3) and (2.4).

$$n = 2595 log_{10}\left(1 + {}^{f}/_{700}\right). \quad (2.3)$$

$$f = 700\left(10^{{}^{n}/_{2595}} - 1\right) \quad (2.4)$$

C. A logarithm is calculated for each filtered power spectrum.

D. Applying Discrete Cosine Transform (DCT) to obtain 1-14 coefficients. The DCT can be computed by the below equation (2.5).

$$X(k) = \sum_{n=0}^{N-1} x_n * COS\left(\,{}^{2\pi jnk}/_{N}\,\right), \qquad k = 1,2,3 \ldots N-1 \quad (2.5)$$

where x_n is a discrete signal and N is the length of the signal.

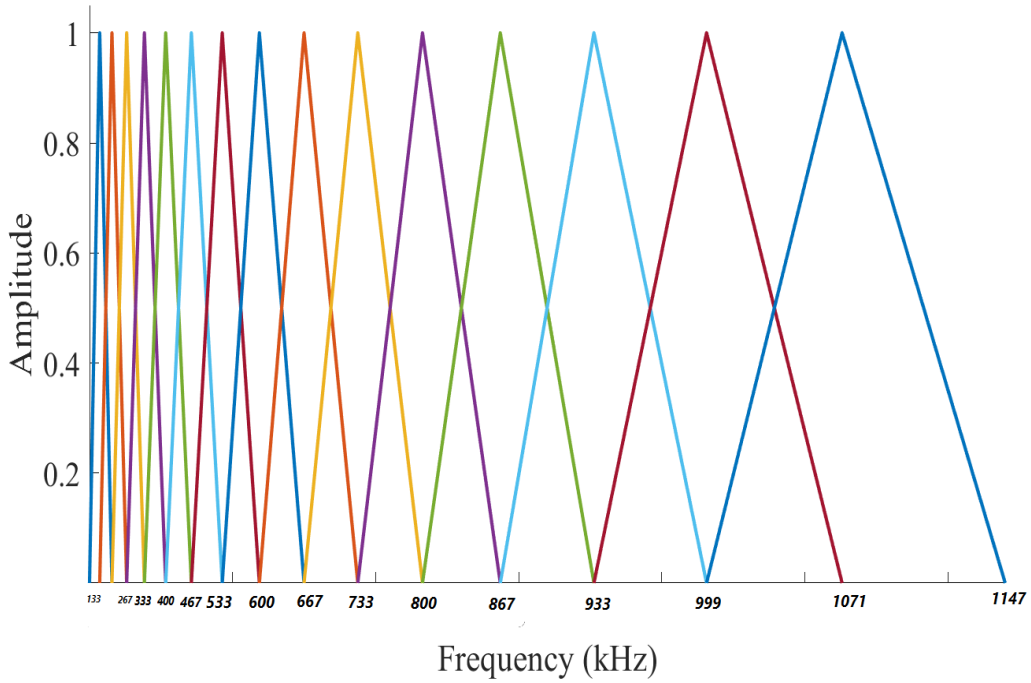

Figure 2.5:Mel frequency filters(Abdul *et al.*, 2020).

II. Gammatone Cepstral Coefficients

GTCC is a sisterhood technique with MFCC and has been developed for extracting features from audio signals. Most of the processes of conducting MFCC and GTCC are the same including applying DFT, calculating the logarithm for each of the filtered power spectrum and applying the DCT as shown in Figure 2.6. The main difference between them is the type of filter bank that is used for obtaining filtered power spectrum. A gammatone filter bank is used in the GTCC technique. The filter bank consists of 32 filters and each of the filters is described by an impulse response, which is composed of the product of sinusoidal tone and the gamma distribution (Adiga *et al.*, 2013). The impulse response or Gamma-tone filter is computed using the equation (2.6) and shown in Figure 2.7.

$$g(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c * t + \mu). \quad (2.6)$$

where, f_c is the midpoint frequency and μ is the phase, which is normally set to be 0. n is the degree of freedom of the filter which is typically set to be equal to or less than 4. a is a constant used to control the gain, and lastly, b is a factor associated to f_c and is formulated by equation (2.7).

$$\mathrm{b} = 1.019 * 24.7 * \left(4.37 * \frac{f_c}{1000} + 1\right). \quad (2.7)$$

Different f_c leads to a set of Gamma-tone filters which are named a Gamma-tone filter bank. The Gamma-tone bank filter is suitable to extract the characteristics of the signal at various frequencies (Qi *et al.*, 2013).

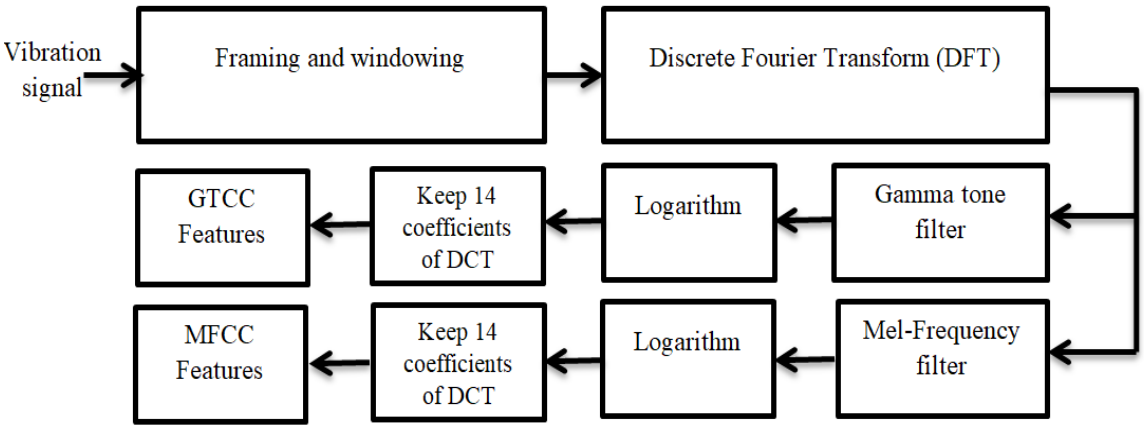

Figure 2.6: Feature extraction process for GTCC and MFCC(Abdul *et al.*, 2020)

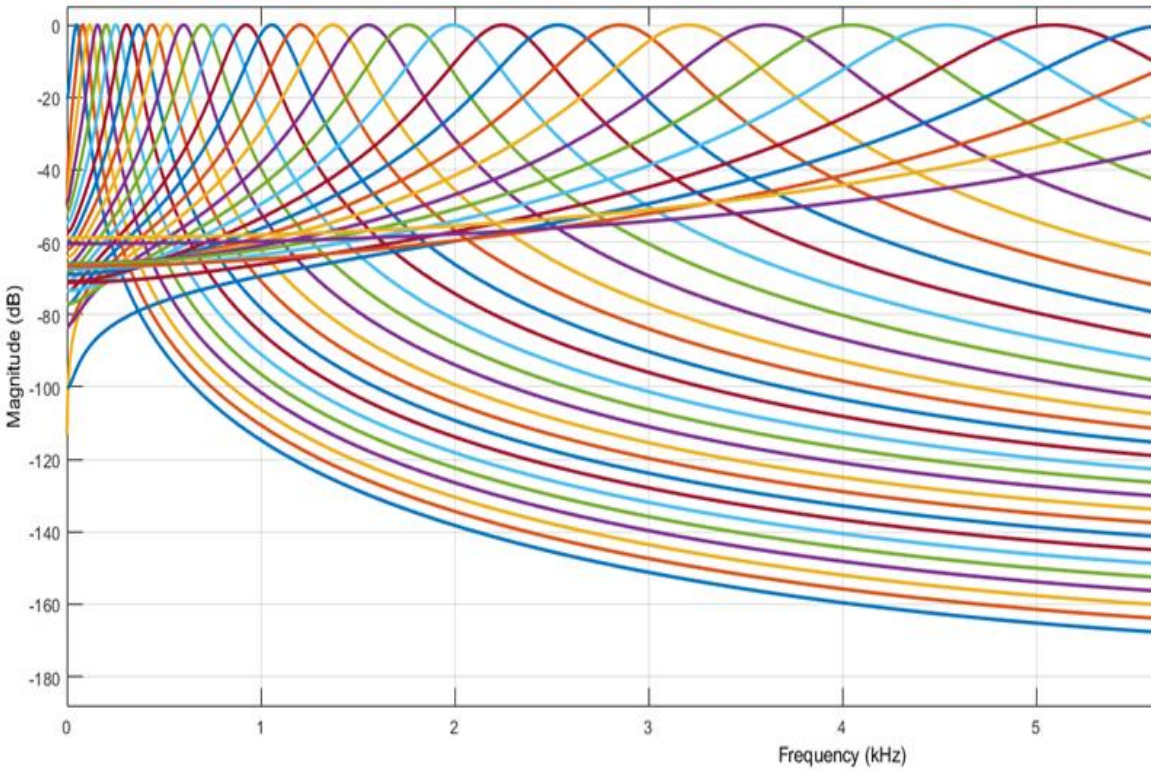

Figure 2.7: Gamma-tone filters(Matlab, 2022)

2.4.2 Classification Model

A classification model is a type of machine learning model that aims to categorize or classify input data into distinct classes or categories. In the context of supervised learning, a classification model is trained using labeled data, where each data point is associated with a specific class or category. The trained model can then predict the class of new, previously unseen data points based on the patterns and relationships it has learned during the training process. In this work, we have used two kinds of supervised learning algorithms based on varying features over a time period namely time series model and none-time series model.

A. Long Short-Term Memory (LSTM)

The LSTM is an improvement to the Recurrent Neural Network (RNN). The LSTM has a more complicated repeating module in the network structure. The main difference between the traditional RNN and LSTM is that the LSTM can capture long-term dependencies (Yang *et al.*, 2018). It adopts a structure that can exceed the problem of gradient vanishing. The LSTM is a deep learning method that classifies and regresses time-series data such as voices and vibrations in consideration of feature changes at each time step (Yoshimatsu *et al.*, 2018). The structure of the LSTM includes four parts in each repeating module; cell state (c), input gate (i), forget gate (f) and output gate (o) (see Figure 2.8). The cell state c_t at time t is computed using the following equation (2.8):

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{2.8}$$

where c_{t-1} is the previous cell state, $\odot$ denotes elementwise multiplication and f_t, i_t and g_t are computed using the equations listed below:

$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \tag{2.9}$$

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \tag{2.10}$$

$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \tag{2.11}$$

where σ_c and σ_g are gate activation functions, choosing a tanh and sigmoid function respectively. The parameters W, R and b are the input weights, the recurrent weights, and the bias of each component, respectively.

The hidden state is also updated using equation (2.12):

$$h_t = o_t \odot \sigma_g(c_t) \tag{2.12}$$

where:

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \quad (2.13)$$

The cell state is responsible for remembering a value during the recurrent connection. Updating the remembered values and then forgetting them is crucial in the learning process of the module. The updated input value remembers values in the memory, while the forgetting gates determine when the remembered input is no longer important. The output gate is capable of knowing when the cell state gives the output value. The output of the previous step during the computation at each gate and cell state will be an input for the next step. Hence, the LSTM module learns how to maintain its memory as a function of previous values. It is reported that the LSTM can capture defect gear information based on a temporal feature that was combined with MFCC and GTCC features. However, the time consumption for training the LSTM model is quite high (Abdul et al., 2020).

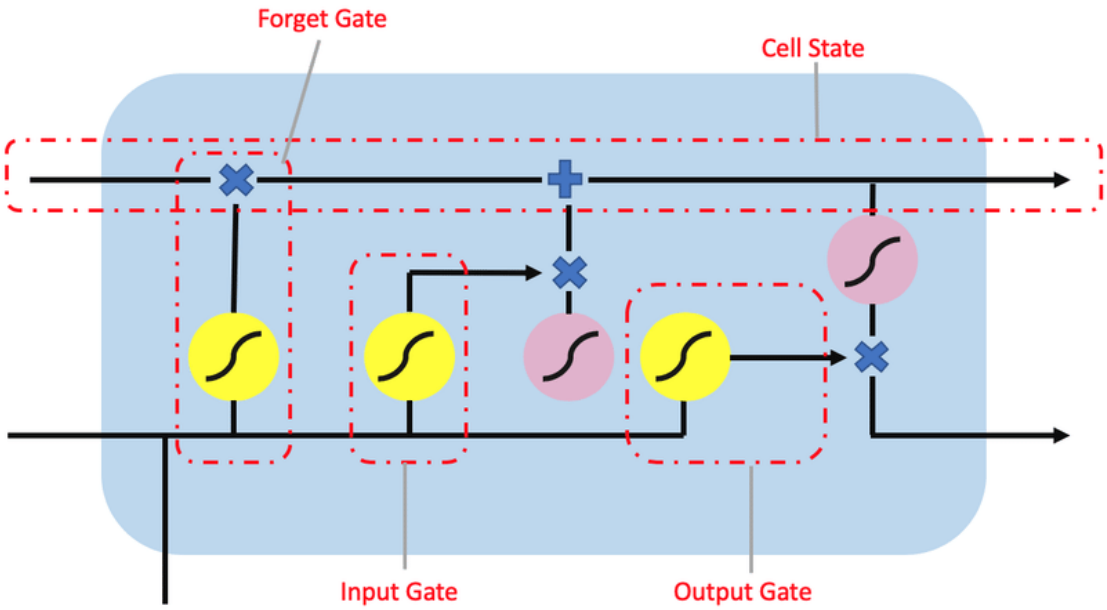

Figure 2.8: LSTM node (Xiang *et al.*, 2020)

B. Echo State Network (ESN)

Another popular progress of the RNN model is Echo State Network, which was invented by Jaeger and Haas (Jaeger and Haas, 2004). Similar to the LSTM model, the ESN is able to overcome the vanishing gradient problem that leads to reduced training time significantly compared to the RNN and LSTM (Dai *et al.*, 2009). The structure of ESN consists of three main sections, input, reservoir, and readout section as shown in Figure 2.9. The weights of the input section are fixed and completely selected randomly. The connection of the reservoir nodes is sparse random connections, and their weights are fixed and random as well. The only weights, which need to be trained, are the readout weights which connect the reservoir to the output neurons. The training time is reduced significantly as the weights of two of the three sections are assigned untrained and selected randomly. In the training stage, the inputs are fed to the reservoir and the output of the reservoir (teacher output) will be fed to the output layer (Ismail Fawaz *et al.*, 2019). It is obvious that the reservoirs include sparse random connections, which help to "echo" the previous states. Hence, when a new input that is similar to something it trained on reaches the network reservoir, the reservoir will dynamically start to follow the activation trajectory which is suitable for the input and consequently can provide a matching signal to what it trained on. The generalization will be performed if it is well-trained from what it has already seen, following activation trajectories that would make sense given the input signal driving the reservoir (Inc, 2023).

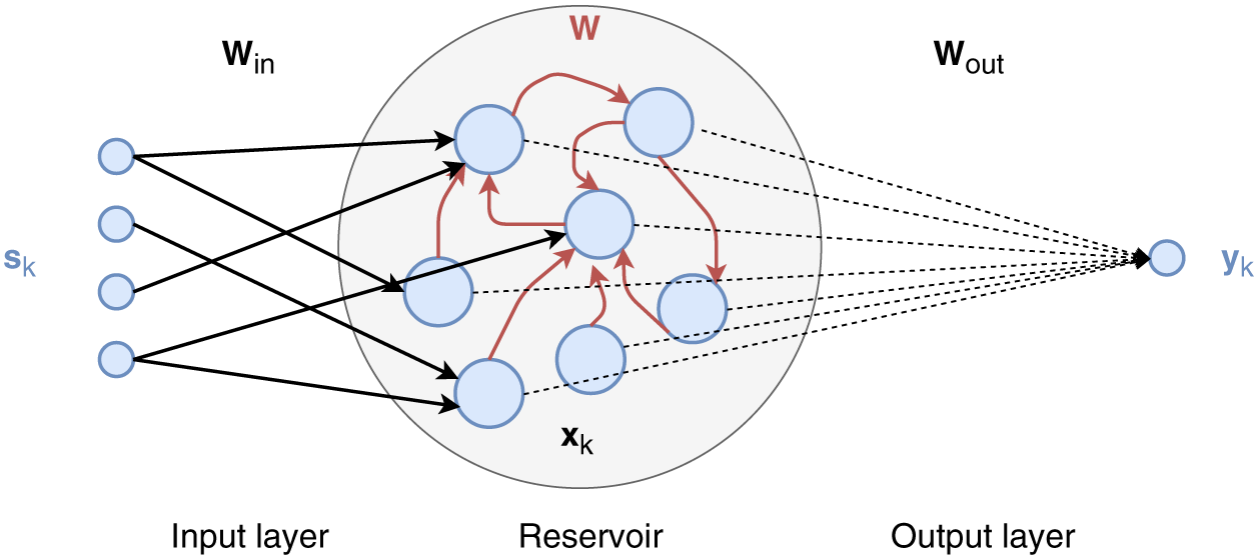

Figure 2.9: ESN structure (Verzelli et al., 2019)

C. Support Vector Machine

One of the powerful machine learnings is the support vector machine that was developed by Vapnik (Farhat, 1992). There is a similarity between single-layer perceptron and SVM as they separate classes linearly, however, the SVM provides the line, the plane, or the hyperplane with maximum margin. The SVM aims to identify an optimal hyperplane between binary classes as shown in Figure 2.10. The formula of the hyperplane can be set as follows:

$$\begin{cases} w^T x_i + b \geq 1 & for\ y_i = +1, \\ w^T x_i + b \leq 1 & for\ y_i = -1, \end{cases} \tag{2.14}$$

where, (x_i, y_i) is the training data and its labels, the bias b is a scalar, and w is the weight vector. The object equation (2.15) with a constraint is used for maximizing the margin of the hyperplane.

$$\begin{aligned} &Minimize && \phi(w) = {}^{1}/_{2} * w^T w \\ &Subject\ to && y_i(w^T x_i + b) \geq 1 - \zeta \quad \zeta \geq 0, i \\ & && = 1,2, \dots, n \end{aligned} \tag{2.15}$$

Then by conducting the Lagrange optimization methods, a set of multipliers are introduced including αi, βi for constraints. The objective function and the constrained optimization problem become like equations (2.16) and (2.17) respectively.

$$Minimize\ \phi(\alpha) = \sum_{i=1}^{n} \alpha_i - {}^{1}/_{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x^T x_j \tag{2.16}$$

$$Subject\ to \quad \sum_{i=1}^{n} \alpha_i y_j = 0, \qquad 0 \leq \alpha_i \leq C\ , i = 1,2, \dots, n \tag{2.17}$$

The corresponding data points α_i are named as support vectors if $0 \leq \alpha_i \leq C$.

Many state-of-arts studies have reported that the SVM is a very influential classifier in high-dimensional problems that leads to achieving good accuracy in the high-dimensional feature space (Al-Talabani *et al.*, 2015). In this dissertation, the dimensionality of the feature space is increased by concatenating all of the frames of both MFCC and GTCC. Then the obtained feature is used to train the SVM classifiers.

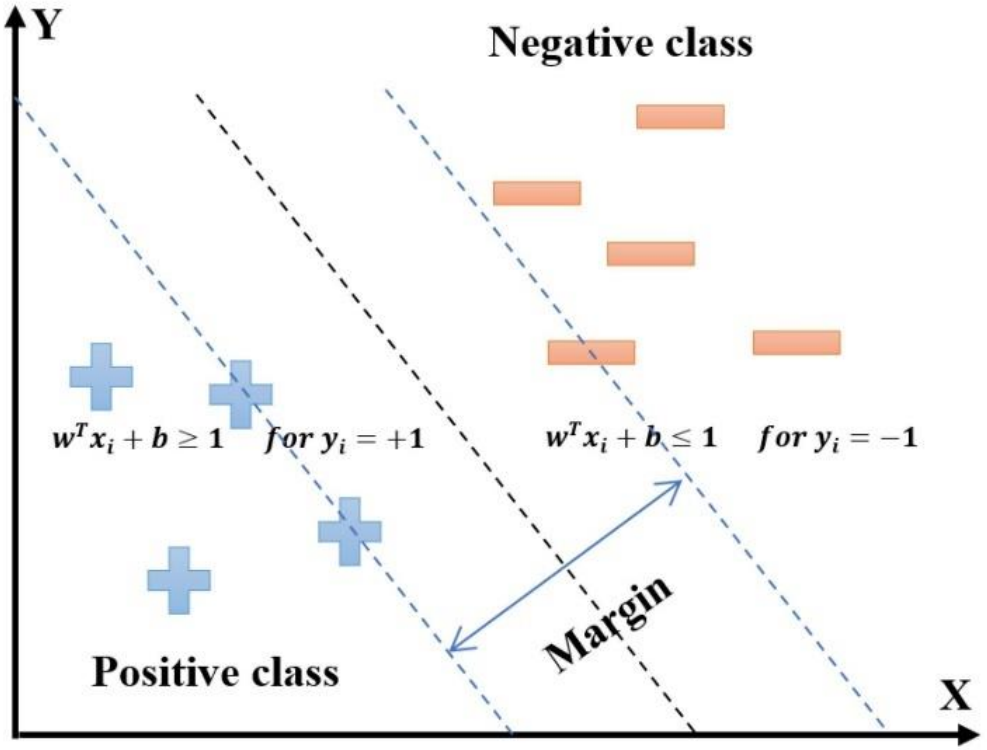

Figure 2.10: Classification of data by SVM

SVMs can handle both linear and non-linear data by using kernel functions, which transform the input data into a higher-dimensional space where classes can be separated more easily. Four common kernel functions have been used in SVMs including linear kernel, polynomial kernel, Radial Basis Function (RBF), and Sigmoid Kernel

Choosing the appropriate kernel function for an SVM depends on the characteristics of the data and the problem being solved. The linear and RBF kernels are widely used due to their effectiveness and flexibility, while the polynomial and sigmoid kernels might be more suitable for specific problem domains or when prior knowledge about the data structure is available.

D. Optimization Algorithm for Tuning Hyperparameters

Optimization algorithms can be categorized into two main types namely deterministic and stochastic. Deterministic algorithms follow mathematical formulae and computational simulations and have been used to solve many optimization problems. However, deterministic algorithms are sometimes stuck in the local minimum. The issue has been addressed in metaheuristic algorithms, which are kinds of stochastic algorithms, by balancing between randomization and local search (Abdullah, 2019). Many metaheuristic algorithms have been developed by researchers to find out a perfect method to deal with limited resources problems within various restrictions. Recently, metaheuristic algorithms have been utilized in various fields such as engineering problems, browsing the Internet, financial profit, bioinformatics, geophysics, and optimizing features for pattern recognition applications (Salih *et al.*, 2022).

Feature optimization can be done either by selecting a proper feature set, which is called feature selection or by optimizing the parameter within the feature technique itself. In 2020, Manik Sharma and Prableen Kaur reviewed 176 articles about the use of metaheuristic algorithms for the feature selection process and the metaheuristic algorithms were classified by their inspiration like inspired birds, animals, humans, insects, and creatures (Sharma and Kaur, 2021). Ibrahim et al. proposed an optimal features selection mechanism using the Invasive Weed Optimization algorithm (IWO) for diagnosing the faults under different load conditions. In the beginning, Matching Pursuit (MP) and Discrete Wavelet Transform (DWT) were extracted, and some statistical features were adopted from the MP and DWT. Then, the IWO was applied to select the optimal feature set for diagnosing the faults (Ibrahim *et al.*, 2022).

Regarding optimizing the parameter inside the feature techniques, the initial parameters of the wavelet neural network were optimized using Improved Grey Wolf Optimization (IGWO) for addressing the problem of diagnostic accuracy and stability degradation. The optimizing initial parameters of WNN led to achieving higher performance in detecting fault by an amount of 1.15% higher compared to WNN where its parameters are selected randomly (Pan *et al.*, 2022). Zhou Guifan optimized all parameters in variational modal decomposition (VMD) for bearing fault detection using a hybrid metaheuristic algorithm in which the idea of the algorithm has come from two other algorithms including a nondominated sorting genetic algorithm and multi-objective particle swarm optimization. The result shows that the optimized VMD is more robust for detecting faults compared with the conventional VMD (Guifan, 2022).

In summary, the purpose of optimization algorithms is to find the best possible solution to a given problem by systematically searching through the solution space. Tuning hyperparameters is an essential step in the process of training machine learning models. Hyperparameters are the settings of the learning algorithm that determine its performance and generalization capabilities. Optimization algorithms can be used to search for the best hyperparameter values that lead to improved model performance. Many optimization algorithms have been developed which have been used to solve plenty of optimization problems including tuning hyperparameters in condition monitoring systems.

I. Bayes Optimization

Bayesian optimization is a global optimization technique used for finding the maximum or minimum of an objective function that is expensive to evaluate

or does not have an easily derivable gradient. It is particularly useful in situations where evaluating the function is time-consuming, costly, or noisy. Bayesian optimization is based on the principles of Bayesian statistics and incorporates prior knowledge about the objective function to guide the optimization process. There are two main components of Bayesian optimization. Firstly, the Surrogate model: A surrogate model is used to approximate the objective function, which is often a Gaussian Process (GP) or a Bayesian Neural Network. The surrogate model captures the uncertainty in the function evaluations and is computationally cheaper to evaluate than the actual objective function. Secondly, Acquisition function: The acquisition function is a utility function that guides the optimization process by determining the next point to sample in the search space. It balances exploration (searching in areas with high uncertainty) and exploitation (sampling near the current best solution). Some popular acquisition functions include Expected Improvement (EI), Probability of Improvement (PI), and Upper Confidence Bound (UCB) (Ibrahim *et al.*, 2021a). The Bayesian optimization process generally follows these steps:

1. Define the objective function: The objective function is the target function you want to optimize.
2. Choose a surrogate model: Select an appropriate surrogate model to approximate the objective function, e.g., Gaussian Process or Bayesian Neural Network.
3. Define an acquisition function: Choose an acquisition function to guide the exploration-exploitation trade-off during the optimization process.
4. Initialize with a set of points: Start with a few initial points in the search space, which can be selected either randomly or using prior knowledge.
5. Iterate until convergence or a stopping criterion is met:

a. Fit the surrogate model to the available data.
b. Find the next point to sample by maximizing the acquisition function.
c. Evaluate the objective function at the selected point and update the dataset.
d. Repeat steps a-c until the optimization process converges or a predefined stopping criterion is met.

Bayesian optimization has been successfully applied to various optimization problems, including hyperparameter tuning in machine learning models, experimental design, and optimization of expensive simulations.

II. Grey Wolf Optimization

GWO is a type of swarm intelligence that mimics the hunting behavior and hierarchical leadership of the grey wolf. These wolves work in groups of 5 to 12 members. The GWO has been compared with different metaheuristic algorithms and it shows that GWO has advantages over others in terms of simplicity in operations, convergence speed, and achieving accurate solutions. The hierarchical leadership of wolves is divided into four categories of wolves which are alpha (α), beta (β), delta (δ), and omega (ω). The leader of the wolves is called alpha while supporters are called beta. Then, delta is a follower type that follows the instructions of alpha and beta. The final one is called omega which looks after the group requirements (see Figure 2.11). (Seema and Kumar, 2016), (Mittal *et al.*, 2016)(Mirjalili *et al.*, 2014).

In the beginning, GWO tries to use encircle mechanism to chase the prey after the initialization of the population. The encircling mechanism of the prey can be represented by the following equations (Mohammed *et al.*, 2021):

$$\vec{D} = \left| \overrightarrow{C_1} \cdot \overrightarrow{X_{p(\iota t)}} - \overrightarrow{X_{\iota t}} \right|,$$
$$\overrightarrow{X_{(\iota t+1)}} = \overrightarrow{X_{p(\iota t)}} - \overrightarrow{A} \cdot \vec{D} \ , \tag{2.18}$$

Distance $\vec{D}$ can be found by using the current position and the position of the other three types of wolves (alpha, beta, and delta). These positions are represented by $X_{p(it)}$. Where $X_{p(it)}$ is the current position and $X_{(it+1)}$ is the new position that can be found based on the distance equation for each type of wolf. $\vec{A}$ and $\vec{C}$, are coefficient vectors that are calculated as follows(Mohammed *et al.*, 2021):

$$\vec{A} = 2 \cdot \vec{a} \cdot \overrightarrow{r_1} + \vec{a} \tag{2.19}$$

$$\vec{C} = 2 \cdot \overrightarrow{r_2} \tag{2.20}$$

Where $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are random numbers that are distributed in [0,1] and a is the control parameter that is decreased linearly from 2 to 0 throughout iterations. After the encircling mechanism, the hunting method is done by using alpha, beta, and delta wolves. GWO assumes that each type of wolf has enough knowledge about the position of the prey, so they are participating in the hunting process to find the new position. So, this method can be done by using the three best solutions to update the current position. Mathematical equations are presented as follows:

$$\overrightarrow{X}(it+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{2.21}$$

Each of $\overrightarrow{X_1}, \overrightarrow{X_2}, and\ \overrightarrow{X_3}$ is calculated by using equations (2.26,2.27 and 2.28) (Mohammed *et al.*, 2021):

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \overrightarrow{D_\alpha} \ , \tag{2.22}$$

$$\overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \overrightarrow{D_\beta} \ , \tag{2.23}$$

$$\overrightarrow{X_3} = \overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot \overrightarrow{D_\delta} \tag{2.24}$$

Where $X_\alpha, X_\beta \ and \ X_\delta$ is the position of alpha, beta, and delta wolves. $\overrightarrow{D\alpha}, \overrightarrow{D_\beta} \ and \ \overrightarrow{D_\delta}$ the distance can be found by using the following equations (2.29) in order to hunt the prey(Mohammed *et al.*, 2021):

$$\begin{aligned} \overrightarrow{D\alpha} &= \left| \ \overrightarrow{C_1} \cdot \overrightarrow{X\alpha} - \overrightarrow{X_{\imath t}} \ \right|, \\ \overrightarrow{D_\beta} &= \left| \ \overrightarrow{C_2} \cdot \overrightarrow{X_\beta} - \overrightarrow{X_{\imath t}} \ \right|, \\ \overrightarrow{D_\delta} &= \left| \ \overrightarrow{C_3} \cdot \overrightarrow{X_\delta} - \overrightarrow{X_{\imath t}} \ \right| \end{aligned} \tag{2.25}$$

Decreasing the value of $\vec{a}$ from 2 to 0 has a great effect on $\overrightarrow{A}$ because it has a range value between [-a, a]. As a result, GWO attacks the prey if the absolute value of $\vec{A} < 1$. Although, if the value of $\left|\vec{A}\right| > 1$, the wolves explore new areas (Mirjalili *et al.*, 2014; Panda and Das, 2019).

Algorithm 1 illustrates the GWO pseudo-code.

Algorithm 1: Grey Wolf Optimizer Pseudo-code

Initialize population *Xn (n=1, 2, ..., m), Xα Xβ and Xδ*

initialize *a and A* and *C* vector

finding *fitness* value

While (*it< Maxiteration*)

For each search agent

Update the location of omega search agent by Equation (2.29)

End For

Update a, A, and C

Fitness evaluation

Update X_{α}, $X\beta$ *and* $X\delta$

$it=it+1$

End while

return Xα

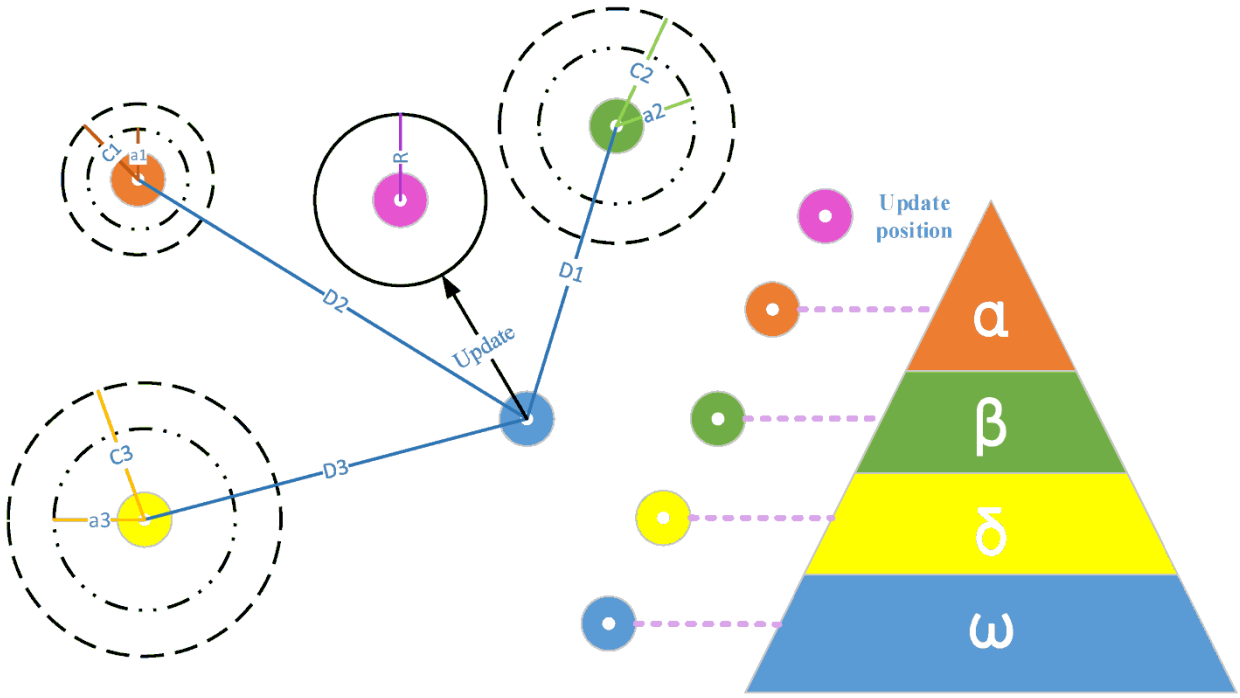

Figure 2.11: Gray wolf optimization process, where, alpha (α), beta (β), delta (δ), and omega (ω). (Dai et al., 2018)

III. Fitness Dependent Optimizer

The Fitness Dependent Optimizer (FDO) is one of the newest metaheuristic algorithms that has been developed by Abdullah and Rashid in 2019. The FDO working is based on a mechanism that bee swarming tries to find out food and it is slightly similar to the Particle Swarm Optimization (PSO) algorithm. The FDO is one of the metaheuristic algorithms that has been used by researchers to solve various applications, especially for engineering design problems (Salih *et al.*, 2022). As with all other metaheuristic algorithms, the FDO starts by initializing an artificial scout population in the search space randomly (see Figure 2.12). Each of the scout bees in the scout population represents a solution (hive) and the scout bee attempts to find a better solution by randomly searching for more positions. When a better solution is found, the previously discovered solution is ignored. So, in each iteration, the algorithm identifies a better new solution. However, if the new solution is not better than the previous solution, the FDO will continue with its previous solution and catch a better new solution. In Evert iteration, an artificial scout bee goes by adding pace to the current position and this movement of the scout bee can be calculated by equation (2.32)(Abdullah, 2019).

$$X_{i+1}=X_{i,t}+\text{pace} \tag{2.26}$$

where x denotes the scout bee, i represents the current scout bee, t is the current iteration, and pace value is dependent on Fitness Weight (fw) and random mechanism. The (fw) is obtained by equation (2.31)(Abdullah, 2019).

$$fw=\left[\frac{X^{*}_{i,t\ \text{fitness}}}{X_{i,t\ \text{fitness}}}\right]\text{-wf} \tag{2.27}$$

where wf is a weight factor, $X^{*}_{i,t\ \text{fitness}}$ and $X_{i,t\ \text{fitness}}$ represent the global best and current best solution respectively. FDO has three different conditions for

calculating pace firstly when the fw is equal to one value as it is illustrated in equation (2.34) (Salih *et al.*, 2022).

$$pace = X_{i,t} * r \tag{2.28}$$

Second, when the fw is equal to zero (see equation (2.35)).

$$pace = distance_{best\ bee} * r \tag{2.29}$$

Third, if fw>0 and fw<1. then the pace is calculated using equation (2.36).

$$pace = \begin{cases} \left(X_{i,t} - X^*_{i,t}\right) * fw^* - 1 & r < 0 \\ \left(X_{i,t} - X^*_{i,t}\right) * fw & r \geq 0 \end{cases} \tag{2.30}$$

where r is a random number between [0 1] and $distance_{best\ bee}$ indicates the variation in the current agent from the best agent, which can be calculated through equation (2.37) (Abdullah, 2019):

$$distance_{best\ bee} = X^* - X_{i,t} \tag{2.31}$$

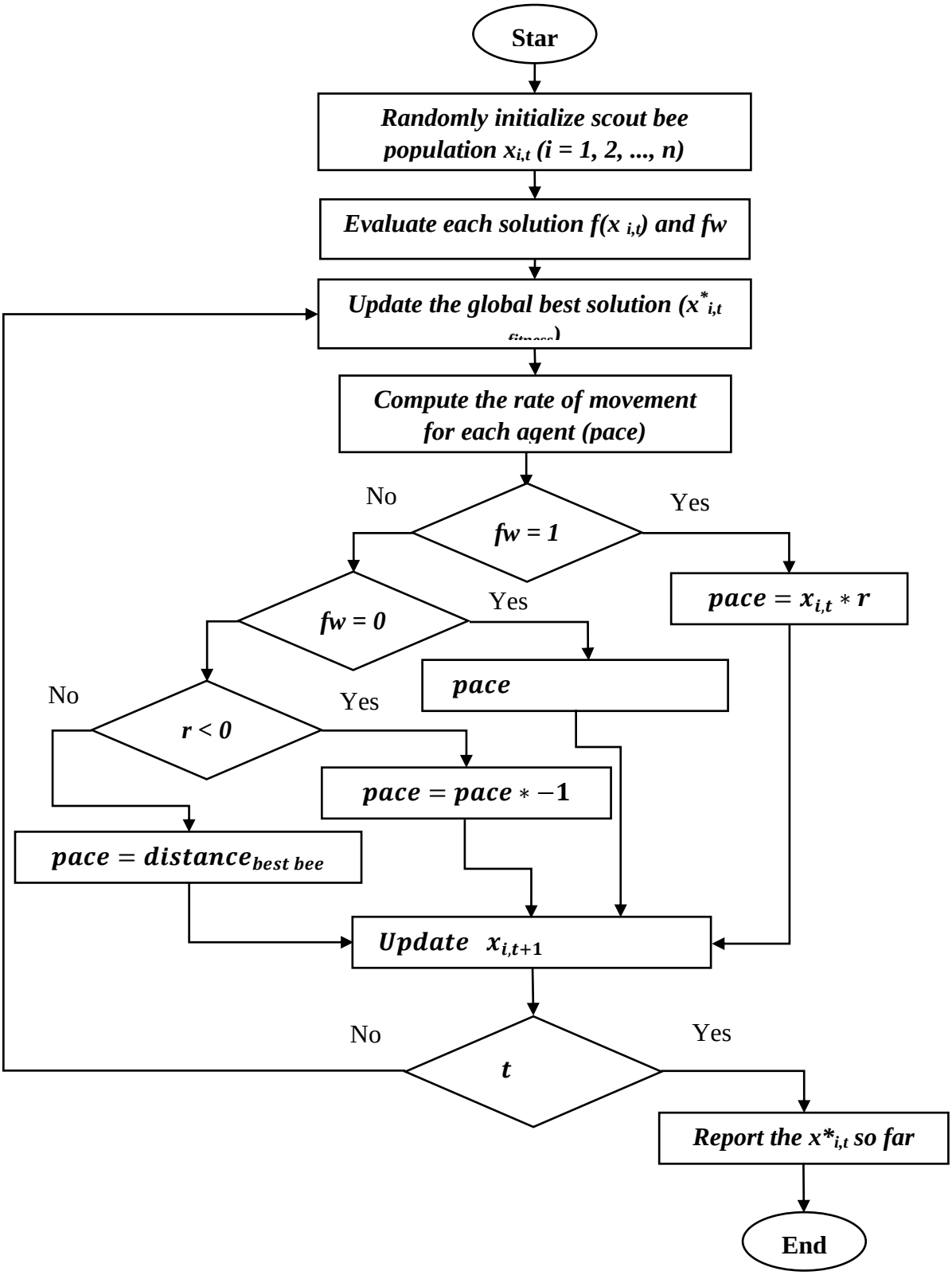

Figure 2.12: FDO flowchart

2.5 Related Works.

Based on the works that have been conducted in the area of machine learning-based gear fault detection, we can decompose the related work into two main sections namely fault detection based on time series model and non-time series model. Time series models and non-time series models are two different approaches to analyzing data, depending on the nature of the data.

2.5.1 Related Work Based on Time Series Models.

Lately, the application of Artificial Neural Network (ANN) has expanded in the domains of pattern identification and classification. Some benefits of these networks include deriving weights through neural computation and providing more objective diagnostic results compared to traditional techniques. Back-Propagation Neural Network (BPNN) is a type of these networks and has been extensively employed for fault diagnosis.

Sorsa et al. (Sorsa *et al.*, 1991) utilized a multilayer perceptron network featuring a hyperbolic tangent function, with 14 noisy measurement features and 10 faults as inputs and outputs, to identify faults in a real-world heat exchanger-continuous stirred tank reactor system. Li et al (Suphioglu *et al.*, 1998) employed frequency features of rolling bearings as neural network inputs to diagnose bearing defect types in motor-bearing systems. Kang et al. achieved optimal diagnostic results by extracting frequency features of vibration signals and using a fuzzy neural network (FNN) in motor systems(Kang *et al.*, 2006).

Although BPNNs or FNNs with frequency features can yield ideal diagnostic results, challenges arise when windowing signals for Fourier transformation, as the vibration signals may become distorted. Additionally, the frequency

spectrum distribution can be unclear for shock vibration signals, and determining true features can be difficult due to frequency sideband distributions in frequency modulation conditions. As a result, diagnostic results may not align with human expectations.

Furthermore, the accuracy of diagnostic results is influenced by the relation matrix between frequency features and fault types, which is established through expert knowledge or experimentation. The implementation of time series analysis can help address these limitations (Samanta, 2004).

Regarding time series features which are called temporal features, they have the potential to capture the dynamics and evolution of the gear's condition over time. These features can provide information about the gear's performance during the operation time (Fulcher, 2017). Many researchers proposed a gear fault detection system based on temporal features. For example, the standard RNN faces challenges in effectively training due to issues with gradient vanishing and exploding, despite its ability to handle time sequences. In contrast, the LSTM network, which incorporates forget, input, and output gates to filter information (Abdul *et al.*, 2020), overcomes these problems and is a popular variant of RNN. Similarly, the Gated Recurrent Unit (GRU) neural network utilizes reset and update gates to achieve better training results based on the same theory. Both LSTM and GRU are suitable for many applications, but they may not be ideal for building very deep structures or analyzing high-dimensional data (Li *et al.*, 2022).

In the field of machinery fault detection, simple and affordable models are desirable. To address this need, the ESN has emerged as a promising alternative to RNN. Instead of using hidden layer neurons, ESN employs a randomly generated reservoir as its basic processing unit (Jaeger and Haas, 2004). This approach allows for a simple structure and low data requirements, making ESN suitable for various applications, including emotion recognition (Ibrahim *et al.*,

2021b) and fault diagnosis (Abdul and Talabani, 2022). However, ESN may not be effective in mining deep or spatial information.

2.5.2 Related Work Based on Non-time series Models.

In the literature, one can find two modern ways to develop automatic fault detection models, one of which depends on extracting information from the time series signal (Cabrera, *et al*, 2017; Ismail Fawaz *et al.*, 2019a; Wang *et al.*, 2017), whereas the other way is to extract information from the non-time series representation of the signal (Chen, Liu, *et al.*, 2019; Yu and Zhou, 2020; Zamanian and Ohadi, 2016). The time series representation of the signal is adopted when the sequence of the information along the signal frames has important meaning to the application. On the other side, adopting the non-time series representation refers to the importance of the global information of the input signal rather than the order of the frame-based information (Cabrera *et al.* , 2017; (Ismail Fawaz *et al.*, 2019). One of the research questions that this study tries to investigate is the usefulness of these two representations for fault detection from vibration signal when a set of specific handcrafted features is adopted. Both of the approaches can be used together with the traditional machine leaning or deep learning algorithms. Since the sequence is not important in the non-time series representation, a question will be raised regarding the representation design of the features along various frames. Traditionally, statistics will be computed along the frames, which have the advantages of unifying the length and reducing the dimensions. However, statistics may lead to loss of information, in addition to the fact that the data along the frames may not have known probability distribution, which may lead to produce non-representative feature vector (Eyben *et al.*, 2009). These approaches can be applied on fault detection models using various types of

signals such as vibration, acoustic, lubrication and temperatures (Isermann, 1993). For fault detection in rotating machinery, the most effective signal among all is the vibration signal, since it carries significant information that are related to the condition of the rotating machinery (Edwards *et al.*, 1998).

Condition monitoring in early works of fault diagnosis is limited to those faults that can be observed from stationary signals. Moreover, power spectrum, cepstrum, adaptive noise cancellation, time-domain averaging, and time-series analysis are also established approaches in the time domain or frequency domain to effectively deal with stationary signals (Jena *et al.*, 2014)(Riaz *et al.*, 2017). However, there are some difficulties with these approaches related to some faults such as cracking gear teeth because they are incapable of disclosing the inherent information of the non-stationary signals, and some faults can be detected by analysis of the non-stationary signals. Therefore, there are new approaches for dealing with non-stationary signals such as time-frequency analysis (Wei *et al.*, 2019).

Due to the increase in the use of artificial intelligence, there have been significant improvements in most industrial technologies including fault detection in rotation machinery. Two methods of machine learning have been used for fault detection and isolation; traditional machine learning and deep learning (Saravanan *et al.*, 2008). Both machine learning algorithms have brought up many significant improvements in various applications such as vehicle production systems (Luo and Wang, 2018), fault detection in industrial systems (Chen, Li, *et al.*, 2019)(Kumar and Hati, 2020), robotic (Zeng *et al.*, 2021), medical application (Zeng *et al.*, 2018) and speech analysis (Gaikwad *et al.*, 2010). Traditional machine-learning algorithms are being criticized by researchers for requiring a handcraft feature (Lei *et al.*, 2019). However, using handcrafted features leads to developing models with less complexity compared to the deep leaned features.

Feature extraction is key to improving the performance of the machine learning-based model of fault detection and diagnosis. At the features level, some researchers have input statistics to the classification methods such as the mean, standard division, skewness, kurtosis and root mean square. For example, there was a fusion of some statistics and the adoption of a time derivative and a high-low pass filter. The fused features are selected using a genetic algorithm and then classified by SVM and an Artificial Neural Network (ANN) (Samanta, 2004). In (Saravanan *et al.*, 2008), the authors evaluated both the SVM and Proximal Support Vector Machines (PSVM) based on the statistical features obtained from morlet wavelet coefficients. While in (Rafiee *et al.*, 2010), four statistical values were adapted from continuous wavelet coefficients in the origin signal.

The DWT has been widely exploited and applied to the vibration signal for fault detection purposes. For example, the DWT feature is used with the Decision Tree (DT) classification in (Saravanan and Ramachandran, 2009). The normalized wavelet energy of post-fault voltage and current fault were computed together to feed the SVM classification (Livani and Evrenosoglu, 2012). In another study, a BPNN based on Clarke's transformation was trained by the DWT-based features (Asuhaimi Mohd Zin *et al.*, 2015). The DWT is also proposed to feed k-Nearest Neighbor (k-NN) (Manohar *et al.*, 2018). The results show that DWT is a powerful and flexible feature for decomposing linear and non-linear processing.

Although the Local Binary Pattern (LBP) is a well-known feature in image-based applications, it has also been adopted for fault detection. A two-dimensional LBP (2D-LBP) was extracted from textures in the 2D vibration images with the k-NN classifier to detect bearing faults (Khan and Kim, 2016). One-dimensional LPB (1D-LBP) was also employed to detect different types of gear faults. The 1D-LBP features obtained from the original signals are used

as input to the k-NN and SVM classifiers (Abdul *et al.*, 2016). The 1-D LBP algorithm was engaged to extract the LBP features of the wavelet coefficients for weld defect classification. Then the features were assessed by k-NN (Hu *et al.*, 2018). The LBP indicates the ability to extract effective and relevant features for detecting a fault in rotation machinery.

The MFCC is a well-known feature that has been widely used in speech analysis such as speech recognition, speaker identification and emotion recognition from speech (Abdul, 2019; Gaikwad *et al.*, 2010; Kathiresan and Dellwo, 2019). However, the MFCC is also proposed for fault detection and diagnosis. In (Zhang *et al.*, 2018), Zhang et al. used the MFCC, the wavelet packet energy decomposition and the zero-crossing rate features together with the SVM classifier. Researchers in (Benkedjouh T., Chettibi T., Saadouni Y., 2018) confirm that the defect information in rotation machinery is mainly included in the first three MFCC coefficients. The model has been evaluated using the SVM. Zhang et al. (Zhang *et al.*, 2018) extracted MFCC from the linear combination of wavelet packet decomposition and zero-crossing rate of vibration signal and then fed it to the SVM classifier. Benkedjouh et al. (Benkedjouh T., Chettibi T., Saadouni Y., 2018) extracted a set of features including some temporal indicators and MFCC coefficients then the SVM was utilized for classifying the gear and bearing faults. Based on the result of (Benkedjouh T., Chettibi T., Saadouni Y., 2018), defect information of gear fault is mainly included in the first three MFCC coefficients. However, other studies show that all of the 14 coefficients of MFCC include more relevant information to gear fault detection than the first three coefficients alone (Abdul *et al.*, 2020). In this work, we have adopted the use of 14 MFCCs and 14 GTCCs along the frames to be used to train the SVM classifier. Benkedjouh et al. extracted the MFCC feature and fed it to the SVM and claimed that the first three MFCC components contain the most defect information of gears

(Benkedjouh T., Chettibi T., Saadouni Y., 2018). However, based on the research of Abdul et al., 1-13 MFCCs are more effective to be taken to train LSTM (Abdul *et al.*, 2020) and Jin et al. evaluated some sets of MFCCs (16, 21, 26, 31, 36, 41 MFCCs) by feeding them into a CNN model individually and the result shows that 41 MFCCs outperformed the others for gear fault detection. In conclusion, the dimension of the MFCC features is important for increasing the classification rate (Jin *et al.*, 2021). There are some other researchers who used MFCC as feature for fault detection as shown in Table 2.1.

Table 2.1: Summarizing of those papers that use MFCC for fault detection

Classification	Feature	Advantage	Disadvantage
SVM	MFCC(Benkedjouh T., Chettibi T., Saadouni Y., 2018), Fusion MFCC, Wavelet Packet Decomposition Energy features and the Zero-crossing rate (ZCR) (Zhang *et al.*, 2018). MFCC and its delta (Akpudo and Hur, 2021)	Most influential coefficients are the first three MFCC coefficients(Benkedjouh T., Chettibi T., Saadouni Y., 2018). For Fusion feature, higher classification accuracy was achieved even under noisy environment(Zhang *et al.*, 2018). Rank-based recursive feature elimination was used to select proper feature among them and the selected feature are reliable for fault detection in electromagnetic pumps(Akpudo and Hur, 2021).	Other features like Wavelet Packet Decomposition and Zero-crossing rate were used to support the MFCC to gain better performance in the noisy background(Zhang *et al.*, 2018).
GMM	MFCC(Fang and Liu, 2020),MFCC+ Kurtosis(Nelwamondo and Marwala, 2006),	MFCC can hold 95% of defect information and adding Kurtosis leads to the improvement the accuracy by	The model was not checked in noisy environment

		4%(Nelwamondo and Marwala, 2006)	
XGboost algorithm	An improvement of MFCC was named adaptive frequency cepstrum coefficient (AFCC) (Qi *et al.*, 2021).	Adaptive MFCC improves the performance of the fault detection system	MFCC method is limited for detecting bearing fault as Mel scale designed based speech signal and is not suitable for the frequency distribution of rolling bearing vibration signal
Similarities analysis	MFCC (Kemalkar and Bairagi, 2017)	MFCC suitable to show some various faults detection	The proposed model was composed into two parts. Fault detection and fault classification. The MFCC was only used for fault classification

With the progress of machine learning, deep learning methods have been conducted in gear fault detection and received extensive attention. The deep learning models may be fed by a raw signal and/or hand-crafted features. Regarding the works that use the raw signal, researchers exploited the vibration raw signal directly and fed it to the CNN model, where there is no need to have any prior knowledge of the related features to the application (Yang *et al.*, 2019)-(Jing *et al.*, 2017). On the other hand, Shao et al. (Shao *et al.*, 2019) used images created by the time-frequency distributions by conducting wavelet transform to convert raw input to image data. Authors in (Wang *et al.*, 2017) proposed local feature-based gated recurrent unit networks that are trained by handcrafted features and automatically learned features. Abdul et al.(Abdul *et al.*, 2020) extracted a temporal feature, which is based on the MFCC and GTCC and fed the LSTM to classify gear faults for helical gears and parallel gearbox alone. Saufi el at. (Saufi *et al.*, 2020) proposed a deep learning model based on

a Stacked Sparse Auto Encoder (SSAE) and then combined it with t-stochastic neighbor embedding (t-SNE). The deep learning model has sole rewards in fault classification due to its potential automatic learning and extracting feature without human interaction (Tang *et al.*, 2020b). In (Hoang and Kang, 2019), the authors use it to isolate the bearing faults, where the raw data vibration signal is given to the CNN model, as CNN is the featureless model. The authors in (Ma *et al.*, 2019) developed a model where the pyramid wavelet packet decomposition of the signal became an input for the proposed CNN model. Shao et al. (Shao *et al.*, 2019) adopted a pre-trained CNN model specifically (VGG-16) for fault detection. The CNN results show a significant improvement in the classification rate. However, there are two main trades of issues in conducting deep learning models which are: 1) the complexity of the model and 2) setting the hyperparameters of the model. The complexity issue is very problematic, however, the hyperparameters of the deep learning model can be optimized. For example, Tang et al. optimized the hyperparameters of a CNN model by using Bayesian optimization (CNN-B) for detecting faults in a hydraulic pump. Continuous wavelet transform was taken from the vibration signal and fed to the CNN-B model and the result indicates that optimizing hyperparameters of the CNN can improve the accuracy of the fault detection system (Tang *et al.*, 2022).

2.6 Performance Measurement

Accuracy is a metric used to evaluate the performance of a machine-learning model. It is defined as the proportion of correct predictions made by the model out of all predictions made. In other words, accuracy measures how well the model is able to correctly classify the input data. The accuracy of a model is typically calculated by comparing its predicted output to the true

output for a set of test data. The accuracy is calculated as the number of correct predictions divided by the total number of predictions made (Poveda-Martínez and Ramis-Soriano, 2020). Mathematically, accuracy can be expressed as:

$$Accuracy = (Number\ of\ correct\ predictions) / (Total\ number\ of\ predictions) \quad (2.32)$$

Accuracy is an important metric for fault detection because it measures the proportion of correct predictions made by a model out of all predictions made. In fault detection, accurate detection of faults is critical for preventing equipment damage and ensuring safety. A high accuracy score indicates that the model is correctly identifying faults and anomalies and minimizing false positives and false negatives. False positives occur when the model predicts a fault when none exists, while false negatives occur when the model fails to detect a fault that actually exists. Both false positives and false negatives can have serious consequences in fault detection. False positives can result in unnecessary maintenance and downtime, while false negatives can lead to equipment failure or safety hazards (Decker and Lewicki, 2003).

The accuracy metric provides a clear and objective measure of how well a fault detection model is performing, which can be used to compare different models or fine-tune the parameters of a model to improve its performance. Other metrics, such as precision, recall, and F1-score, are also commonly used in fault detection, but accuracy is often the most straightforward and intuitive metric to interpret.

2.7 Summary of the Chapter

In this chapter, two main topics are explained. Firstly, a background of these tools that have been used in this work such as both features (MFCC and GTCC), with their implementation and some classifiers (SVM, LSTM, and ESN). Secondly, a literature review about two types of fault detection based on machine learning algorithms such as those that can deal with time series features and others that can identify patterns based on non-time series features.

CHAPTER THREE

3 GEAR FAULT DETECTION USING TIME SERIES MODEL

Fault detection based on time series models is a valuable approach used across industries to recognize abnormalities or deviations from anticipated patterns in systems or processes. These models effectively capture the temporal dependencies and patterns found in sequential data, making them ideal for detecting faults in systems that evolve over time. The fundamental concept behind fault detection using time series models revolves around capturing comprehensive dynamic information at regular intervals. Machine learning models that capture the temporal features of the vibration signal need to be capable of detecting the dependencies between time steps. Traditionally, Hidden Markov Model (HHM) and DTW are among the most used machine learning models for this purpose. Due to the advancement of deep learning methods RNN based models are frequently adopted by researchers. However, the problem of gradient vanishing that appears in RNN leads to the development of new modified versions, two of which (LSTM and ESN) are utilized in this chapter (see Figure 3.1). In the following sections, the adopted methodologies of both models are presented, in addition to the implementation and experimental setups.

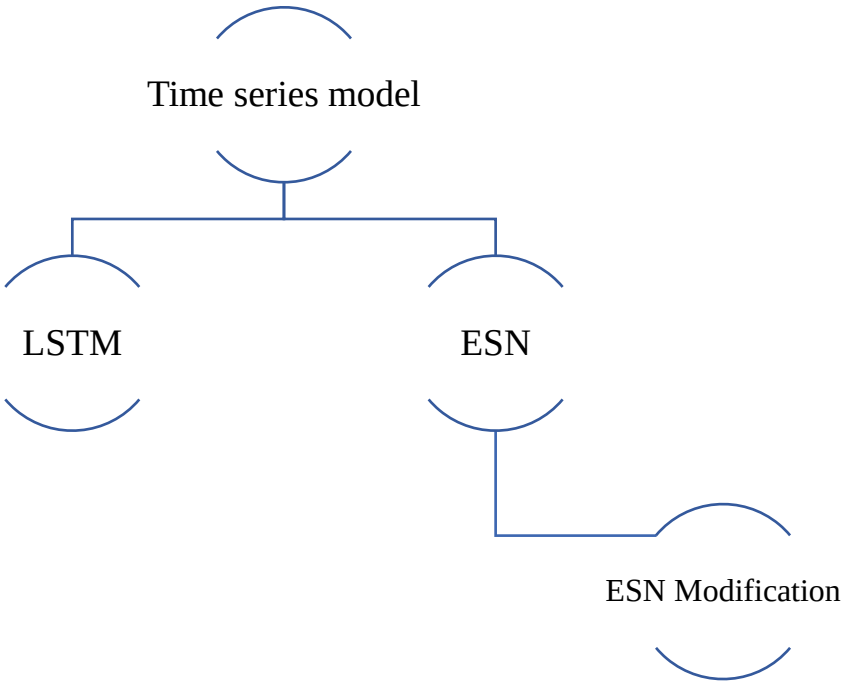

Figure 3.1: procedure of this chapter.

3.1 Gear Fault Detection Using LSTM.

3.1.1 Motivation of Using LSTM for Gear Fault Detection.

LSTM was introduced by Hochreiter and Schmidhuber in 1997, (Jaeger and Haas, 2004) as a modified architecture of RNN. It serves as an extension to standard RNNs and offers several advantages. LSTM is a well-known and useful model for time series-based pattern recognition problems due to its capability to handle long-term dependencies, retain relevant information in their memory cell, and accommodate variable-length sequences. Specifically, when training RNNs on extended sequences, LSTMs address the vanishing gradient problem, which can block learning. Their memory cell enables the retention of information over extended periods, which facilitates to capture and utilize long-term dependencies within the data samples (Bai, 2021).

Moreover, LSTMs possess the capacity to selectively remember or forget information from previous time steps using their memory cell. This capability ensures that significant information is retained while irrelevant or redundant data is discarded. Thus, it enhances the learning process and ultimately improves model performance. Additionally, the flexibility of LSTMs allows them to handle sequences of varying lengths effectively. This feature is particularly valuable when working with sequential data that lack fixed lengths, enabling the model to adapt and process such data dynamically. These characteristics of LSTM make a suitable and effective candidate model to be proposed for the vibration signal applications such as gear fault detection (Xiang *et al.*, 2020).

3.1.2 Dataset Description

I. Dataset-PHM09

One of the utilized datasets was collected by the Prognostics and Health Management Society, recognized as Prognostic Health Monitoring (PHM09). This database has been used widely by many researchers for detecting faults in gearbox equipment (Wei *et al.*, 2019)(Abdul *et al.*, 2016)(Shao *et al.*, 2019). Three sensors were tied in a gearbox with two acceleration sensors and a tachometer sensor for recording the vibration signals and speed shaft. Two typical gearboxes were used during the data acquisition process. Both gearboxes were different in terms of the utilized gear, helical or spur. Each gearbox ran under different shaft speeds; 30, 35, 40, 45, and 50 Hz, and different variance loads. The length of each record is four seconds with a 66667-sample rate. In this dissertation, four gear cases are considered including normal, chipped, broken, and mixed faults. Hence, the data comprises

120 samples with lengths of 266666-digit rate. Table 3.1 and 3.2 illustrate the four cases for helical and spur gear respectively and Figure 3.2 illustrates the structure of the gearbox.

Table 3.1: Status of the helical gears in PHM09.

Case	16Teeth	48Teeth	24 Teeth	40 Teeth
Helical 1	Good	Good	Good	Good
Helical 2	Good	Good	Chipped	Good
Helical 3	Good	Good	Broken	Good
Helical 4	Good	Good	Broken	Chipped

Table 3.2: Status of the Spur gears in PHM09

Case	32T	96T	48T	80T
Spur 1	Good	Good	Good	Good
Spur 2	Chipped	Good	Eccentric	Good
Spur 3	Good	Good	Eccentric	Good
Spur 4	Good	Good	Good	Broken

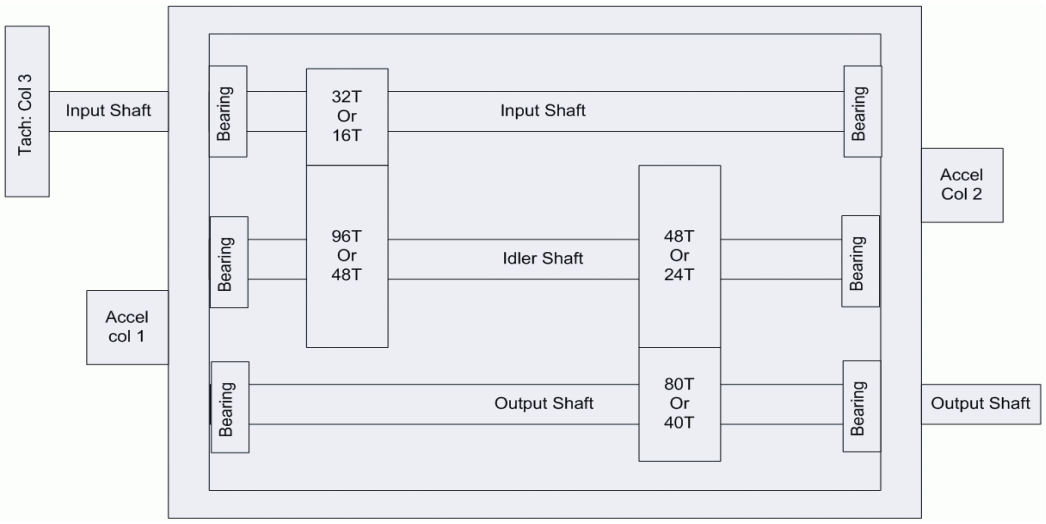

Figure 3.2: Structure of PHM09 gearbox where the T is tooth of the gear (PHM, 2009).

II. Dataset_ DDS.

Gearbox datasets were collected from a Drivetrain Dynamic Simulator (DDS) in Southeast University, China. A DDS was used to diagnose the faults in the gearbox under two different operation conditions such as load configuration and rotating speed (0V-20Hz and 2V-30Hz). The DDS consists of four main parts; parallel gearbox, planetary gearbox, brake, and motor (see Figure 3.3). The types of gear fault employed in the experiments were missing one of the gear feet (miss), crack in the root of feet (root), surface wear (surface) and feet crack (chipped). Therefore, to diagnose the gear fault four different data sets were created each with five classes (health and four fault conditions).

One of the accelerometer sensors was employed to measure the planetary gearbox vibration in the x, y, and z directions. Another one of the accelerometer sensors was used to measure the vibration of the gearbox in different directions. The last accelerometer sensor was used to measure the motor vibration. A

torque sensor was used in the experiments for measuring the load. The sampling window and sampling frequency of the data acquisition instrument are 512 s and 1024 Hz, respectively (Wang *et al.*, 2017)(Shao *et al.*, 2019).

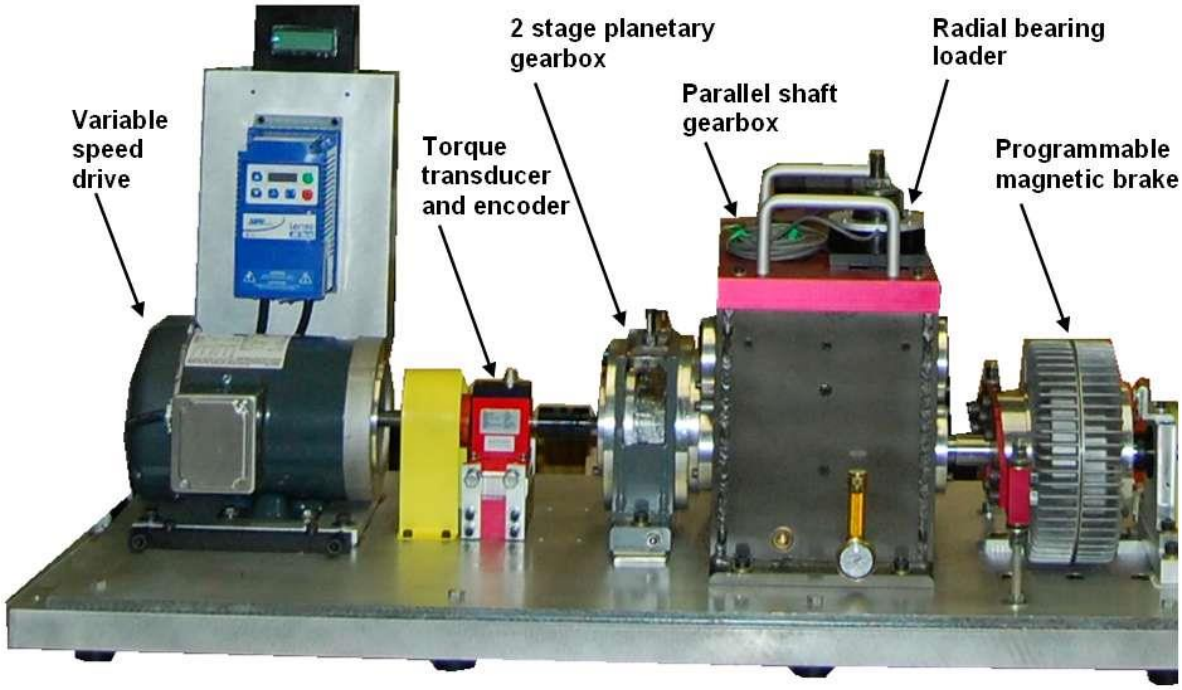

Figure 3.3: Drivetrain Dynamic Simulator (DDS) (Shao et al., 2018).

3.1.3 Feature Extraction and Representation

The adopted extracted features from the vibration data in this dissertation are the MFCCs and the GTCC. The process of feature extraction is started by windowing the raw signal into a number of frames, which depends on the frequency rate (fs) of the vibration signal. The length of the constructed vibration signal samples in both datasets is set to be ($2 * 10^4$) and the total

utilized number of samples are (416) and (750) for PHM09 and DDS datasets respectively. Therefore, the obtained frames for each sample are 17 in both datasets (dataset_PHM09 and dataset_DDS). For both the MFCC and GTCC features, 14 coefficients of each frame are computed where the default parameters for the MFCC and GTCC are fixed as shown in Table 3.3.

Table 3.3: The MFCC and GTCC Parameters

Parameters	MFCC	GTCC
Delta Window Length	2	2
Frequency Range	133 to 6864 Hz	50 Hz to fs /2
Number of Filters	40	32
Number of Coefficients	14	14
Log Energy	15	15
Window Length	0.03* freq. rate	0.03* freq. rate

Both features represent the input samples and are used for training the classifiers in a concatenated form (feature fusion). The adopted time series representation of features is producing a matrix of size (t×r) (denoted as matrix A) for each sample, where t is the number of windows (time step) and r is the number of features (dimensions). The matrix $A_{t\times r}$ for both MFCC and GTCC is combined to produce a matric of size $(t \times 2r)$ and then feed the LSTM.

3.1.4 LSTM Architecture

The adopted LSTM model in this chapter consists of two LSTM layers and one fully connected layer (see Figure 3.4). Due to the lack of resources, the hyperparameter of the adopted LSTM model was fixed at two LSTM layers each with 228 nodes and a ReLU function that feeds a fully connected layer with a number of neurons equal to the number of classes and followed by a SoftMax layer. In the training phase, the batch size is fixed to be 300, the number of epochs is chosen to be 30 and the adopted optimizer is selected to be Adam.

In order to reduce the complexity of the gear fault model based on the LSTM architecture, which was adopted by (Abdul *et al.*, 2020), the adopted size of the input sample in this model is reduced since the extracted feature for each sensor feeds the classification model in an independent way instead of taking the concatenated form of both features as an input to the classification model (see Figure 3.5). For instance, when the LSTM model in (Abdul *et al.*, 2020) was fed by PHM09, the features (MFCC and GTCC) were extracted from both input and output sensors and then concatenated to feed the classification model. The same idea is adopted for the DDS when the feature is extracted from each direction x, y, and z individually then feeds the LSTM model.

To evaluate the conducted experiments, a 10-fold cross-validation approach is applied to assess the performance of the adopted LSTM model and evaluate the different conditions related to the nature of the vibration data.

In the whole of this dissertation, chi-square, and T-tests have been adopted as a statistical test to know how significant the models are. The Chi-Square test is a statistical test used to determine whether there is a significant association between two categorical variables. Chi-Square is suitable to know the significant association between the output of two models for a specific case.

The T-test is used to show a significant association between the output of two models for several cases.

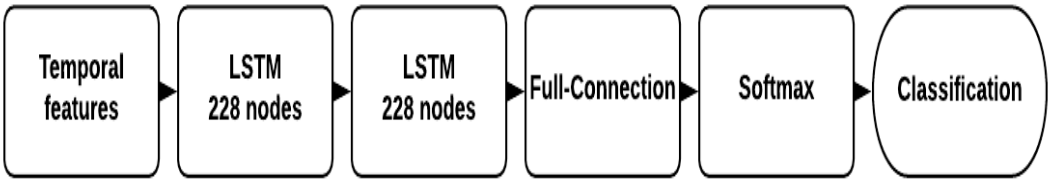

Figure 3.4: LSTM structure

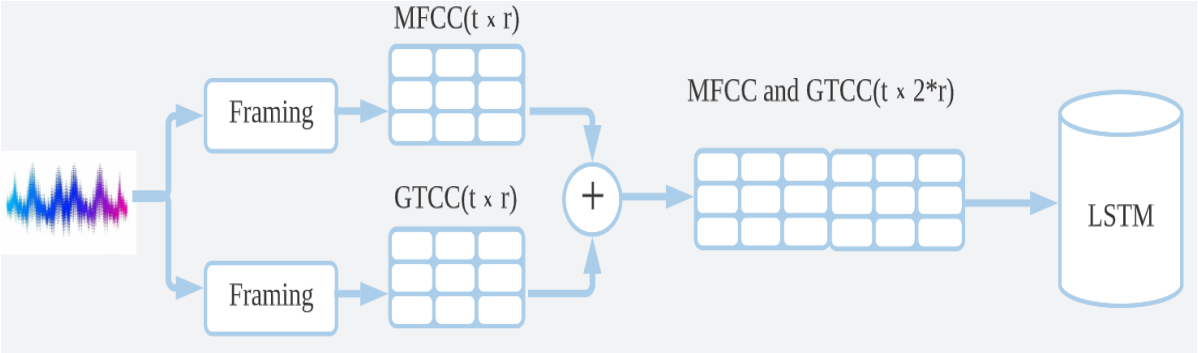

Figure 3.5: Gear fault detection based on LSTM model where we have three coefficients (r=3) and three-time steps (t=3).

3.2 Gear Fault Detection Using ESN

3.2.1 Motivation of ESN

The ESN is a form of RNN that was introduced by Jaeger in 2004 (Jaeger and Haas, 2004). Its purpose is to effectively process and analyze time-series-

based data, much like LSTM networks. Nevertheless, ESNs exhibit distinct characteristics and advantages in comparison to LSTMs. Within an ESN, the recurrent portion of the network called the reservoir, comprises randomly initialized recurrent nodes or units. These units are connected in a sparsely interconnected manner, establishing a multitude of recurrent connections. ESNs possess several key advantages, such as easy training and memory capability. Firstly, easy training implies that training an ESN involves a simple learning step solely for the readout layer. The recurrent connections within the reservoir are fixed and randomly initialized, which remains unaltered during training. Consequently, this approach renders the training process computationally efficient and less susceptible to overfitting. Secondly, ESNs possess memory capability, which enables them to capture and utilize information from past time steps. The recurrent connections within the reservoir retain the historical context of the input data, which facilitates the network's recognition of temporal patterns and dependencies (Dai *et al.*, 2009; Inc, 2023). Due to the challenge of the lack of computational resources that this work faced (as mentioned in (1.3)), ESN becomes a preferable candidate model for this work.

3.2.2 Time Complexity of ESN And LSTM

Comparing the time complexity of ESN and LSTM networks can provide insights into their computational efficiency for sequence modeling tasks. Big O notation of ESN can be computed based on several factors

- Training Time Complexity: O(N^2) for initialization, where N is the number of reservoir units. O(T) for training, where T is the length of the training data sequence. Overall, the training time complexity is dominated by the initialization step and is relatively efficient, especially for long sequences.

- Inference (Prediction) Time Complexity:O(T), where T is the length of the input sequence. ESNs are computationally efficient during inference because they involve fixed random connections and simple recurrent computations.

Regarding the time complexity of the LSTM can be computed as follows.

- Training Time Complexity: O(T * N^2) for the forward and backward passes during training, where T is the sequence length, and N is the number of LSTM units in a layer. Training LSTMs can be more computationally expensive, especially with large sequence lengths.

- Inference (Prediction) Time Complexity: O(T * N^2) for the forward pass during inference. LSTMs involve complex computations at each time step, making them potentially less efficient during inference compared to ESNs.

In summary, ESNs are often preferred for applications where computational efficiency is crucial, especially during training. LSTMs, on the other hand, offer more modeling flexibility and can capture complex temporal dependencies but come with higher computational costs. The choice between the two depends on the specific requirements of the task and the available computational resources.

3.2.3 The Adopted ESN Architecture

In recent years, several researchers have independently expanded upon the basic architecture of ESN by developing more advanced reservoirs, readouts, or input representations. To assess the combined effectiveness of these modifications in the context of multi-time series classification, the authors (Bianchi *et al.*, 2021), introduced a comprehensive framework for classification of the multivariate time series. This framework generalizes

various Reservoir Computing (RC) architectures by integrating four key modules, a reservoir module, a dimensionality reduction module, a representation module, and a readout module. Each module serves a specific purpose, and its descriptions are provided below. Figure 3.6 provides an overview of the models that can be implemented within this framework.

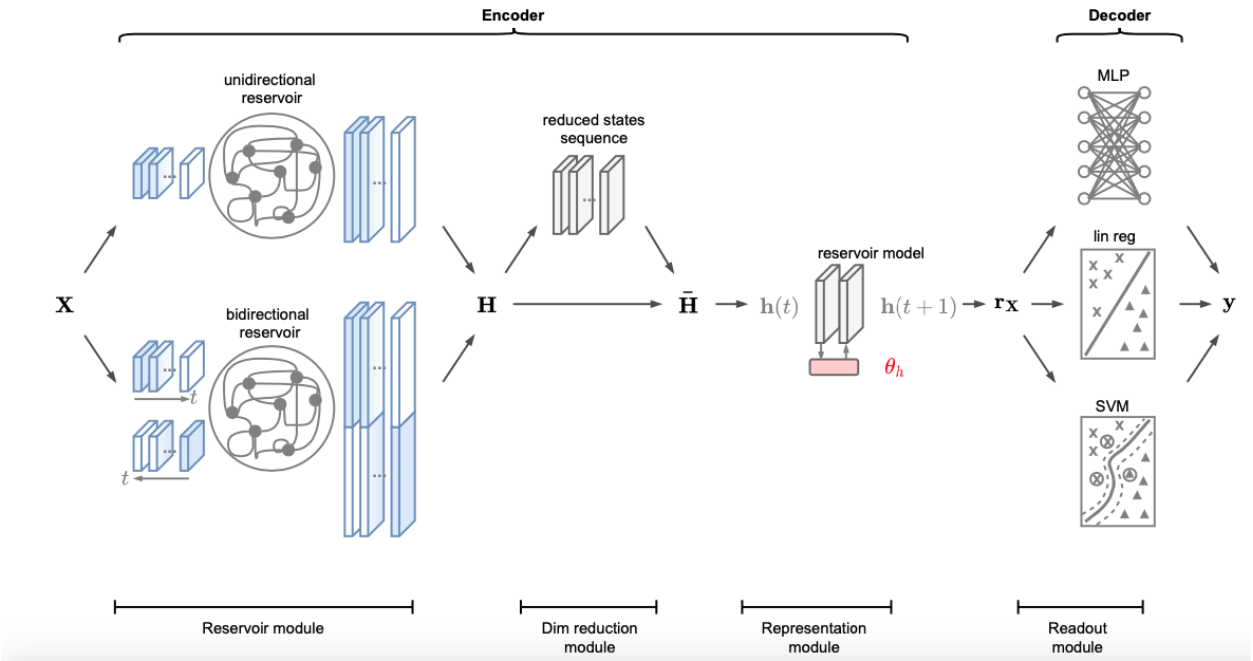

Figure 3.6: ESN structure for one reservoir (Bianchi et al., 2021).

3.2.3.1 Reservoir Module

Bidirectional reservoir was adopted by (Bianchi *et al.*, 2021) which refers to a reform of the traditional ESN architecture that incorporates bidirectional connections within the reservoir layer. In a standard ESN, the connections between neurons in the reservoir layer are unidirectional which

means the information flows only in one direction. Bidirectional reservoirs in ESNs differ from the standard architecture by including additional connections that facilitate the bidirectional flow of information within the reservoir layer. This enables reservoir neurons to capture both past and future inputs, resulting in a more comprehensive representation of the input sequence. It is important to note, though, that the incorporation of bidirectional reservoirs can introduce computational complexity due to the increased number of connections and the necessity to propagate information in both directions during training and inference. Bidirectionality is implemented, as shown in equation 3.1, by feeding an input sequence into the same reservoir both in straight and reverse orders.

$$\begin{aligned} \vec{h}(t) &= w_{in^*}x(t) + w_r * \vec{h}(t-1) \\ \bar{h}(t) &= w_{in^*}x(t) + w_r * \bar{h}(t-1) \end{aligned} \tag{3.1}$$

where h (t) is the RNN state at time t that depends on its previous value h (t − 1) and the current input x (t), f (·) is a non-linear activation function and the matrices W_{in} and W_r are the weights of the input and recurrent connections, respectively. The full state is obtained by concatenating the two-state vectors as illustrated by equation (2.14) and can capture longer time dependencies by summarizing at every step in both recent and past information.

3.2.3.2 Dimension Reduction Using Principal Component Analysis

The output of the reservoir layers is high dimensional and sparse, which can lead to overfitting and high computational costs. To address this issue, the well-known trainable dimension reduction technique Principal Component Analysis (PCA) was used to reduce dimension and convert the output into a

non-correlated representation. This can help improve the performance of machine learning models, reduce storage and computation requirements, and assist in visualizing high-dimensional data (Bianchi et al., 2021).

PCA works by selecting eigenvectors of the covariance matrix with the greatest eigenvalues, resulting in a data-dependent projection that de-correlates the data from the original input. The dimension of the output data is then reduced to a pre-fixed number, which can be either fixed or optimized through one of several approaches (Bianchi et al., 2021). This dimension reduction step helps to minimize the computation of the reservoir model space and produces a new sequence $h_1^{\sim}$, $h_2^{\sim}$, and $h_3^{\sim}$, which are used as input to the reservoir model.

3.2.3.3 Reservoir Module Space

The reservoir model space method, originally introduced in (Bianchi *et al.*, 2021), aims to capture the generative process of the reservoir sequence and establish a metric relationship among the samples. In the approach, the output of the reservoir module representation is not solely based on the final reservoir state, it rather depends on the entire sequence of states generated over time. As a result, the dataset is conveniently described as a three-mode tensor ($\mathcal{H} \in \mathcal{R}^{N \times T \times R}$), requiring a transformation that maps ($R \rightarrow D_{s.t}, D \ll R$) while keeping the other dimensions unchanged. To achieve dimensionality reduction on high-order tensors, the authors in (Bianchi *et al.*, 2021) suggest utilizing Tucker decomposition. This decomposition involves expressing the tensor as a core tensor multiplied by a matrix along each mode. When modifying only one dimension of $\mathcal{H}$, the Tucker decomposition is essentially equivalent to applying a 2D PCA on a specific matrix representation of $\mathcal{H}$. Particularly, to reduce the third dimension (R), one computes the mode-3 matricization of $\mathcal{H}$ by arranging the mode-3 fiber (high-order analog of matrix rows/columns) to be the rows of

a resulting matrix $H_{(3)} \in \mathcal{R}^{NT \times R}$. Then, the standard PCA projects the rows of $H_{(3)}$ on the eigenvectors associated with the largest D eigenvalues of the covariance matrix $C \in \mathcal{R}^{R \times R}$, which is defined a

$$C = \frac{1}{NT-1}\sum_{i=1}^{NT}(h_i - \overline{h})(h_i - \overline{h})^T \tag{3.2}$$

Where, h_i is the i-th row of $H_{(3)}$ and $\overline{h} = \frac{1}{N}\sum_{i=1}^{NT} h_i$. As a result of the concatenation of the first two dimensions in $\mathcal{H}$, the reservoir states across all samples and time steps are simultaneously captured in the tensor $\mathcal{H}$ and the matrix C evaluates the variation of components over time. However, this approach results in the loss of both the original data structure and the temporal orderings. The reservoir states from different samples and generated at different time steps get mixed together, potentially causing a loss in representation capability (Bianchi *et al.*, 2021). This disregards the presence of variations in time courses within individual samples. To overcome this limitation, they address each matrix $H_n \in \mathcal{R}^{NT \times R}$, obtained by slicing $\mathcal{H}$ along its first dimension (N), as an individual sample. With this modification, they introduced a sample covariance matrix that is defined as follows:

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(H_n - \overline{H})(H_n - \overline{H})^T \tag{3.3}$$

The first D leading eigenvectors of S are stacked in a matrix obtained as $E \in R^{R \times D}$ and the desired tensor of reduced dimensionality is obtained $\overline{\mathcal{H}} = \mathcal{H} \times_3 E$. Where, $\times_3$ is the three-mode product, C, S denotes the variations of the variables in the reservoir. However, since the whole sequence of reservoir states is treated as a single observation, the temporal ordering in different MTS is preserved. After dimensionality reduction, the model representation becomes

$$\tilde{h}(t+1) = U_h \tilde{h}(t) + u_h \quad (3.4)$$

where $\tilde{h}(.)$ are the columns of a frontal slice $\overline{H}$ of $\overline{\mathcal{H}}$, $U_h \in R^{D\times D}$, and $u_h \in R^D$. The representation will now coincide with the parameters vector $r_x = \theta_h = [vec(U_h);\ u_h] \in R^{D(D+1)}$, as shown in Figure 3.7.

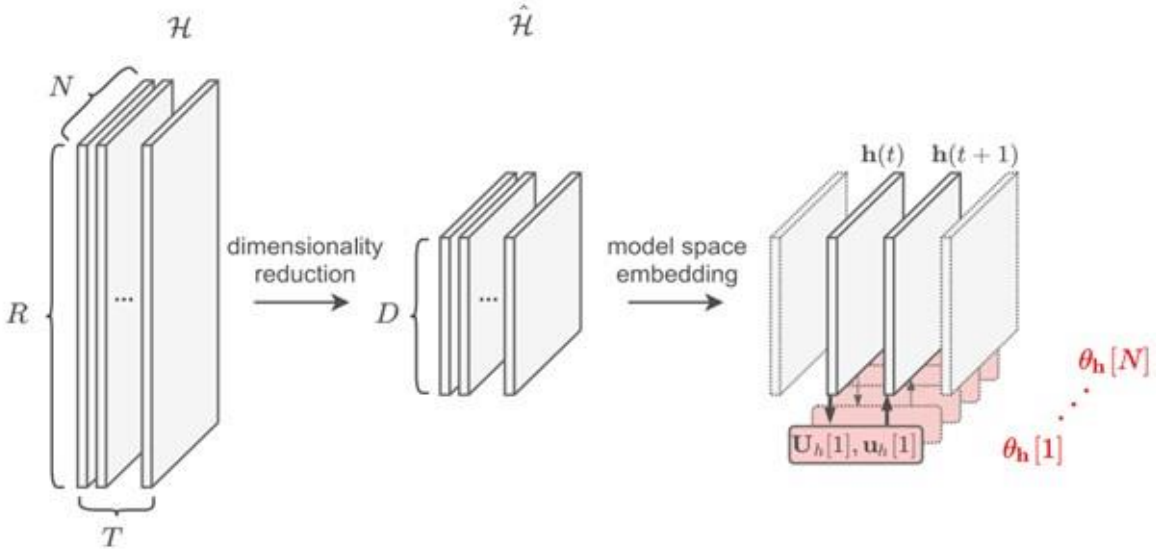

Figure 3.7:Reservoir model representation(Bianchi *et al.*, 2021)

3.2.3.4 Readout

The readout module classifies the representations and is either implemented as a linear readout or SVM classifier or a multilayer perceptron (MLP). In a standard ESN, the readout is linear and is quickly trained by solving a convex optimization problem. Based on their result, the linear readout outperformed the SVM and MLP.

3.2.4 Experimental Setup

The features are obtained from the vibration signal of each sensor. For example, in the PHM09 dataset, features are extracted from the input and the output sensor where each sensor is located on one of the sides of the shaft in the gearbox as shown in Figure 3.1. Regarding the DDS dataset, the features are extracted from each of the sensors (direction x, y, and z). Consequently, 14 coefficients of each frame for both features (MFCC and GTCC) are the output from this process such that the obtained features are formed by a $t \times r$ matrix (A) for each input sample, where r is the number of features (dimensions) and t is the number of time steps. Regarding time series ESN models with one reservoir which is illustrated in Figure 3.8, the matrix A ($t \times r$) for both MFCC and GTCC are combined to produce a matrix of size ($t \times 2r$), where r represents time step. The obtained feature (matrix A) is fed into the reservoir, and its dynamics transform the input into a higher-dimensional representation. The classifier is trained using the output of the reservoir after reducing high dimensional output using PCA. The 10-folds cross-validation techniques are conducted to assess the performance of all adopted models.

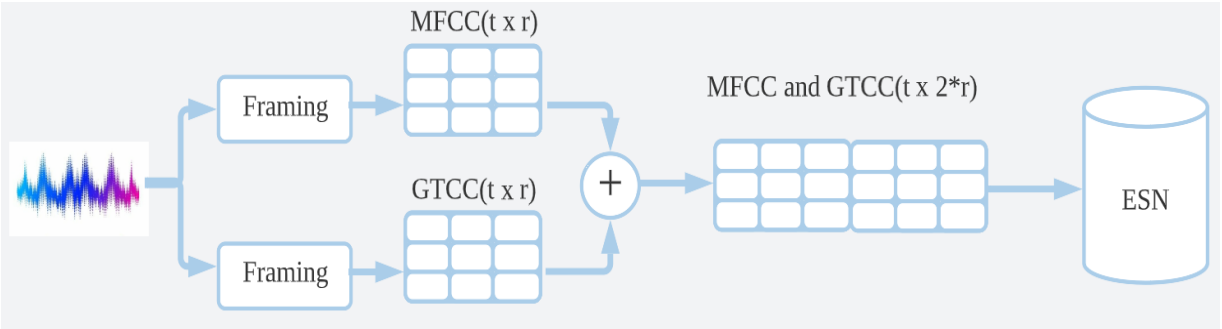

Figure 3.8: Gear fault detection model based on ESN where we have three coefficients (r=3) and three-time steps (t=3).

3.2.5 Modified ESN Model

According to the architecture of the ESN, model explained above, only one reservoir is adopted for each input direction. In this section, we are going to present a new modification that adopts three reservoirs' modules instead of one as shown in Figure 3.9. This can serve input data of multi-channel especially when the focus of the model is dealt with each channel in an independent way at the reservoir module level.

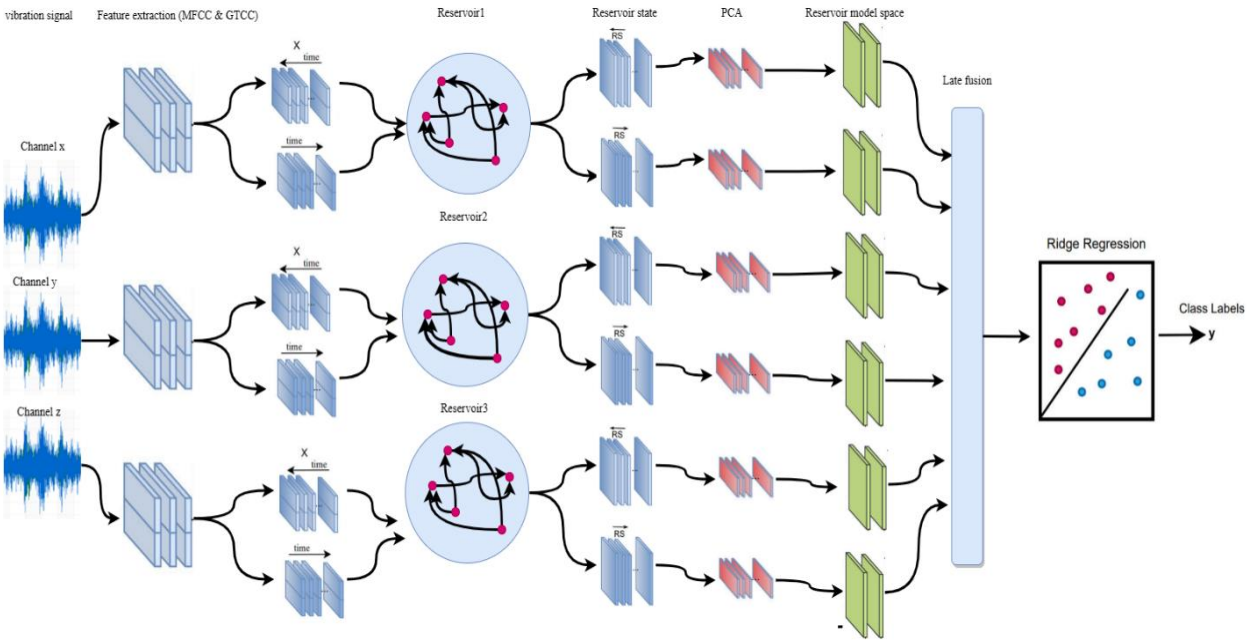

Figure 3.9: ESN with three reservoirs

In this proposed model, we utilized the three reservoirs-ESN approach, which employs three reservoirs with a late fusion. By processing data from different reservoirs independently, we can obtain diverse relational information about

the sequences. Combining the outputs of the three reservoirs using reservoir model space fusion yields more varied representations of the data and emphasizes the unique characteristics of each reservoir. Consequently, we modified the equation from the original to incorporate three separate outputs from PCA, which are presented in the following equations (3.5)

$$\begin{aligned} h_1^{\sim}(t+1) &= U_{h_1} h_1^{\sim}(t) + uh_1 \\ h_2^{\sim}(t+1) &= U_{h_2} h_2^{\sim}(t) + uh_2 \\ h_3^{\sim}(t+1) &= U_{h_3} h_3^{\sim}(t) + uh_3 \end{aligned} \tag{3.5}$$

where $h_1^{\sim}(t), h_2^{\sim}(t), and\ h_3^{\sim}(t)$ are the columns of a frontal slice H˜1 and H˜2 respectively, $U_{h_1} h_1^{\sim}, U_{h_2} h_2^{\sim}, and\ U_{h_3} h_3^{\sim} \in R^{\mathrm{D}\times\mathrm{D}}$ and $uh_1, uh_2, and\ uh_3 \in R^D$. The late fusion of triple reservoirs will be applied in this stage by combining all three outputs from $rx_1, rx_2, and\ rx_3$

$$\begin{aligned} rx_1 &= \theta_{h_1} = [vec(U_{h_1});\ uh_1] \\ rx_2 &= \theta_{h_2} = [vec(U_{h_2});\ uh_2] \\ rx_3 &= \theta_{h_3} = [vec(U_{h_3});\ uh_3] \end{aligned} \tag{3.6}$$

$$rx = [rx_1;\ rx_2;\ rx_3] \tag{3.7}$$

Regarding the readout layer, three classification models were adopted by (Bianchi *et al.*, 2021) including SVM, LR, and MLP. We employed all three methods, but LR is preferable for this work as it is mentioned in the experiment setup and evaluation.

3.2.6 Experimental Setup

The proposed ESN model with three reservoirs model for gear fault detection is evaluated based on two different gearboxes including planetary and

parallel gearboxes (see Figure 3.3). We extracted 14 MFCCs and 14 GTCCs as features from the vibration signal for each individual direction (x, y, and z) (see Figure 3.10) and fed them to different reservoir model. The features are extracted from the vibration signal for each window with a length of 30ms and an overlap of 20ms. These features were chosen to capture the essential characteristics of vibration signals. The MFCC and GTCC are fed to the ESN model individually and as a fusion. The fusion feature's size is 28 coefficients with 17-time steps, with each frame representing a time step while the individual size of each feature (MFCC and GTCC) is 14 coefficients with 17-time steps.

The proposed model was evaluated on two types of gearboxes - planetary and parallel gearboxes. As mentioned in section 3.1.2, data was collected under two different configurations, namely 20 Hz–0 V and 30 Hz–2 V. As a result, four configurations were examined for gear fault detection, comprising planetary gearboxes under 20 Hz–0 V and 30 Hz–2 V, and parallel gearboxes under 20 Hz–0 V and 30 Hz–2 V. In addition, two distinct scenarios were employed to assess the ESN model. The first scenario employed a single reservoir ESN, while the second scenario used our proposed model which featured three reservoirs. For the first scenario (one reservoir ESN), the extracted feature from all channels (x,y, and z) are combined and fed to the ESN model based on one reservoir. While, the proposed model consists of three reservoirs and each channel (x,y, and z) is given to the specific reservoir.

The output of reservoirs is very high dimension and therefore, PCA is used to reduce the output dimension. Flattening of the output reservoir is applied to the output of the late channel fusion and three classifiers are employed including SVM, Linear Regression (LR), and MLP. The performance of LR and MLP outperformed the performance of SVM with the P-value (5.90133E-13, 1.53519E-09) respectively. As shown in Figure 3.10, there is not a significant

difference between the performance of LR and MLP with (P-value 0.162). However, the MPL is very slow and time-consuming compared to LR. Thus, the LR is fixed in the proposed model for classifying classes.

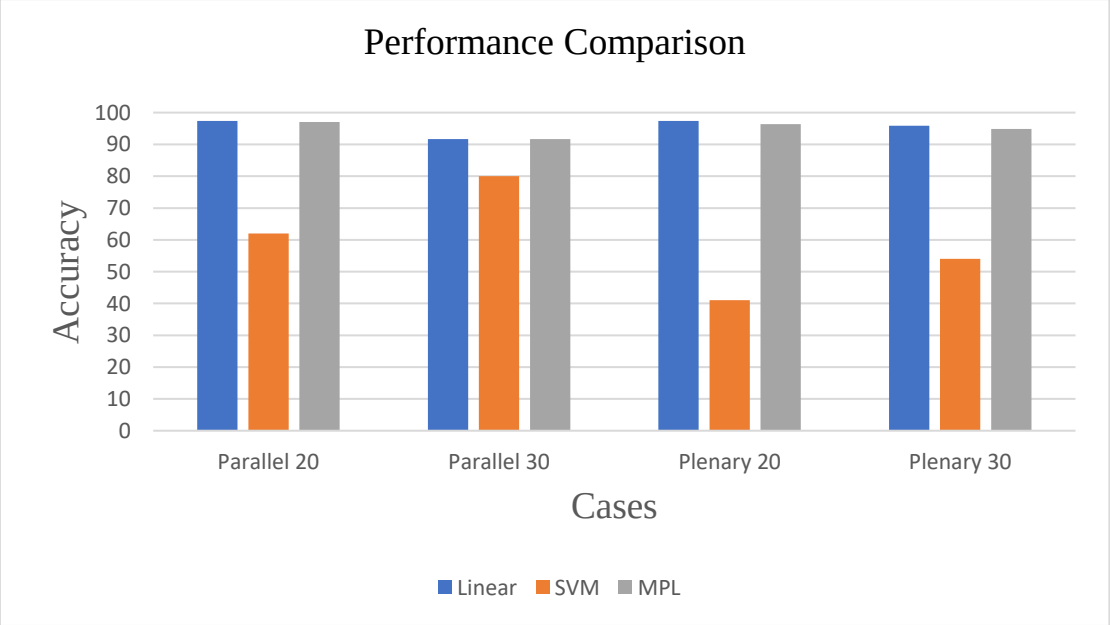

Figure 3.10: Accuracy of modified ESN for DDS dataset based on Linear, SVM, and MLP

3.2.7 ESN Hyperparameters

ESN hyperparameters play a critical role in determining the performance of the ESN model, and many studies have reported challenges in identifying the optimal values for these parameters. In most cases, researchers have relied on manual parameter assignment or previous experiences to determine the ESN hyperparameters. In this study, we aimed to optimize several critical ESN hyperparameters, including the size of the reservoir state, spectral radius,

connectivity size, input scaling, leakage in the reservoir state update, and number of dropouts. For this purpose, Bayesian optimization is applied to optimize these parameters as according to a comparison conducted in reference, Bayesian optimization was found to be more efficient than other gradient optimizations [48]. Table 3.4 shows all optimized parameters and Table 3.5 illustrates the performance of the model in each of the 10-fold cross-validation as the 10-fold cross-validating is used to optimize parameters based on overall vibration data.

Table 3.4: The parameters utilized in the proposed method have been optimized using the Bayesian optimizer.

Positions	Name of parameters
Reservoir	Internal units which is a number of units in the reservoir
PCA	Dimension which is number of the features
Model space	Regularization of the ridge regression is the value which lead to reduce penalty function
Readout	Regularization of the ridge regression in readout is also the value which lead to reduce penalty function

Table 3.5: Accuracy of each of the cases based on each fold during the validation.

Cases\No. fold	1	2	3	4	5	6	7	8	9	10
planetary 20-0	0.94	0.96	0.97	0.91	1.00	1.00	1.00	1.00	0.96	1.00
planetary 30-2	1.00	0.94	0.96	1.00	0.92	0.90	0.95	1.00	1.00	0.92

Parallel 20-0	0.95	1.00	1.00	1.00	0.92	0.96	0.96	1.00	0.94	1.00
Parallel 30-2	0.96	1.00	0.74	0.77	0.95	0.96	0.91	1.00	0.91	0.96

3.3 Result of Gear Fault Detection Based on Time Series Model

3.3.1 LSTM Result

Regarding the results of the PHM09 and based on Table 3.6, the average accuracy of the LSTM model, when the gear is helical and trained by individual sensors, is less compared to the model accuracy fed by the concatenation of the features of both sensors. However, because the fault detection in spur gear is more predictable since spur gear has a symmetrical tooth profile and single pair engagement, the performance of the model in both cases is equally performing when spur gear is utilized.

Table 3.6 Accuracy of the LSTM model for PHM09 dataset (helical and spur gear)

	LSTM (%)	LSTM (input & output %)
Input Helical	92.79	99.32
Output Helical	99.52	
Input spur	100	100
Output spur	100	

Regarding the DDS dataset, authors (Abdul *et al.*, 2020) extracted the feature from the signal channels (x, y, and z), then concatenated them. In this work, the features of individual sensors (x, y, and z) are extracted and fed to the LSTM model individually. As illustrated in Tables 3.7 and 3.8, significant degradation of the average accuracy of individual channels is observed in both gearboxes including planetary and parallel gearboxes. The reason for this degradation in both gearboxes might be due to the lack of input data to train the LSTM model. Moreover, the LSTM model requires a large amount of computation and memory space during the training phase, which can be a challenge on resource-constrained devices. To avoid this issue, we may need a low computation model, as we shall see in the next section that adopting the ESN model can be beneficial in that sense.

Table 3.7: Accuracy of the LSTM model for DDS dataset (planetary gearbox)

Case name	LSTM (%)	LSTM (x, y, and z)
planetary _x_20	85.6	97.87
planetary _y_20	96	
planetary _z_20	80.4	
planetary _x_30	85.6	94
planetary _y_30	75.2	
planetary _z_30	75.2	

Table 3.8: Performance of the LSTM model for DDS dataset (parallel gearbox)

Case name	LSTM (%)	LSTM (x, y, and z)
parallel_x_20	100	97.45
parallel_y_20	89.6	
parallel_z_20	98.4	
parallel_x_30	96.4	95.7
parallel_y_30	83.6	
parallel_z_30	92.4	

3.3.2 ESN Result

Two ESN structures have been implemented including ESN with one reservoir and ESN with three reservoirs. The ESN model with one reservoir was developed by (Bianchi *et al.*, 2021), while the ESN model with three parallel reservoirs is one of the contributions of this dissertation. In this chapter, the ESN with one reservoir is trained and tested either by obtained features (MFCC and GTCC) from the individual sensor (input and output sensor in PHM09 dataset and (x, y, and z) sensor in DDS dataset) or by the concatenated representation of the features in all sensors. From the obtained results, one can observe that the model with single sensors significantly outperforms the model that uses the concatenation of the channels. To exploit the independence of the provided information of each channel, we proposed to add parallelly two other reservoirs to the original ESN model. Each reservoir handles and manipulates one of the channels (x, y, and z). As shown in Table 3.9 and 3.10, a significant improvement has been observed with p-value= 0.02 compared to the ESN

model with one reservoir. The reason for this improvement is adding a reservoir which leads to generating complex and high-dimensional dynamics in response to input features. It also allows the network to efficiently process temporal data and capture long-term dependencies. Further information is presented in Figure 3.11 which is the confusion matrix of the modified ESN model for one of the folds when the gearbox is parallel with configuration 20-0 (DDS dataset).

Table 3.9: Accuracy of the ESN model for DDS dataset (planetary gearbox)

	ESN three reservoirs	ESN Combined signals	ESN channel x	ESN channel y	ESN channel z
Planetary 20-0	97.41	71.3	95.8	90.5	98.4
Planetary 30-2	95.90	64.4	95.8	84.1	92.1

Table 3.10: Accuracy of the ESN model for DDS dataset (parallel gearbox)

	ESN three reservoirs	ESN Combined signals	ESN channel x	ESN channel y	ESN channel z
Parallel 20-0	97.34	71.8	98.4	93.7	98.4
Parallel 30-2	91.65	78.7	98.4	96.8	88.9

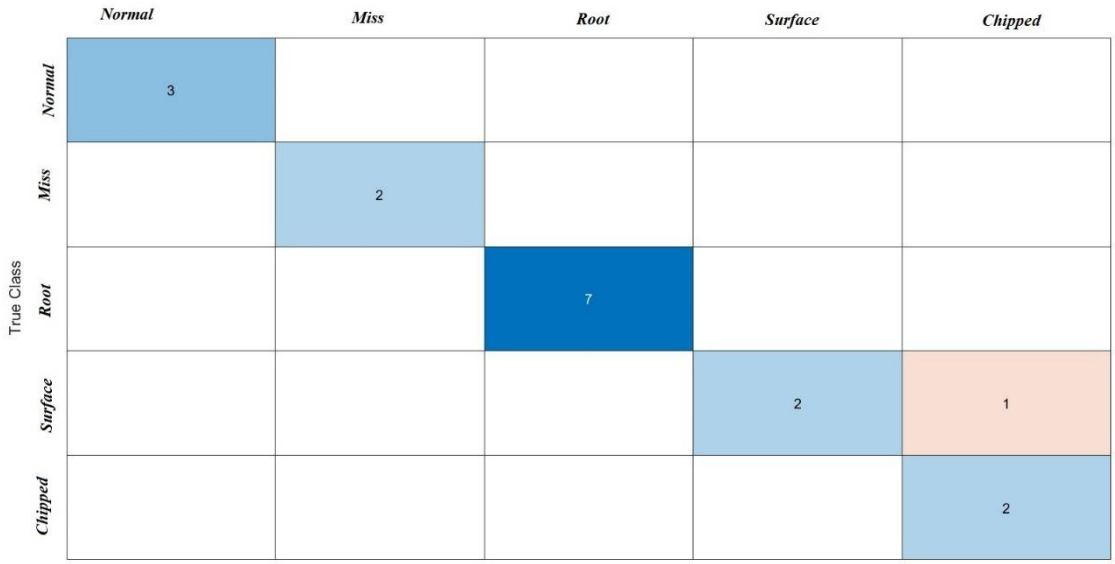

Figure 3.11: Confusion matrix of modified ESN model for parallel 20-0 (DDS dataset)

3.3.3 Comparison with State of Art Studies

In this work, to make a comparison between the models (LSTM and ESNs) with the best-achieved result and the state-of-the-art studies, we have presented a number of works that were applied to the same adopted datasets.

Regarding the DDS dataset where our proposed models have been applied to fault detection in the parallel gearbox, Table 3.11 shows five-time series models, which have been found in the literature, all of which used the same DDS dataset to feed models. For instance, Authors of (Wang *et al.*, 2017) used four-time series models including the Stacked Autoencoders-Deep Neural Network (SAE-DNN), GRU, Bidirectional Gated Recurrent Units (BiGRU), and Local Feature-Based Gated Recurrent Unit Networks (LFGRU) which enhanced gated recurrent unit network (GRU) that was based on a hybrid handcraft feature and learned feature to detect fault. Abdul et al.(Abdul *et al.*, 2020) extracted a temporal feature from all sensors, which is based on the MFCC and GTCC and fed the LSTM to classify gear faults for helical gears

and parallel gearbox alone. The result of our proposed model (ESN with one reservoir and LSTM when they feed by the individual sensor) is able to outperform all the presented state-of-the-art works. Moreover, ESN with three reservoirs outperforms SAE-DNN, Gated Recurrent Units (GRU), BiGRU, and LFGRU. However, when the configuration of the parallel gearbox is 30-2, the ESN with three reservoirs cannot achieve a better accuracy (in some sensors) compared to LSTM, which was adopted by Abdul et al.(Abdul *et al.*, 2020). Moreover, there is no guarantee to know the location of a gear fault during the operation time. Consequently, the importance of the vibration direction remains unknown. Therefore, the ESN with three reservoirs is still reasonable to be used as it is stronger and less time-consuming compared to the LSTM model.

Table 3.11: The accuracies of adopted time series models and the state-of-arts models for parallel gearbox.

Sources	Fault diagnosis method		20-0	30-2
Wang et al.(Wang *et al.*, 2017)	SAE-DNN		92.70	91.90
	GRU		93.80	90.50
	BiGRU		93.80	90.70
	LFGRU		94.80	95.80
Abdul et al.	MFCC-GTCC-LSTM		97.45	95.70
LSTM	MFCC-GTCC-LSTM	x	**100**	**98.4**
		y	89.60	96.8
		z	98.40	88.9
ESN	ESN with one reservoir	x	98.4	96.40
		y	93.7	83.60
		z	98.4	92.40

	ESN with one reservoir (Combined signals)	71.8	78.7
	ESN with three reservoirs	97.34	91.65

Regarding the case of the helical gear data (in the PHM09 dataset), authors in (Abdul *et al.*, 2016) extracted a handcraft feature and fed it to the LSTM classifiers. The performance of our work has not outperformed their achieved result as shown in Table 3.12. However, the difference is not statically significant (p-value=0.08)

Table 3.12: The accuracies of the ESN and the state-of-arts models for helical gear.

	Features	Accuracy%
Abdul et al (Abdul *et al.*, 2020)	MFCC-GTCC-LSTM	99.30
ESN with one reservoir	MFCC-GTCC from input sensor	**98.80**
	MFCC-GTCC from output sensor	**98.80**

Concerning spur gear (in PHM09 dataset), C. Wu et al. (Wu *et al.*, 2019) used the same dataset adopted in this work and proposed 1-D CNN model which was

fed by the raw data directly and their result shows that the performance of the ESN is better than the 1-D CNN (see Table 3.13).

Table 3.13: The accuracies of the ESN and the state-of-arts models for spur gear.

	Features	Accuracy%
C. Wu et al. (Wu *et al.*, 2019)	1-DCNN	99.33
ESN with one reservoir	MFCC-GTCC from input sensor	**100**
	MFCC-GTCC from output sensor	**100**

Regarding the ESN model with three reservoirs, it is not applicable to be evaluated by the PHM09 dataset as the data was collected by an accelerometer sensor which was a 1-axis accelerometer. This sensor can record only one axis (one channel). Therefore, the PHM09 dataset does not have three different vibration channels as the DDS dataset has, as it mentioned in section 3.1.2.

3.4 Summary of Chapter Three

In this chapter, two models were employed: LSTM and ESN. Initially, the LSTM model was trained by concatenating both features and utilizing all channels of the vibration signal for gear fault detection. Although the model's performance was reasonable, it is a time-consuming model during the training phase and required significant memory space for implementation. To address this issue, a single channel of vibration signal was used for training the LSTM model, but this led to a slight degradation in its average performance, which may be due to the underfitting.

To overcome the time-consuming issue of the LSTM model, attention was shifted toward the ESN model, an enhanced version of the RNN. The ESN model offered a notable advantage as certain parameters were selected randomly, resulting in significantly reduced time requirements. The ESN model was trained using concatenated features (MFCC and GTCC) and the complete set of vibration signal channels. However, the accuracy of the ESN model for gear fault detection was not satisfactory enough for real-life applications.

Subsequently, the ESN model was trained individually for each channel of the vibration signal, resulting in a significant improvement in performance. Certain channels demonstrated superior performance, potentially due to a higher occurrence of vibrations in those specific channels. However, it was challenging to determine the exact location of the fault solely based on these results. To ensure reliable results irrespective of the fault location, the ESN model was enhanced by adding two additional parallel reservoirs to the standard ESN model, which originally had only one reservoir. Each reservoir was responsible for increasing the variance of the data.

The enhanced ESN model utilized the three reservoirs, with each reservoir being fed data from one of the channels. This approach resulted in achieving more reliable results for gear fault detection, considering that the location of the fault was not known in advance.

CHAPTER FOUR

4 GEAR FAULT DETECTION USING NON-TIME SERIES MODEL.

In this chapter, fault detection based on non-time series feature representation has been investigated. For this purpose, the implementation of three different approaches using non-time series features have been presented namely, an approach that adopt optimizing features, the use of statistical representation of features hyperparameters, and brute force usage of all values by concatenating time steps feature values where the utilized features are MFCC and GTCC.

Time series-based classifiers pay attention to the dependencies of the time step values, however, there are still two main reasons that motivate the use of features firstly non-time series representation considers the overall behavior and characteristics of the vibration signal, rather than focusing on individual data points or local patterns. Second, non-time series features may also be useful to enhance the robustness of fault detection systems by reducing the impact of noise or fluctuations in individual data points. By aggregating information across the entire system, non-time series features can provide a more stable and reliable indication of potential faults.

4.1 The Motivation for Using the SVM

Support Vector Machines (SVM) is a supervised machine learning algorithm used for classification and regression tasks. They operate by creating hyperplanes in a high-dimensional feature space. The objective of SVMs is to identify an optimal hyperplane that not only separates the classes but also maximizes the margin, indicating a higher level of confidence in the classification process. The motivation behind the use of SVM in this

dissertation can be summarized in two main points. First, SVM performs well in high-dimensional spaces, even when the number of features (dimensions) is much greater than the number of samples. This makes SVM more suitable for handling complex, high-dimensional classification problems compared to alternative algorithms such as k-Nearest Neighbors, Naïve Bayes and DT. Second, SVM is inherently less affected by outliers since the margin-based approach focuses on the critical support vectors closest to the decision boundary.

4.2 Optimized Features

As introduced in Chapter 3, the same two feature techniques have been adopted in this chapter, which are MFCC and GTCC. Both features were originally developed for extracting features from speech signals. Consequently, the default hyperparameter values of the MFCC and GTCC are based on speech features, and they may not be optimal for gear fault detection. To overcome this limitation and enhance the effectiveness of MFCC and GTCC as a feature for gear fault detection, in this section, we employ the GWO optimization algorithm. But the performance of optimized MFCC as we will see in 4.5.1 does not change compared to default MFCC. So, we focus on optimizing the hyperparameters of GTCC. The GWO and FDO algorithms are utilized to fine-tune three important GTCC hyperparameters: the number of coefficients, the minimum frequency, and the maximum frequency. By adjusting these parameters, the gear fault detection system accuracy has been improved. In the following section, we will explain the methodology of optimizing the GTCC hyperparameter step by step.

4.2.1 Initializing GWO and FDO Parameters

Based on the literature, GWO and FDO optimization algorithm have the potential to get a solution in a short time as it is called speed convergency and it has the ability to achieve accurate solutions compared to other metaheuristic algorithms (Salih *et al.*, 2022)(Panda and Das, 2019). For any metaheuristic algorithm, there are three fundamental parameters that need to be identified before conducting the algorithm including an objective function which is also called the fitness function, the number of iterations, and the number of agents to find out the solution. In this work, optimizing the GTCC hyperparameter using GWO and FDO minimizes the classification error rate which is based on the SVM classification method as illustrated in Figure 4.1. One of the initial parameters in both GWO and FDO, which is the number of iterations, is set to 100. The minimum and maximum of all optimized parameters of CTCC including number of coefficients, range of minimum and maximum frequency, are set to (10-20), (15-100) and (100-frequency rate/2) respectively.

In the proposed model, the number of iterations of GWO and FDO is set to 100. Moreover, 500 iterations are also adopted but there is not a significant change in tuning the parameters as shown in Figure 4.2. Regarding GWO, this confirms that GWO is able to converge in a fast way. The number of agents must not be less than three as the GWO works are based on at least three agents namely alpha (α), beta (β), and delta (δ) (Faris *et al.*, 2018). So, we set the number of agents to 10 agents for both algorithms (GWO and FDO). The main reason behind choosing GWO for this work is that GWO is simpler in terms of formulation compared to FDO. The number of parameters in GOW is also minimal to avoid any complexities. Consequently, the GWO can be implemented without consuming more iterations throughout the training of classifiers (Mansor *et al.*, 2021).

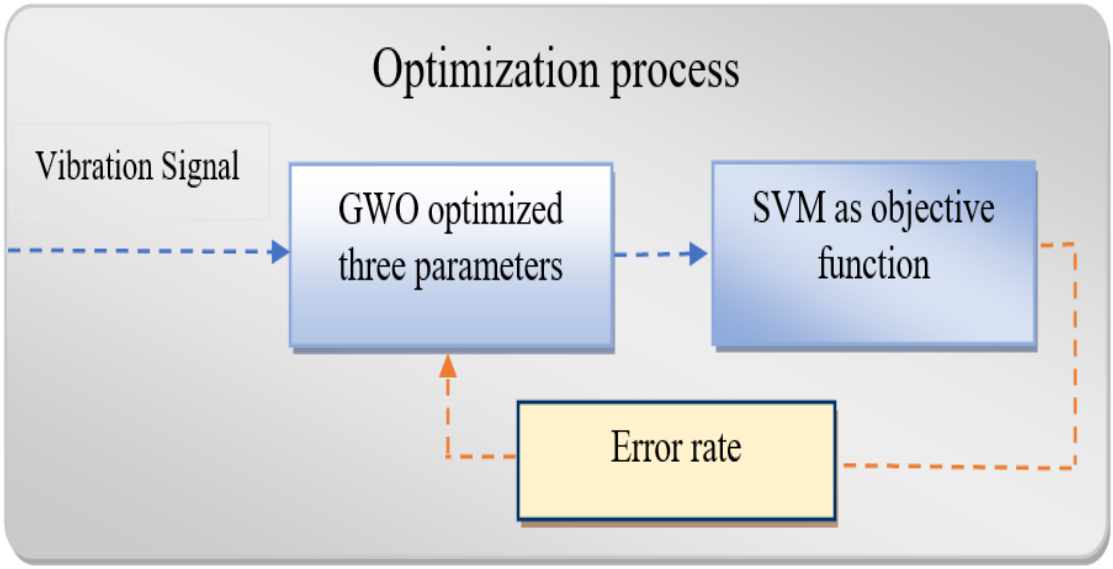

Figure 4.1: Optimization process based on GWO and SVM

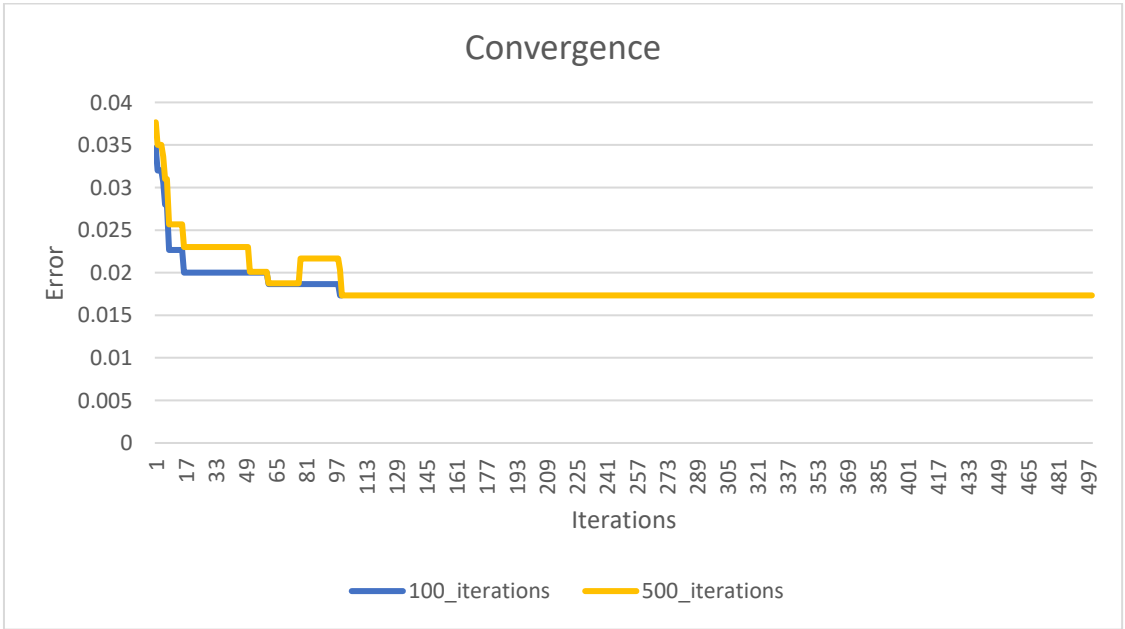

Figure 4.2: convergence line of GWO

4.2.2 Model Design

In this section, the processes of the proposed method are described as also shown in Figure 4.3. These processes are decomposed into three main processes namely feature extraction, applying GWO, and classification method. The vibration signal length in both datasets is set to be (2×10^4). The total sample is (416) and (750) for PHM09 and DDS datasets respectively. Various speeds in both datasets are utilized to evaluate the proposed model and all three channels (x, y, z) in the DDS dataset are concatenated.

In this section, GTCC is adopted to be extracted from the samples and the optimization process is applied to tune its main parameter since it is initially derived for speech signals. In other words, optimizing these parameters is motivated by the fact that the default GTCC parameters' values have been set based on the speech feature. Thus, the parameters need to be tuned to be a more effective feature for gear fault detection. To address this problem, the GWO optimization algorithm is used in this section to adjust three GTCC hyperparameters including the number of coefficients, minimum, and maximum frequency which can be effective in gear fault detection system development. GWO has optimized these three parameters based on the accuracy of SVM performance as shown in Figure 4.3. Later, the GTCC frames are concatenated to accommodate harmonic characteristics on vibration signal such that the size of each feature vector is equal to $(1 \times (frame \times number\ of\ coefficients))$, which feeds the SVM to classify various faults. As a consequence of optimizing these parameters, the size of the feature will be changed because of the change in the number of coefficients and frequency range during the optimization process. The number of coefficients will affect the dimensional of each feature and the frequency range will impact the number of frames. For example, after optimizing these parameters, the optimum

number of coefficients becomes 17, and the number of frames is optimized to be 6. While in the default case, the number of coefficients and frames are equal to 14 and 17 respectively. Consequently, the number of dimensions of the feature degrades from (1×238) to (1×102). The accuracy obtained by the SVM classifier is utilized as a fitness function and fault classifier during the optimization process and classification process respectively. The process of parameter optimization validation is done by 10- fold cross-validation.

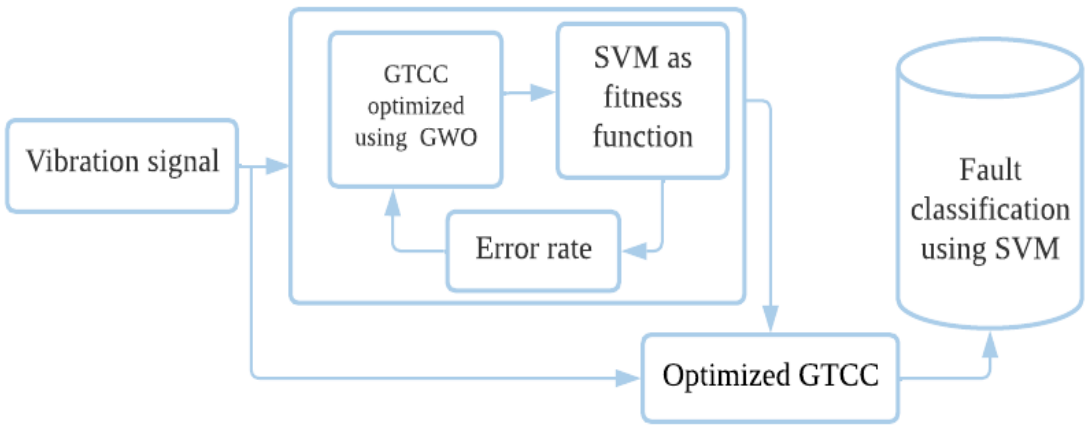

Figure 4.3: The proposed model with optimized feature

4.3 SVM with Statistic Feature (Stat-SVM)

Windowing time-series signals lead to capturing the short time information, however, various ways are adopted to globalize the time-series feature representation and consequently lead to unifying the length of the signal representation regardless of its original length. Statistical features are one of the ways that offer numerous advantages including dimension reduction of the data, unifying the length, and easing the interpretation of the feature

representation. Statistical features allow us to summarize large amounts of data into a few key measures, providing a concise representation of the dataset including information about the distribution of data and patterns within the dataset. Statistical features provide quantitative metrics that summarize or describe various aspects of the data. These metrics are making them interpretable by individuals with a basic understanding of statistics.

The feature extraction process involves two main steps: extracting MFCC and GTCC. Each process starts with breaking down the raw signal into frames using a technique called framing or windowing. The feature extraction step generates a (t × r) matrix (A) for each input sample, where t is the number of frames and r is the number of coefficients. From this matrix, nine statistical values are extracted including mean, variance, standard deviation, root mean square, min, max, skewness, and kurtosis, resulting in 126 features (9 statistics × 14 coefficients (r)). The statistical features for both MFCC and GTCC are then concatenated, resulting in 252 statistic features, which are fed to the SVM as shown in Figure 4.4.

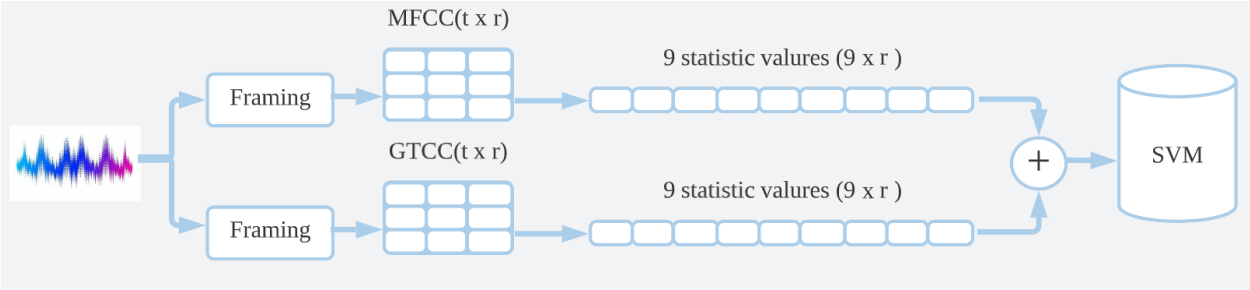

Figure 4.4: The design of the Stat-SVM model where we have three coefficients (r=3) and three-time steps (t=3).

However, the statistical features reduce complex data, potentially leading to an oversimplification of the underlying information. By reducing the data to a few key statistics, certain nuances and detailed patterns may be lost, limiting the richness of the analysis. That leads to a situation where the statistical features alone may struggle to detect faults that manifest as subtle or non-linear changes across different dimensions. To address this problem, we shall see in the next section that all the frames of the features (MFCC and GTCC) have been concatenated in order to keep all the information that is obtained from the features.

4.4 SVM with Concatenated Feature (Concat-SVM)

Windowing the vibration signal is a crucial step in both adopted feature extraction techniques, namely MFCC and GTCC. This step holds significant importance as it effectively reduces spectral leakage, enhances frequency resolution, and minimizes side lobes and border effects. These advantages contribute to more precise and reliable feature extraction for various signal-processing applications, including fault detection. In order to avoid losing any information inside each window as we have faced in the Stat-SVM, we propose a model based on the concatenation of the whole feature values for all of the windows, then it is fed to the SVM (called Concat-SVM). Figure 4.5 shows the design of the Concat-SVM model.

The features are extracted from the vibration signal, and this produces a t×r matrix (A) for each input sample, where t is the number of time steps (frames) and r is the number of coefficients. The matrix A is flattened into a one-dimensional vector of 1×P, where P is equal to t×r. The flattened representation of both features of the size of 2×P is fused and used to train the SVM for classifying the faults of the gears.

The hyperparameters of SVM have been selected based on Bayes optimization algorithms including Box-Constraint (the c parameter) and kernel function. Based on the result of the bayes optimizer, Box-Constraint is optimized to be 1 while the kernel function is chosen to be Linear. Moreover, the 10-fold cross-validation technique is used for evaluating the performance of the model.

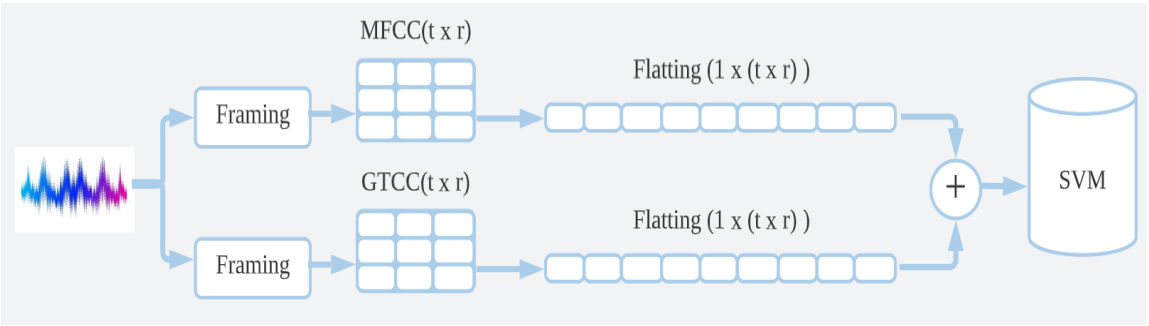

Figure 4.5: The design of the Concat-SVM where we have three coefficients (r=3) and three-time steps (t=3).

4.5 Result of Non-time series Model

4.5.1 Result of Optimized Feature

In this section, the proposed model is evaluated based on two different datasets (PHM09 and DDS). Three GTCC parameters (minimum frequency, maximum frequency, and the number of coefficients) are optimized using the GWO and FDO optimizers. Regarding the convergence of the GWO, according to the obtained result, it is changed based on the load of the data samples. For instance, the convergence in the case of 20Hz-0V-load in the DDS dataset is significantly faster, and the error rate is lower than what is in the case of the 30Hz-2V-load as shown in Figure 4.6. Achieving a lower error rate for 20Hz-

0 load is also confirmed in (Wang *et al.*, 2017) (Shao *et al.*, 2019) (Saufi *et al.*, 2020) and (Abdul *et al.*, 2020). This might be due to the increasing overall acceleration levels of the vibration signal when the load is high. Additionally, the amplitude of each component of the vibration signal increases proportionally to the load variation (Sharma and Parey, 2017).

The minimum frequency is one of the important parameters because of the characteristics of the severe gear fault location in low-frequency regions (Watson *et al.*, 2007). According to the optimized parameters using GWO, the minimum frequency in five out of six cases for both datasets are in the range of 31 to 43 Hz, while the default minimum frequency is set to 50 Hz.

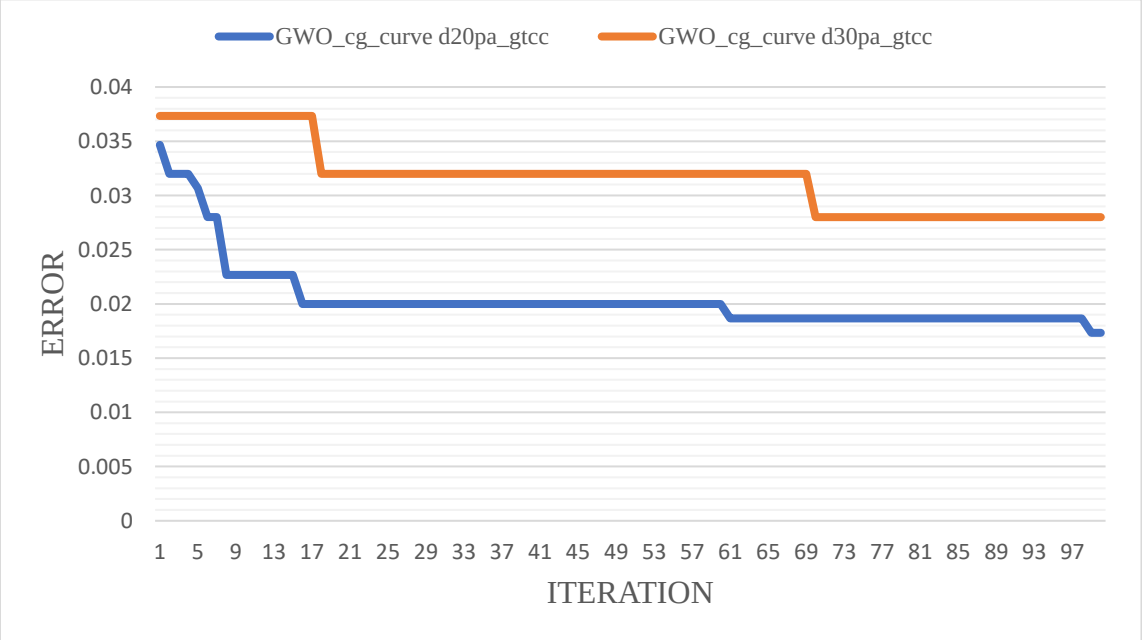

Figure 4.6: Convergence curve of GWO for both configurations (20Hz-0V-load and 30Hz-2V load).

Regarding the maximum frequency, the default maximum frequency range is equal to ($\frac{Frequency\ rate}{2}$) which is about 33333 Hz for the adopted cases in this chapter. In contrast, the optimized maximum frequency is less than half of the default range value as it is shown in Table 4.1 and 4.2 for both DDS and PHM09 datasets respectively. Moreover, the optimized maximum frequency has increased according to the speed of the input shaft as GMF is proportional to the speed. The change in the GMF and its sidebands is one of the well-known indicators for fault appearance and the GMF depends on the speed of the input shaft and the number of the teeth gear. Equation 4.1 illustrates the relationship between speed and gear mesh frequency where,

$$\mathrm{GMF} = \mathrm{T} \times \mathrm{RPM} \qquad (4.1)$$

T is the number of gear teeth and RPM is the speed of the input shaft.

Table 4.1: Optimized values for three parameters in DDS.

Speed-load	No. Coefficient	Min Frequency (Hz)	Max Frequency (Hz)
20Hz-0V	17	34.07	8059.95
30Hz-2V	17	43.66	9517.48

Table 4.2: Optimized values for three parameters in PHM09.

Speed	No. Coefficient	Min Frequency (Hz)	Max Frequency (Hz)
30 Hz	16	31.00	7991.06
35 Hz	17	32.19	9489.32
40 Hz	17	31.27	10061.14

45 Hz	17	55.07	12037.44
50 Hz	17	31.19	12309.94

The three optimized parameter values are used to extract GTCC from the vibration signal. The obtained handcraft feature (GTCC) is then fed to the SVM model to discriminate gear faults. Once again, the SVM is validated using 10-fold cross-validation. Based on the result for the PHM09 dataset, the use of the optimized parameters increases the accuracy of SVM as shown in Table 4.3 and Figure 4.7 shows the confusion matrix of the model when the gear speed is 50Hz. This improvement indicates that the default parameters are not the perfect values for the gear fault detection application, and this supports the claim that choosing speech-based parameters may not be suitable for other applications. Although the FDO optimizer was able to improve the result by optimizing the parameters in four cases out of five cases, GWO is still outperforming the default in all the cases and the FDO in most of the cases.

Table 4.3: Accuracy result for PHM09

Speed	GTCC		
	Default (100%)	GWO (100%)	FDO (100%)
30 Hz	99.04	100	100
35 Hz	98.08	100	99.52
40 Hz	97.60	99.76	100
45 Hz	100	100	99.28
50 Hz	98.32	100	99.52

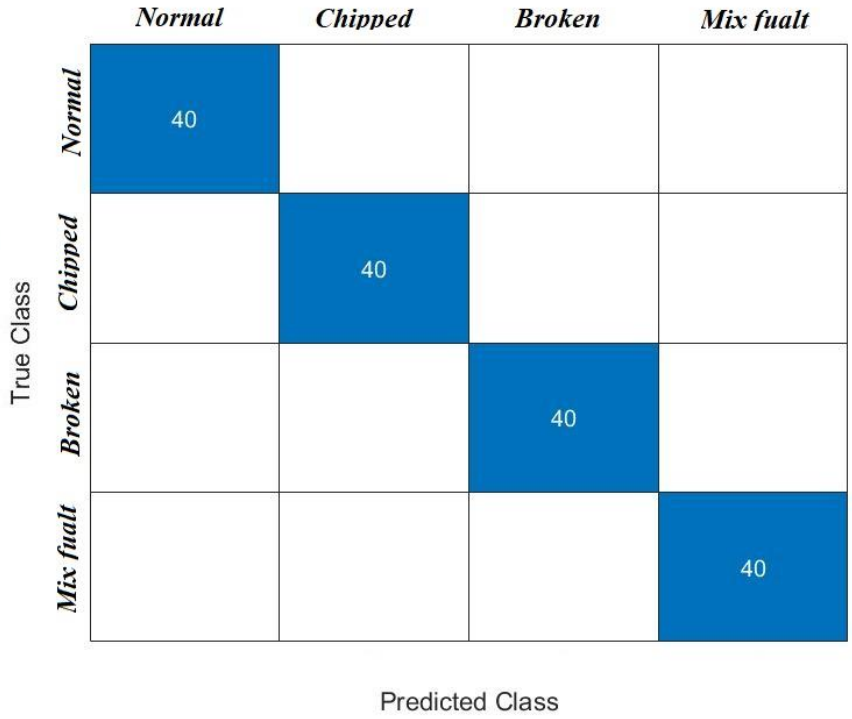

Figure 4.7: Confusion matrix of optimized GTCC model for PHM09 (speed 50Hz)

Regarding the DDS dataset, a significant improvement is observed using GWO as displayed in Table 4.4. In addition to the improvement of the performance of the SVM to classify gear faults, the optimization of these parameters leads to a reduction in the dimension of the GTCC feature as all frames will be flattened eventually. Consequently, the number of frames decreased from 17 frames to 6 frames. However, the number of coefficients increased from 14 to 17. As an output of the optimization process, the feature vector is reduced from $(1 \times (14 \times 17) = 238)$ to $(1 \times (17 \times 6) = 102)$.

The use of FDO in this dataset has not shown improvement compared to GWO. The reason might be that the FDO suffers from two main issues including low exploitation and slow convergence (Salih *et al.*, 2022).

Table 4.4: Accuracy result for the DDS dataset

Various speed	GTCC		
	default	GWO	FDO
Planetary 20 Hz-0V	93.33	97.90	93.73
Planetary 30 Hz-2V	94.70	96.30	93.20
Parallel 20 Hz-0V	94.20	98.3	93.30
Parallel 30 Hz-2V	95.10	97.1	94.70

Regarding the PHM09 dataset, the proposed model achieves a higher accuracy rate compared to the state of arts as shown in Table 4.13. This confirms the usefulness of the proposed optimized GTCC and SVM and is an indication of the potential of the optimized GTCC to extract information on the gear faults from the vibration signal.

Regarding the MFCC feature, the mentioned optimizers have been used for optimizing its hyperparameter. However, the performance of the adopted model for gear fault detection does not change and no significant improvement is achieved during GWO iteration as shown in below Figure 4.8, which is a pilot test for planetary gear box under 20 Hz-0v configurations. This might be due to the fact that tuning these parameters does not affect the bandwidth of the Mel filter bank because the bandwidth in MFCC is only determined by the frequency spacing, which is uniform and overlaps between adjacent filters.

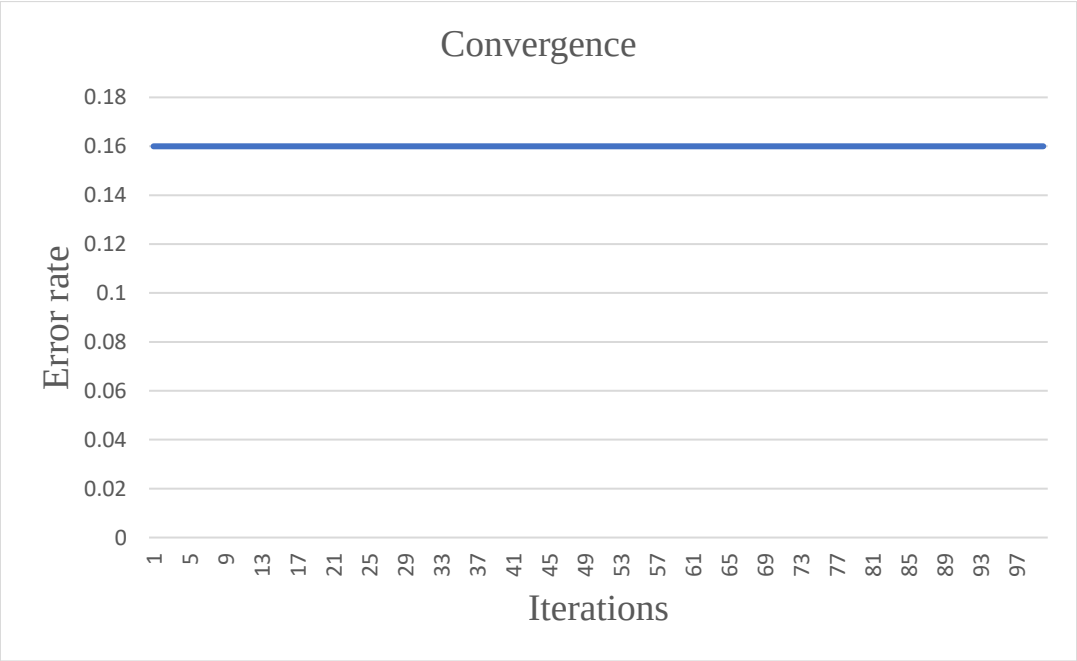

Figure 4.8:Convergence line of GWO during MFCC optimization

4.5.2 Result of Stat-SVM and Concat-SVM Models

The result of the PHM09 dataset is presented in Table 4.5 for each of the gears (Spur and Helical). The vibration signal for the spur gear gets reflected easily for any fault due to the low contact ratio. However, because of the large contact ratio in helical gears, the faults do weakly affect the vibration signal (Amarnath and Praveen Krishna, 2014). This has been confirmed by our results as well. The Concat-SVM and Stat-SVM models achieved 100% of accuracy for both input and output spur cases, while the input and output helical cases record slightly lower accuracy. However, the performance of the concat-SVM method for helical gear achieves better accuracy than the Stat-SVM method.

Table 4.5: Accuracy of Concat-SVM and Stat-SVM for PHM09 dataset

Cases	Concat-SVM (%)	Stat-SVM (%)
Input Helical	99.62	99.62
Output Helical	100	99.81
Input spur	100	100
Output spur	100	100

The result of the DDS dataset is also presented for each of the gearboxes (planetary and parallel gearbox). In the planetary gearbox, there are several gears that are attached to each other, such as sun, ring, and multi-planet gears. Multiple planet gears produce similar vibrations with differently phased gear meshes coupled with each other, consequently, some of the excitations of multiple gear meshes, which are very helpful for detecting faults, may be neutralized or cancelled (Lei *et al.*, 2012). Thus, fault detection and diagnosis in the planetary gearbox is more complex compared to fault detection in parallel gearboxes. This is also observed from the performance of the stat-SVM model, which is degraded significantly. However, the concat-SVM model shows the capability of effectively handling gear fault detection and the performance outperformed the other adopted models (see Table 4.8). The reason for the good performance of the concat-SVM may be the capability of having a comprehensive representation of the vibration signal features. In that sense, the concat-SVM outperforms the statistical representation of the time step feature values and the sequence-related information that fed the LSTM and ESN model in the previous chapter. Figure 4.9 shows the confusion matrix of Stat-SVM for the best-case achievement (channel z and confirmation 20-0).

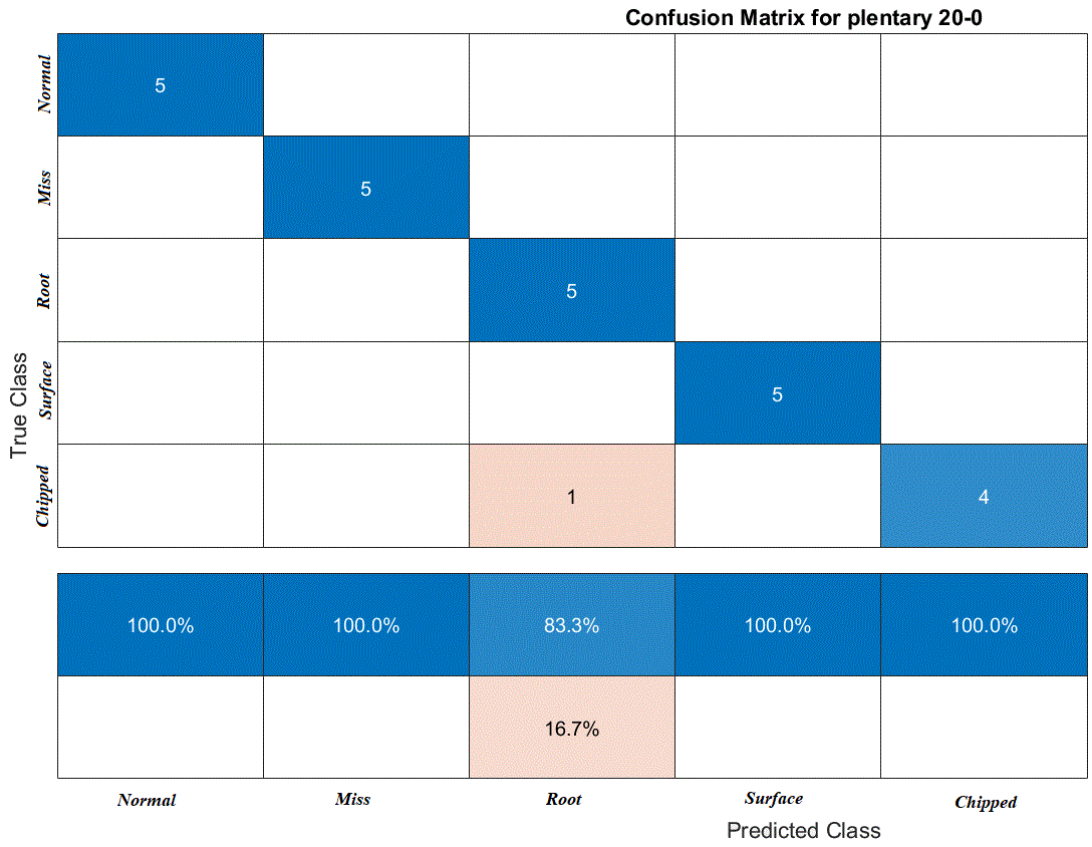

Figure 4.9: Confusion matrix of Stat-SVM model for DDS (planetary 20-0)

In parallel gearbox, as shown in Table 4.9, it is obvious that the concat-SVM has outperformed other models with a p-value less than 0.05 (p-value of 0.01 against the stat-SVM). The concatenation of the feature values along all the frames has performed better than the adopted statistics representation which might be an indication of losing information by the statistics. The computed statistics (mean, variance, standard deviation, root mean square, min, max, skewness, and kurtosis) are a sort of dimension reduction and globalizing the data representation. However, they are able to capture global information based on the characteristics of the adopted statistics and the distribution of the values along the time steps. Some of the adopted statistics are more useful to represent data with Gaussian distribution such as mean, variance, and standard deviation.

Some others are detecting more of the non-Gaussian properties of the data such as skewness and kurtosis. Consequently, the distribution of the data may not be a serious issue since various statistics have been adopted. However, it leads to discarding some data values that may carry related information to fault gear application. Moreover, based on the feature distribution (as shown in Table 4.6 and 4.7), some statistical values like (mean, variance, standard division, and root to mean square) are less effective if the distribution of the feature is not normally distributed.

Table 4.6: Normality test for MFCC and GTCC coefficients (PHM09 dataset)

Kolmogorov-Smirnova					
Coefficients	Speed-30	Speed-35	Speed-40	Speed-45	Speed-50
	P-value	P-value	P-value	P-value	P-value
GTCC1	0	0	0	0	0
GTCC2	4.93E-92	1.30E-150	1.30E-150	1.20E-163	3.47E-89
GTCC3	3.70E-260	0	0	0	0
GTCC4	9.20E-47	5.40E-44	5.40E-44	1.96E-36	1.26E-40
GTCC5	3.40E-67	4.94E-29	4.94E-29	3.27E-34	4.79E-16
GTCC6	8.30E-167	1.40E-297	1.40E-297	9.80E-133	4.30E-186
GTCC7	7.04E-05	1.67E-33	1.67E-33	2.12E-19	2.78E-22
GTCC8	3.62E-16	8.30E-143	8.30E-143	4.99E-44	4.21E-53
GTCC9	3.94E-25	2.20E-105	2.20E-105	4.40E-130	1.11E-96
GTCC10	5.99E-46	3.85E-16	3.85E-16	1.53E-06	5.21E-26
GTCC11	1.12E-15	3.90E-102	3.90E-102	4.05E-36	2.60E-28
GTCC12	2.59E-70	0.137426	0.137426	4.03E-11	4.59E-19
GTCC13	0.2	1.58E-07	1.58E-07	6.62E-09	0.008212
GTCC14	7.34E-65	1.07E-14	1.07E-14	1.50E-63	1.03E-52
MFCC1	0	0	0	0	0

MFCC2	6.93E-30	1.10E-17	1.10E-17	1.88E-12	1.12E-18
MFCC3	1.30E-27	1.69E-13	1.69E-13	0.015297	3.03E-07
MFCC4	4.14E-12	1.55E-10	1.55E-10	5.89E-13	1.22E-25
MFCC5	3.15E-11	3.92E-29	3.92E-29	5.44E-12	0.095177
MFCC6	1.55E-37	3.82E-09	3.82E-09	2.89E-13	0.000103
MFCC7	3.48E-05	1.47E-05	1.47E-05	1.19E-06	2.16E-16
MFCC8	4.29E-18	2.85E-10	2.85E-10	1.26E-12	5.47E-12
MFCC9	1.35E-28	0.030251	0.030251	0.2	1.16E-14
MFCC10	4.44E-06	0.000473	0.000473	0.045631	1.61E-09
MFCC11	1.84E-07	5.26E-22	5.26E-22	4.77E-68	1.70E-35
MFCC12	3.92E-09	2.13E-09	2.13E-09	3.20E-48	7.18E-16
MFCC13	0.001824	8.50E-05	8.50E-05	0.114221	0.048262
MFCC14	3.43E-06	3.67E-10	3.67E-10	0.032757	2.10E-10

Table 4.7: Normality test for MFCC and GTCC coefficients (DDS dataset)

Kolmogorov-Smirnova				
Coefficients	Parallel 20-0	Parallel 30-2	Planetary 20-0	Planetary 30-2
	P-value	P-value	P-value	P-value
GTCC1	0	1.90E-206	0	0
GTCC2	1.50E-168	9.70E-106	2.83E-58	7.90E-294
GTCC3	0	1.63E-05	0.003	3.24E-05
GTCC4	5.19E-78	0.2	0.011	1.65E-07
GTCC5	1.21E-74	0.2	0.001	3.61E-05
GTCC6	4.60E-249	4.13E-05	0.141	0.2
GTCC7	3.40E-121	0.0188	0.09	0.2
GTCC8	1.17E-69	0.08	0.0241	0.2
GTCC9	5.56E-24	0.2	0.048	0.2
GTCC10	3.03E-10	0.003	0.2	0.2
GTCC11	1.61E-39	0.2	0.2	0.2
GTCC12	1.48E-72	0.0005	0.147	0.2

GTCC13	0.005118	0.2	0.018	0.129
GTCC14	9.05E-41	0.2	0.2	0.081
MFCC1	0	1.90E-206	0	0
MFCC2	6.54E-10	5.70E-61	5.39E-20	5.30E-294
MFCC3	3.76E-72	0.007	0.009	0.002
MFCC4	6.04E-14	0.098	0.2	0.2
MFCC5	3.05E-29	0.2	0.2	0.2
MFCC6	3.46E-12	0.2	0.2	0.2
MFCC7	4.10E-11	0.03	0.2	0.2
MFCC8	0.06446	0.08	0.11	0.1
MFCC9	1.31E-05	0.13	0.08	0.2
MFCC10	0.004575	0.15	0.2	0.09
MFCC11	6.63E-10	0.2	0.2	0.2
MFCC12	2.09E-12	0.2	0.2	0.2
MFCC13	1.79E-26	0.2	0.2	0.2
MFCC14	6.41E-06	0.2	0.2	0.2

In spite of the fact that the concatenation increases the dimension of the data but fortunately, the SVM classifier has been reported to be useful with high dimensional data(Al-Talabani *et al.*, 2015). The performance of the concat-SVM model tells us that the sequence base information of the vibration signal is not mandatory for detecting gear faults.

Table 4.8:Accuracy of Concat-SVM and Stat-SVM for the planetary gearbox.

case name	concat -SVM (%)	Stat-SVM (%)
planetary _x_20	**99.60**	95.20

planetary _x_30	**100**	93.80
planetary _y_20	**100**	94.00
planetary _y_30	**99.60**	86.00
planetary _z_20	**100**	98.00
planetary _z_30	**99.60**	88.00
mean	99.80	92.50

Table 4.9: Accuracy of Concat-SVM and Stat-SVM for parallel gearbox.

case name	concat-SVM(%)	Stat-SVM(%)
Parallel_x_20	**100**	99.40
Parallel_x_30	**100**	96.60
Parallel_y_20	**100**	97.00
Parallel_y_30	**100**	92.40
Parallel_z_20	**100**	97.80
Parallel_z_30	**100**	98.00
mean	**100**	96.87

4.5.3 Comparison With State of Art Studies

In order to make a comparison between the model with the best-achieved result in this work and the state-of-the-art studies, we have presented a number of previous works that were applied to the same adopted dataset.

Regarding GTCC optimization, as shown in Table 4.10, the state-of-the-art results for the DDS dataset include various approaches and achieve higher results, however, the proposed model records more accuracy at 98% and 96.30% for the 20Hz-0V and 30Hz-2V loads respectively except Pre-trained model which was developed by Shao et al.(Shao *et al.*, 2019). Nevertheless, there is a significant difference between our proposed model and the pre-trained model in terms of the complexity of the model. The pre-trained model, nonetheless, has two blocks of convolution layers, and three fully connected layers, which are required to be trained to consume a lot of memory space and time. For both blocks of convolution layers, the learning parameters are equal to 12979200 and the number of weights for three fully connected layers is 119576336 weights. The proposed architecture in this dissertation is much simpler; however, it achieves a comparable result to the work in (Shao et al., 2019). Moreover, the proposed work demonstrates notable improvement over the approach presented in (Abdul et al., 2020), which utilized the same GTCC feature, however, in a time series-based representation without optimizing its hyperparameters.

Table 4.10: Comparison between the Achieved Results of the DDS Dataset and the State-of-the-Art Results

Sources	Fault diagnosis method	20Hz-0V	30Hz-2V
Wang et al.(Wang *et al.*, 2017)	SAE-DNN	92.7	91.9
	GRU	93.8	90.5
	BiGRU	93.8	90.7
	LFGRU	94.8	95.8

Shao et al.(Shao *et al.*, 2019)	Pre-trained model	99.64	99.02
Saufi et al.(Saufi *et al.*, 2020)	SSAE+PSO+fast kurtogram+tSNE	97.53	96.16
Abdul et al.(Abdul *et*	GTCC-LSTM	96.55	93.80
Proposed method	Optimized GTCC-SVM	98	96.30

Regarding the Concat-SVM, the best achievement recorded by our proposed models has been applied to fault detection in parallel gearbox, Table 4.11 shows four state of art studies where all used three channels (x, y, and z) together to feed their models. The only two works that use the channel separately are our proposed models and the work by Saufi et al.(Saufi *et al.*, 2020), where the mean accuracy of the three channels is presented in Table 4.12. Authors of (Wang *et al.*, 2017) proposed an enhanced gated recurrent unit network (GRU) which was based on a hybrid handcraft feature and learned feature to detect fault whereas, Shao et al. (Shao *et al.*, 2019) used time-frequency distributions as an image input by conducting wavelet transform to convert vibration signal to image and then fed to a pre-trained network. Additionally, Saufi et al. (Saufi *et al.*, 2020) proposed a deep learning model based on a stacked sparse autoencoder and then combined it with t-SNE. And finally, Abdul et al.(Abdul *et al.*, 2020) extracted a temporal feature from all sensors, which is based on the MFCC and GTCC, and fed the LSTM to classify gear faults for helical gears and parallel gearbox alone. The result of our proposed model (concat-SVM) is able to outperform all the presented state-of-the-art works.

Table 4.11: The accuracies of the cancat-SVM model and the state-of-art models for parallel gearbox

Sources	Fault diagnosis method	20-0	30-2
Wang et al.(Wang *et al.*, 2017)	SAE-DNN	92.70	91.90
	GRU	93.80	90.50
	BiGRU	93.80	90.70
	LFGRU	94.80	95.80
Shao et al.(Shao *et al.*, 2019)	CNN	98.70	94.14
	Pre-trained model	99.64	99.02
Saufi et al.(Saufi *et al.*, 2020)	SSAE+PSO+fast kurtogram+tSNE	97.53	96.16
Abdul et al.(Abdul *et al.*, 2020)	GTCC-LSTM	96.55	93.80
	MFCC-GTCC-LSTM	99.73	99.20
Proposed	Optimized GTCC-SVM	98	96.30
Proposed method	**Concat-SVM**	**100**	**100**

Regarding dealing with the channels (x,y, and z) separately, tables (4.12 and 4.13) show the accuracies of each channel for both of our proposed model contact-SVM and the adopted model by Saufi et al.(Saufi *et al.*, 2020) for both configurations 30hz-2 -20hz-0 respectively.

Table 4.12: The accuracies of the models using individual channels for 30hz-2 configuration.

		planetary gearbox			parallel gearbox		
		x	y	z	x	y	z
Saufi et al.(Saufi *et al.*, 2020)	DNN	77.6	67.2	80.8	76	63.6	65.2
	CNN	80.4	75.6	81.6	82	79.6	83.2
	SSAE+PSO+fast kurtogram+tSNE	97.5	95.5	**100**	95	97.1	96.4
proposed method (concat-SVM)		**100**	**99.6**	99.6	**100**	**100**	**100**

Table 4.13: The accuracies of the models using individual channels for 20hz-0 configuration.

		planetary gearbox			parallel gearbox		
		x	y	z	x	y	z
Saufi et al.(Saufi *et al.*, 2020)	DNN	80.4	86.8	84.8	80	75.6	79.8
	CNN	87.6	81.2	83.6	76.8	81.6	84.4
	SSAE+PSO+fast kurtogram+tSNE	97.5	97.5	98.8	96.8	97	98.8
proposed method (concat-SVM)		**99.6**	**100**	**100**	**100**	**100**	**100**

Regarding the case of the helical gear data (in the PHM09 dataset), authors in (Abdul *et al.*, 2016) extracted a handcraft feature and fed it to the SVM and k-

NN classifiers, however, the performance of our work has outperformed their achieved result. In another research, Abdul et al (Abdul *et al.*, 2020) used a temporal feature based on MFCC and GTCC that was extracted from both input and output accelerometer sensors and then fed to the LSTM model. L. Jing et al. (Jing *et al.*, 2017) developed a CNN model to learn features from frequency data of vibration signals instead of the raw data directly. The whole achieved accuracies have been presented in Table 4.14 and the proposed concat-SVM model shows the capability to achieve the best result.

Table 4.14: The accuracies of the cancat-SVM and the state-of-art models for helical gear

	Features	Accuracy%
Abdul et al. (Abdul *et al.*, 2016)	LBP_SVM	81.00
	LBP_k-NN	90.00
Abdul et al (Abdul *et al.*, 2020)	MFCC-GTCC-LSTM	99.32
L. Jing et al. (Jing *et al.*, 2017)	CNN + Learning feature	99.33
	CNN + manual feature	91.21
Proposed model(GTCC optimized-SVM)	**GTCC**	**99.75**
Proposed model (concat-SVM)	**Concatenating MFCC and GTCC with SVM**	**99.81**

For the spur case (in the PHM09 dataset), C. Wu et al. (Wu *et al.*, 2019) used the same dataset adopted in this work and proposed a 1-D CNN model which was fed by the raw data directly and their result shows that the performance of the 1-D CNN is better than the traditional machine learning algorithm for a

fixed-shaft gearbox and planetary gearbox fault diagnosis. However, the proposed concat-SVM shows better performance than this model as well (see Table 4.14).

Table 4.15: The accuracies of the cancat-SVM and the state-of-art models for spur gear

model	Features	Accuracy%
C. Wu et al. (Wu *et al.*, 2019)	1-DCNN	99.33
The proposed (concat-SVM)	**Concatenating MFCC and GTCC with SVM**	**100**
The proposed (GTCC optimized-SVM)	**GTCC**	**99.75**

4.6 Summary of the Chapter

In this chapter, three models for automatic fault detection have been suggested: optimized feature-SVM, Stat-SVM, and Concat-SVM. The features used, namely MFCC and GTCC, were originally designed for extracting features from speech signals and their parameters were fine-tuned accordingly. Our proposal involves optimizing these parameters to make them more suitable for gear fault detection. We utilized two optimization methods, GWO and FDO, and found that the performance of MFCC remained unchanged. However, the optimized GTCC exhibited superior performance compared to the default GTCC.

This led us to expect that the statistical values of both features could be valuable. We extracted statistical values from the concatenated MFCC and

GTCC and fed them to the SVM. The results indicated that the statistical features may lose some relevant information since certain statistical values are sensitive to feature distribution, such as mean, average, standard deviation, variance, and root mean square.

To prevent information loss, we proposed a technique called "frame concatenation," where all frames of each feature (MFCC and GTCC) are linked together in one series. The concatenated frames of MFCC are then connected to the concatenated frames of GTCC, resulting in an augmented feature. This augmented feature is fed to the SVM for gear fault detection. This approach, known as concat-SVM, not only preserves information from all frames but also increases the feature dimensions, which benefits the SVM as it performs better with higher-dimensional features. Our results demonstrate that concat-SVM outperforms existing literature and all other implemented models in this dissertation.

CHAPTER FIVE

5 CONCLUSION AND FUTURE WORK

5.1 Conclusion

To meet the growing demand for reliable and accurate monitoring systems, there is a significant interest in developing automatic fault detection and diagnosis methods for rotating machinery. These methods are crucial for ensuring high efficiency in modern industrial systems. There are some key factors to develop an automatic fault detection model. One of the key factors in achieving accurate fault detection and diagnosis is the extraction of fault features and the selection of appropriate features to train a machine learning algorithm. The second main factor is selecting a proper machine learning algorithm in terms of the behavior of the utilized feature whether the feature is a time series feature or non-time series feature.

In this dissertation, both time-series and non-time series-based representation models have been investigated using the cepstrum features MFCC and GTCC. Based on the literature, both features (especially MFCC) are useful for many pattern recognition systems such as fault detection and identification. Time series classification models (LSTM and ESN) and non-time series (SVM) classification models have been trained by the obtained feature (MFCC and GTCC) with two different forms including concatenating and statistical form for gear fault detection.

For the time series model, two models have been adopted namely, LSTM and ESN. The model based on the LSTM is trained by concatenating both features. In the beginning, the combination of whole channels of the vibration signal has been used for gear fault detection. The performance of the model is reasonable however, it is very time-consuming in the training phase and takes

more memory space during the implementation. In order to reduce the time consumption, we have used one channel of the vibration signal for training the LSTM model. but the performance of the LSTM faced a little degradation which might be due to an underfitting problem.

To overcome the time-consuming training process of the LSTM model, we have focused on the ESN model, which is a random weight-based reservoir model inspired by the RNN. The ESN model is less time-consuming as some of the parameters are selected randomly. In this dissertation, the ESN model has been fed by the concatenated features (MFCC and GTCC) with the whole channels of the vibration. However, the accuracy of the ESN model for gear fault detection is not sufficiently reliable to be suitable for real-life applications. The ESN model was then trained via individual channels of the vibration signal and the result improved significantly. The performance of the model for some channels outperformed the other channels and this might be due to the high vibration rate occurring in that channel. However, there is no guarantee to know the location of the fault, and this led us to know which of the channels is more related to the fault. To get a reliable result regardless of the location of the fault, we have enhanced the ESN model by adding two other reservoirs to the standard ESN which has only one reservoir. Each of the reservoirs is responsible for increasing the variance of the data. The enhanced ESN fed each of the three reservoirs by one channel and late fusion of the representation is adopted, the resulting improvement proves the usefulness of this approach.

Regarding the non-time series analysis, three automatic fault detection models have been proposed namely optimized feature-SVM, Stat-SVM, and Concat-SVM. Both features (MFCC and GTCC) were originally developed for extracting features from the speech signal and their parameter was tuned based on the same aspect. We proposed to optimize their parameters to become more suitable for gear fault detection. Based on the two optimization methods (GWO

and FDO), the performance of the MFCC does not change. However, the performance of the optimized GTCC outperforms the default GTCC. We then think that the statistical values of both features might be useful. The extracted statical values are extracted from the concatenated MFCC and GTCC and fed to the SVM. Based on the result, statistical features might lose some information from the feature where they are most relevant information since some of the statistical values are sensitive to the distribution of the features for instance mean, average, standard deviation, variance, and root mean square.

To avoid losing the information, we proposed to link all the frames' features (MFCC and GTCC) together in a vector (called frame concatenation). The obtained frame concatenation of MFCC is linked to the obtained frame concatenation of GTCC. The gained feature is then fed to the SVM to detect faults in gears. This technique (concat-SVM) besides keeping information from all frames, shows the capability to increase the dimensions of features which is beneficial for SVM as the SVM works better in high dimensions features. The result shows that the concat-SVM outperforms the state of arts in the literature and also all adopted models that have been implemented in this dissertation. This may ensure the usefulness of all of the frame feature values in the whole signal.

Based on the experimental results in this dissertation, Concat-SVM outperformed all other adopted models and state of arts in literature. This implies that the temporal relationships among the sample values within the vibration signal may not be crucial for detecting gear faults. The Concat-SVM is less time consuming compared to all other model that we have adopted that leads to the proposed model will be useful to be employed in a continuous learning environment.

5.2 Future work

In the future work, we are going to explore the performance of our framework in the fault detection tasks. For example, the ESN can be further enhanced by conducting booting method for the reservoir layers that might be useful to predict gear fault detection in earlier stage. We can also use other sensors to detect the location of fault which is useful to know which of the vibration channels to be utilized for gear fault detection.

REFERENCES

Abdul, Z.K. (2019), “Kurdish speaker identification based on one dimensional convolutional neural network”, Vol. 7 No. 4 (special issue), pp. 566–572.

Abdul, Z.K., Al-Talabani, A. and Abdulrahman, A.O. (2016), “A new feature extraction technique based on 1D local binary pattern for gear fault detection”, *Shock and Vibration*, Vol. 2016, doi: 10.1155/2016/8538165.

Abdul, Z.K., Al-Talabani, A.K. and Ramadan, D.O. (2020), “A hybrid temporal feature for gear fault diagnosis using the long short term memory”, *IEEE Sensors Journal*, Vol. 23, pp. 14444–14452, doi: 10.1109/jsen.2020.3007262.

Abdul, Z.K. and Talabani, A.K. Al. (2022), “Highly Accurate Gear Fault Diagnosis Based on Support Vector Machine”, *Journal of Vibration Engineering & Technologies*, Springer Nature Singapore, No. 0123456789, doi: 10.1007/s42417-022-00768-6.

Abdullah, J.M. (2019), “Fitness Dependent Optimizer : Inspired by the Bee Swarming Reproductive Process”, *IEEE Access*, IEEE, Vol. 7, pp. 43473–43486, doi: 10.1109/ACCESS.2019.2907012.

Adiga, A., Magimai, M. and Seelamantula, C.S. (2013), “Gammatone wavelet cepstral coefficients for robust speech recognition”, *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, doi: 10.1109/TENCON.2013.6718948.

Aherwar, A. (2012), “An investigation on gearbox fault detection using vibration analysis techniques: A review”, *Australian Journal of Mechanical Engineering*, Vol. 10 No. 2, pp. 169–183, doi: 10.7158/m11-830.2012.10.2.

Akpudo, U.E. and Hur, J.W. (2021), “A cost-efficient mfcc-based fault detection and isolation technology for electromagnetic pumps”, *Electronics (Switzerland)*, Vol. 10 No. 4, pp. 1–21, doi: 10.3390/electronics10040439.

Al-Talabani, A. (2015), *Automatic Speech Emotion Recognition-Feature Space Dimensionality and Classification Challenges*, University of Buckingham.

Al-Talabani, A., Sellahewa, H. and Jassim, S.A. (2015), "Emotion recognition from speech: tools and challenges", *Mobile Multimedia/Image Processing, Security, and Applications 2015*, Vol. 9497, p. 94970N.

Amarnath, M. and Praveen Krishna, I.R. (2014), "Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis", *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, Vol. 58 No. December, pp. 154–164, doi: 10.1016/j.measurement.2014.08.015.

Andrew F. Geib, Chung Chieh Kuo, Martin Gawecki, EnShuo Tsau, Je Won Kang, P.R.S. (2014), "MFCC and celp to detect turbine engine faults", Vol. 2 No. 12.

Asuhaimi Mohd Zin, A., Saini, M., Mustafa, M.W., Sultan, A.R. and Rahimuddin. (2015), "New algorithm for detection and fault classification on parallel transmission line using DWT and BPNN based on Clarke's transformation", *Neurocomputing*, Elsevier, Vol. 168, pp. 983–993, doi: 10.1016/j.neucom.2015.05.026.

Bai, M. (2021), "Long Short-Term Memory Network-Based Normal Pattern".

Belkacemi, B., Saad, S., Ghemari, Z., Zaamouche, F. and Khazzane, A. (2020), "Detection of induction motor improper bearing lubrication by discrete wavelet transforms (DWT) decomposition", *Instrumentation Mesure Metrologie*, Vol. 19 No. 5, pp. 347–354, doi: 10.18280/i2m.190504.

Benkedjouh T., Chettibi T., Saadouni Y., A.M. (2018), *Gearbox Fault Diagnosis Bbased on Mel-Frequency Cepstral Coefficients and Support Vector Machine*, Springer International Publishing, doi: 10.1007/978-3-319-89743-1.

Bianchi, F.M., Scardapane, S., Lokse, S. and Jenssen, R. (2021), "Reservoir computing approaches for representation and classification of multivariate time series", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32 No. 5, pp. 2169–2179, doi: 10.1109/TNNLS.2020.3001377.

Brkovic, A., Gajic, D., Gligorijevic, J., Savic-Gajic, I., Georgieva, O. and Di Gennaro, S. (2017), "Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery", *Energy*, Elsevier Ltd, Vol. 136, pp. 63–71, doi: 10.1016/j.energy.2016.08.039.

Cabrera, D., Sancho, F., Li, C., Cerrada, M., Sánchez, R.V., Pacheco, F. and de Oliveira, J.V. (2017), "Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation", *Applied Soft Computing Journal*, Vol. 58, pp. 53–64, doi: 10.1016/j.asoc.2017.04.016.

Candanedo, I.S., Nieves, E.H., González, S.R., Martín, M.T.S. and Briones, A.G. (2018), "Machine learning predictive model for industry 4.0", *Communications in Computer and Information Science*, Vol. 877 No. July, pp. 501–510, doi: 10.1007/978-3-319-95204-8_42.

Cerrada, M., Sánchez, R.V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J., Vásquez, R.E., *et al.* (2017), "Neural network-based survey analysis of risk management practices in new product development", *Mechanical Systems and Signal Processing*, Vol. 72–73 No. 1, pp. 1–18, doi: 10.3390/s17020273.

Chahal, A. and Gulia, P. (2019), "Machine learning and deep learning", *International Journal of Innovative Technology and Exploring Engineering*, Electronic Markets, Vol. 8 No. 12, pp. 4910–4914, doi: 10.35940/ijitee.L3550.1081219.

Chen, B., Li, Y., Zeng, N. and He, W. (2019), "Fractal lifting wavelets for machine fault diagnosis", *IEEE Access*, IEEE, Vol. 7, pp. 50912–50932,

doi: 10.1109/ACCESS.2019.2908213.

Chen, L., Liu, L., He, M. and Liu, D. (2019), “Gearbox fault diagnosis based on VMD and acoustic emission technology”, *I2MTC 2019 - 2019 IEEE International Instrumentation and Measurement Technology Conference, Proceedings*, IEEE, Vol. 2019-May, pp. 1–6, doi: 10.1109/I2MTC.2019.8826954.

Choi, D.J., Han, J.H., Park, S.U. and Hong, S.K. (2020), “Comparative Study of CNN and RNN for Motor fault Diagnosis Using Deep Learning”, *2020 IEEE 7th International Conference on Industrial Engineering and Applications, ICIEA 2020*, pp. 693–696, doi: 10.1109/ICIEA49774.2020.9102072.

Clive Jennings. (2020), “Split Bearings Improve Safety And Reduce Maintenance”, *Modern Slavery Statement*.

Dai, J., Venayagamoorthy, G.K. and Harley, R.G. (2009), “An introduction to the echo state network and its applications in power system”, *2009 15th International Conference on Intelligent System Applications to Power Systems, ISAP ’09*, No. 2, pp. 1–7, doi: 10.1109/ISAP.2009.5352913.

Dai, S., Niu, D. and Li, Y. (2018), “Daily Peak Load Forecasting Based on Complete Adaptive Noise and Support Vector Machine Optimized by Modified Grey Wolf”, doi: 10.3390/en11010163.

Decker, H.J. and Lewicki, D.G. (2003), *Spiral Bevel Pinion Crack Detection in a Helicopter Gearbox, Nasa Tm-2003-212327 Arl-Tr-2958*.

Dhamande, L.S. and Chaudhari, M.B. (2016), “Detection of Combined Gear-Bearing Fault in Single Stage Spur Gear Box Using Artificial Neural Network”, *Procedia Engineering*, Elsevier B.V., Vol. 144, pp. 759–766, doi: 10.1016/j.proeng.2016.05.082.

Dimitriadis, D., Maragos, P. and Potamianos, A. (2011), “On the effects of

filterbank design and energy computation on robust speech recognition", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 19 No. 6, pp. 1504–1516, doi: 10.1109/TASL.2010.2092766.

Edwards, S., Lees, A.W. and Friswell, M.I. (1998), "Fault diagnosis of rotating machinery", *Shock and Vibration Digest*, doi: 10.1177/058310249803000102.

Eyben, F., Wöllmer, M. and Schuller, B. (2009), "OpenEAR - introducing the munich open-source emotion and affect recognition toolkit", *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, doi: 10.1109/ACII.2009.5349350.

Fang, J. and Liu, P. (2020), "Personalized Power Consumption Control Method for Intelligent Buildings That Perceiving Speech Recognition Identity", *IOP Conference Series: Earth and Environmental Science*, Vol. 510 No. 2, doi: 10.1088/1755-1315/510/2/022007.

Farhat, N.H. (1992), "Photonit neural networks and learning mathines the role of electron-trapping materials", *IEEE Expert-Intelligent Systems and Their Applications*, Vol. 7 No. 5, pp. 63–72, doi: 10.1109/64.163674.

Faris, H., Aljarah, I., Al-Betar, M.A. and Mirjalili, S. (2018), "Grey wolf optimizer: a review of recent variants and applications", *Neural Computing and Applications*, Springer London, Vol. 30 No. 2, pp. 413–435, doi: 10.1007/s00521-017-3272-5.

Fulcher, B.D. (2017), "Feature-based time-series analysis", pp. 1–28.

Gaikwad, S.K., Gawali, B.W. and Yannawar, P. (2010), "A review on speech recognition technique", *International Journal of Computer Applications*, Vol. 10 No. 3, pp. 16–24, doi: 10.5120/1462-1976.

Guifan, Z. (2022), "Fault Diagnosis Method of Rotating Machinery Based on Collaborative Hybrid Metaheuristic Algorithm to Optimize VMD", *Journal*

of Sensors, Vol. 2022, doi: 10.1155/2022/8054801.

Henriquez, P., Alonso, J.B., Ferrer, M.A. and Travieso, C.M. (2014), "Review of automatic fault diagnosis systems using audio and vibration signals", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 44 No. 5, pp. 642–652, doi: 10.1109/TSMCC.2013.2257752.

Hoang, D.T. and Kang, H.J. (2019), "A survey on Deep Learning based bearing fault diagnosis", *Neurocomputing*, Elsevier B.V., Vol. 335, pp. 327–335, doi: 10.1016/j.neucom.2018.06.078.

Hu, H., Peng, G., Wang, X. and Zhou, Z. (2018), "Weld defect classification using 1-D LBP feature extraction of ultrasonic signals", *Nondestructive Testing and Evaluation*, Taylor & Francis, Vol. 33 No. 1, pp. 92–108, doi: 10.1080/10589759.2017.1299732.

Ibrahim, A., Anayi, F., Packianather, M. and Alomari, O.A. (2022), "New Hybrid Invasive Weed Optimization and Machine Learning Approach for Fault Detection", *Energies*, Vol. 15 No. 4, doi: 10.3390/en15041488.

Ibrahim, H., Loo, C.K. and Alnajjar, F. (2021a), "Grouped Echo State Network with Late Fusion for Speech Emotion Recognition", pp. 1–12.

Ibrahim, H., Loo, C.K. and Alnajjar, F. (2021b), "Speech emotion recognition by late fusion for bidirectional reservoir computing with random projection", *IEEE Access*, Vol. 9, pp. 122855–122871, doi: 10.1109/ACCESS.2021.3107858.

Inc, S.E. (2023), "What is an intuitive explanation of Echo State Networks?", *Stack Exchange Network*, available at: https://stats.stackexchange.com/q/261735 (accessed 19 November 2021).

Isermann, R. (1993), "Fault diagnosis of machines via parameter estimation and knowledge processing-Tutorial paper", *Automatica*, Vol. 29 No. 4, pp. 815–835, doi: 10.1016/0005-1098(93)90088-B.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.A. (2019), "Deep learning for time series classification: a review", *Data Mining and Knowledge Discovery*, Springer US, Vol. 33 No. 4, pp. 917–963, doi: 10.1007/s10618-019-00619-1.

Jaeger, H. and Haas, H. (2004), "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication", *Science*, Vol. 304 No. 5667, pp. 78–80, doi: 10.1126/science.1091277.

Jena, D.P., Sahoo, S. and Panigrahi, S.N. (2014), "Gear fault diagnosis using active noise cancellation and adaptive wavelet transform", *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, Vol. 47 No. 1, pp. 356–372, doi: 10.1016/j.measurement.2013.09.006.

Jiao, J., Zhao, M., Lin, J. and Zhao, J. (2018), "A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes", *Knowledge-Based Systems*, Elsevier, Vol. 160, pp. 237–250.

Jin, S., Wang, X., Du, L. and He, D. (2021), "Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN", *Applied Acoustics*, Vol. 172, p. 107562, doi: https://doi.org/10.1016/j.apacoust.2020.107562.

Jing, L., Zhao, M., Li, P. and Xu, X. (2017), "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox", *Measurement*, Elsevier Ltd, Vol. 111 No. July, pp. 1–10, doi: 10.1016/j.measurement.2017.07.017.

Kang, Y., Wang, C.-C., Chang, Y.-P., Hsueh, C.-C. and Chang, M.-C. (2006), "Certainty improvement in diagnosis of multiple faults by using versatile membership functions for fuzzy neural networks", *Advances in Neural Networks-ISNN 2006: Third International Symposium on Neural Networks, Chengdu, China, May 28-June 1, 2006, Proceedings, Part III 3*, pp. 370–

375.

Karabacak, Y.E., Özmen, N.G. and Gümü\csel, L. (2022), “Intelligent worm gearbox fault diagnosis under various working conditions using vibration, sound and thermal features”, *Applied Acoustics*, Elsevier, Vol. 186, p. 108463.

Kateris, D., Moshou, D., Pantazi, X.E., Gravalos, I., Sawalhi, N. and Loutridis, S. (2014), “A machine learning approach for the condition monitoring of rotating machinery”, *Journal of Mechanical Science and Technology*, Vol. 28 No. 1, pp. 61–71, doi: 10.1007/s12206-013-1102-y.

Kathiresan, T. and Dellwo, V. (2019), “Cepstral derivatives in MFCCS for emotion recognition”, *2019 IEEE 4th International Conference on Signal and Image Processing, ICSIP 2019*, IEEE, pp. 56–60, doi: 10.1109/SIPROCESS.2019.8868573.

Kemalkar, A.K. and Bairagi, V.K. (2017), “Engine fault diagnosis using sound analysis”, *International Conference on Automatic Control and Dynamic Optimization Techniques, ICACDOT 2016*, pp. 943–946, doi: 10.1109/ICACDOT.2016.7877726.

Khan, S.A. and Kim, J.-M. (2016), “Rotational speed invariant fault diagnosis in bearings using vibration signal imaging and local binary patterns”, *The Journal of the Acoustical Society of America*, Vol. 139 No. 4, pp. EL100–EL104, doi: 10.1121/1.4945818.

Klein, R., Rudyk, E., Masad, E. and Issacharoff, M. (2011), “Model based approach for identification of gears and bearings failure modes”, *International Journal of Prognostics and Health Management*, Vol. 2 No. 2, pp. 1–10.

Kumar, P. and Hati, A.S. (2020), “Review on machine learning algorithm based fault detection in induction motors”, *Archives of Computational Methods in*

Engineering, Springer Netherlands, No. 0123456789, doi: 10.1007/s11831-020-09446-w.

Lei, J., Liu, C. and Jiang, D. (2019), "Fault diagnosis of wind turbine based on long short-term memory networks", *Renewable Energy*, Elsevier Ltd, Vol. 133, pp. 422–432, doi: 10.1016/j.renene.2018.10.031.

Lei, Y., Kong, D., Lin, J. and Zuo, M.J. (2012), "Fault detection of planetary gearboxes using new diagnostic parameters", *Measurement Science and Technology*, Vol. 23 No. 5, doi: 10.1088/0957-0233/23/5/055605.

Li, X., Bi, F., Zhang, L., Lin, J., Bi, X. and Yang, X. (2022), "Rotating machinery faults detection method based on deep echo state network", *Applied Soft Computing*, Elsevier B.V., Vol. 127, p. 109335, doi: 10.1016/j.asoc.2022.109335.

Liu, R., Yang, B., Zio, E. and Chen, X. (2018), "Artificial intelligence for fault diagnosis of rotating machinery: A review", *Mechanical Systems and Signal Processing*, Elsevier Ltd, Vol. 108, pp. 33–47, doi: 10.1016/j.ymssp.2018.02.016.

Livani, H. and Evrenosoglu, C.Y. (2012), "A fault classification method in power systems using DWT and SVM classifier", *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, pp. 1–5, doi: 10.1109/TDC.2012.6281686.

Luo, R.C. and Wang, H. (2018), "Diagnostic and prediction of machines health status as exemplary best practice for vehicle production system", *IEEE Vehicular Technology Conference*, IEEE, Vol. 2018-Augus, pp. 1–5, doi: 10.1109/VTCFall.2018.8690710.

Ma, S., Cai, W., Liu, W., Shang, Z. and Liu, G. (2019), "A lighted deep convolutional neural network based fault diagnosis of rotating machinery", *Sensors (Switzerland)*, Vol. 19 No. 10, pp. 1–20, doi: 10.3390/s19102381.

Manohar, M., Koley, E., Kumar, Y. and Ghosh, S. (2018), "Discrete Wavelet Transform and kNN-Based Fault Detector and Classifier for PV Integrated Microgrid", *Lecture Notes in Networks and Systems*, Vol. 38, pp. 19–28, doi: 10.1007/978-981-10-8360-0_2.

Mansor, M.A., Kasihmuddin, M.S.M. and Sathasivam, S. (2021), "Grey Wolf Optimization algorithm with Discrete Hopfield Neural Network for 3 Satisfiability analysis", *Journal of Physics: Conference Series*, Vol. 1821 No. 1, doi: 10.1088/1742-6596/1821/1/012038.

Matlab. (2022), "gammatoneFilterBank", *Mathwork*, available at: https://uk.mathworks.com/help/audio/ref/gammatonefilterbank-system-object.html.

Mec.Edu teams. (2020), "Gears: Types of Gears & Types of Gear failures – Mechanical Education", *Mechanical Education*.

Mirjalili, S., Mirjalili, S.M. and Lewis, A. (2014), "Grey Wolf Optimizer", *Advances in Engineering Software*, Elsevier Ltd, Vol. 69, pp. 46–61, doi: 10.1016/j.advengsoft.2013.12.007.

Mittal, N., Singh, U. and Sohi, B.S. (2016), "Modified Grey Wolf Optimizer for Global Engineering Optimization", *Applied Computational Intelligence and Soft Computing*, Vol. 2016, pp. 1–16, doi: 10.1155/2016/7950348.

Mohammed, H.M., Abdul, Z.K., Rashid, T.A., Alsadoon, A. and Bacanin, N. (2021), "A new K-means grey wolf algorithm for engineering problems", *World Journal of Engineering*, Emerald Publishing Limited.

Muralidharan, V. and Sugumaran, V. (2013), "Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump", *Measurement: Journal of the International Measurement Confederation*, Elsevier Ltd, Vol. 46 No. 1, pp. 353–359, doi: 10.1016/j.measurement.2012.07.007.

Nelwamondo, F. V. and Marwala, T. (2006), "Faults detection using Gaussian mixture models, mel-frequency cepstral coefficients and kurtosis", *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, Vol. 1 No. 2, pp. 290–295, doi: 10.1109/ICSMC.2006.384397.

Norton, M.P. and Nelson, F.C. (1990), *Fundamentals of Noise and Vibration Analysis for Engineers*, *The Journal of the Acoustical Society of America*, Vol. 88, doi: 10.1121/1.400194.

Pan, C., Jin, A., Yang, W. and Zhang, Y. (2022), "Early Detection of Network Fault Using Improved Gray Wolf Optimization and Wavelet Neural Network", *Mathematical Problems in Engineering*, Vol. 2022, doi: 10.1155/2022/1235229.

Panda, M. and Das, B. (2019), "Grey Wolf Optimizer and Its Applications: A Survey", *Lecture Notes in Electrical Engineering*, Springer, Singapore, pp. 179–194, doi: 10.1007/978-981-13-7091-5_17.

PHM. (2009), "2009 PHM challenge competition data set", *The Prognostics and Health Management Society*, available at: https://phmsociety.org/public-data-sets/ (accessed 25 October 2021).

Poveda-Martínez, P. and Ramis-Soriano, J. (2020), "A comparison between psychoacoustic parameters and condition indicators for machinery fault diagnosis using vibration signals", *Applied Acoustics*, Vol. 166, pp. 1–13, doi: 10.1016/j.apacoust.2020.107364.

Qi, J., Wang, D., Xu, J. and Tejedor, J. (2013), "Bottleneck features based on gammatone frequency cepstral coefficients", *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, No. August, pp. 1751–1755.

Qi, M., Zhou, R., Zhang, Q. and Yang, Y. (2021), "Feature Classification

Method of Frequency Cepstrum Coefficient Based on Weighted Extreme Gradient Boosting", *IEEE Access*, Vol. 9, pp. 72691–72701, doi: 10.1109/ACCESS.2021.3079286.

Qiao, W. and Lu, D. (2015), "A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis - Part I: Components and Subsystems", *IEEE Transactions on Industrial Electronics*, Vol. 62 No. 10, pp. 6536–6545, doi: 10.1109/TIE.2015.2422112.

Qu, Y., Bechhoefer, E., He, D. and Zhu, J. (2013), "A new acoustic emission sensor based gear fault detection approach", *International Journal of Prognostics and Health Management*, Vol. 4 No. SPECIAL ISSUE 2, pp. 1–14.

Rafiee, J., Rafiee, M.A. and Tse, P.W. (2010), "Application of mother wavelet functions for automatic gear and bearing fault diagnosis", *Expert Systems with Applications*, Elsevier Ltd, Vol. 37 No. 6, pp. 4568–4579, doi: 10.1016/j.eswa.2009.12.051.

Revathi, P., Reddy, P.M.M., Srinivas, V.K.S. and Sudhakara, E. (2014), "International Journal of Research in Computer Applications and Robotics Issn 2320-7345 Constrained Seasonal Group", Vol. 2 No. 8, pp. 79–91.

Riaz, S., Elahi, H., Javaid, K. and Shahzad, T. (2017), "Vibration Feature Extraction and Analysis for Fault Diagnosis of Rotating Machinery-A Literature Survey", *Asia Pacific Journal of Multidisciplinary Research*, Vol. 5 No. 1, pp. 103–110.

Salameh, J.P., Cauet, S., Etien, E., Sakout, A. and Rambault, L. (2018), "Gearbox condition monitoring in wind turbines: A review", *Mechanical Systems and Signal Processing*, Elsevier Ltd, Vol. 111, pp. 251–264, doi: 10.1016/j.ymssp.2018.03.052.

Salih, J.F., Mohammed, H.M. and Abdul, Z.K. (2022), "Modified Fitness

Dependent Optimizer for Solving Numerical Optimization Functions", *IEEE Access*, Vol. 10, pp. 83916–83930, doi: 10.1109/ACCESS.2022.3197290.

Samanta, B. (2004), "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms", *Mechanical Systems and Signal Processing*, Vol. 18 No. 3, pp. 625–644, doi: 10.1016/S0888-3270(03)00020-7.

Saravanan, N., Kumar Siddabattuni, V.N.S. and Ramachandran, K.I. (2008), "A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box", *Expert Systems with Applications*, Vol. 35 No. 3, pp. 1351–1366, doi: 10.1016/j.eswa.2007.08.026.

Saravanan, N. and Ramachandran, K.I. (2009), "Fault diagnosis of spur bevel gear box using discrete wavelet features and Decision Tree classification", *Expert Systems with Applications*, Elsevier Ltd, Vol. 36 No. 5, pp. 9564–9573, doi: 10.1016/j.eswa.2008.07.089.

Saufi, S.R., Ahmad, Z.A. Bin, Leong, M.S. and Lim, M.H. (2020), "Gearbox fault diagnosis using a deep learning model with limited data sample", *IEEE Transactions on Industrial Informatics*, Vol. 16 No. 10, pp. 6263–6271, doi: 10.1109/TII.2020.2967822.

Seema and Kumar, V. (2016), "Modified Grey Wolf Algorithm for optimization problems", *2016 International Conference on Inventive Computation Technologies (ICICT)*, Vol. 3, IEEE, pp. 1–5, doi: 10.1109/INVENTIVE.2016.7830162.

Shao, S., McAleer, S., Yan, R. and Baldi, P. (2019), "Highly accurate machine fault diagnosis using deep transfer learning", *IEEE Transactions on Industrial Informatics*, IEEE, Vol. 15 No. 4, pp. 2446–2455, doi: 10.1109/TII.2018.2864759.

Shao, S., Member, S., Mcaleer, S., Yan, R. and Member, S. (2018), "Highly-accurate machine fault diagnosis using deep transfer learning", *IEEE Transactions on Industrial Informatics*, IEEE, Vol. PP No. c, p. 1, doi: 10.1109/TII.2018.2864759.

Sharma, M. and Kaur, P. (2021), "A Comprehensive Analysis of Nature-Inspired Meta-Heuristic Techniques for Feature Selection Problem", *Archives of Computational Methods in Engineering*, Springer Netherlands, Vol. 28 No. 3, pp. 1103–1127, doi: 10.1007/s11831-020-09412-6.

Sharma, V. and Parey, A. (2016), "A review of gear fault diagnosis using various condition indicators", *Procedia Engineering*, The Author(s), Vol. 144, pp. 253–263, doi: 10.1016/j.proeng.2016.05.131.

Sharma, V. and Parey, A. (2017), "Case study on the effectiveness of gear fault diagnosis technique for gear tooth defects under fluctuating speed", *IET Renewable Power Generation*, Vol. 11 No. 14, pp. 1841–1849, doi: 10.1049/iet-rpg.2016.0639.

Smith, J.D. (2003), *Gear Noise and Vibration*, *Gear Noise and Vibration*, doi: 10.1201/9781482276275.

Sorsa, T., Koivo, H.N. and Koivisto, H. (1991), "Neural Networks in Process Fault Diagnosis", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21 No. 4, pp. 815–825, doi: 10.1109/21.108299.

Stamboliska, Z., Rusiński, E. and Moczko, P. (2015), *Proactive Condition Monitoring of Low-Speed Machines*, *Proactive Condition Monitoring of Low-Speed Machines*, doi: 10.1007/978-3-319-10494-2.

Suphioglu, C., Blaher, B., Rolland, J.M., McCluskey, J., Schappi, G., Kenrick, J., Singh, M.B., *et al.* (1998), "Molecular basis of IgE-recognition of Lol p 5, a major allergen of rye- grass pollen", *Molecular Immunology*, Vol. 35 No. 5, pp. 293–305, doi: 10.1016/S0161-5890(98)00050-9.

Tang, S., Yuan, S. and Zhu, Y. (2020a), “Deep learning-based intelligent fault diagnosis methods toward rotating machinery”, *IEEE Access*, Vol. 8, pp. 9335–9346, doi: 10.1109/ACCESS.2019.2963092.

Tang, S., Yuan, S. and Zhu, Y. (2020b), “Convolutional neural network in intelligent fault diagnosis toward rotatory machinery”, *IEEE Access*, doi: 10.1109/ACCESS.2020.2992692.

Tang, S., Zhu, Y. and Yuan, S. (2022), “Intelligent fault diagnosis of hydraulic piston pump based on deep learning and Bayesian optimization”, *ISA Transactions*, Elsevier Ltd, Vol. 129, pp. 555–563, doi: 10.1016/j.isatra.2022.01.013.

Tiwari, V. (2010), “MFCC and its applications in speaker recognition”, *International Journal on Emerging Technologies*, Vol. 1 No. 1, pp. 19–22.

Verzelli, P., Alippi, C. and Livi, L. (2019), “Echo State Networks with Self-Normalizing Activations on the”, *Scientific Reports*, Springer US, pp. 1–14, doi: 10.1038/s41598-019-50158-4.

Wang, J., Zhao, R., Wang, D., Yan, R., Mao, K. and Shen, F. (2017), “Machine health monitoring using local feature-based gated recurrent unit networks”, *IEEE Transactions on Industrial Electronics*, Vol. 65 No. 2, pp. 1539–1548, doi: 10.1109/TIE.2017.2733438.

Wang, T., Han, Q., Chu, F. and Feng, Z. (2019), “Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox : A review”, *Mechanical Systems and Signal Processing*, Elsevier Ltd, Vol. 126, pp. 662–685, doi: 10.1016/j.ymssp.2019.02.051.

Watson, M., Sheldon, J., Amin, S., Lee, H., Byington, C. and Begin, M. (2007), “A comprehensive high frequency vibration monitoring system for incipient fault detection and isolation of gears, bearings and shafts/couplings in turbine engines and accessories”, *Proceedings of the ASME Turbo Expo*,

Vol. 5, pp. 885–894, doi: 10.1115/GT2007-27660.

Wei, D., Wang, K.S., Heyns, S. and Zuo, M.J. (2019), *Convolutional Neural Networks for Fault Diagnosis Using Rotating Speed Normalized Vibration*, *Applied Condition Monitoring*, Vol. 15, Springer International Publishing, doi: 10.1007/978-3-030-11220-2_8.

Wilk, A., Łazarz, B. and Madej, H. (2008), "Gear fault detection using vibration analysis", *Diagnostyka*, Vol. nr 3(47) No. 47, pp. 111–116.

Wu, C., Jiang, P., Ding, C., Feng, F. and Chen, T. (2019), "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network", *Computers in Industry*, Elsevier B.V., Vol. 108, pp. 53–61, doi: 10.1016/j.compind.2018.12.001.

Xiang, S., Qin, Y., Zhu, C., Wang, Y. and Chen, H. (2020), "LSTM networks based on attention ordered neurons for gear remaining life prediction", *ISA Transactions*, Elsevier Ltd, No. xxxx, doi: 10.1016/j.isatra.2020.06.023.

Yang, H.B., Zhang, J.A., Chen, L.L., Zhang, H.L. and Liu, S.L. (2019), "Fault diagnosis of reciprocating compressor based on Cconvolutional neural networks with multisource raw vibration signals", *Mathematical Problems in Engineering*, Vol. 2019, doi: 10.1155/2019/6921975.

Yang, R., Huang, M., Lu, Q. and Zhong, M. (2018), "Rotating Machinery Fault Diagnosis Using Long-short-term Memory Recurrent Neural Network", *IFAC-PapersOnLine*, Elsevier B.V., Vol. 51 No. 24, pp. 228–232, doi: 10.1016/j.ifacol.2018.09.582.

Yoshimatsu, O., Satou, Y. and Shibasaki, K. (2018), "Rolling bearing diagnosis based on CNN-LSTM and various condition dataset", *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*, pp. 1–8.

Yu, J. and Zhou, X. (2020), "One-dimensional residual convolutional

autoencoder based feature learning for gearbox fault diagnosis", *IEEE Transactions on Industrial Informatics*, IEEE, Vol. 16 No. 10, pp. 6347–6358, doi: 10.1109/TII.2020.2966326.

Zamanian, A.H. and Ohadi, A. (2016), "Gearbox fault detection through PSO exact wavelet analysis and SVM classifier", *ArXiv E-Prints*, doi: 10.13140/RG.2.1.4983.3442.

Zeng, C., Su, H., Li, Y., Guo, J. and Yang, C. (2021), "An approach for robotic leaning inspired by biomimetic adaptive control", *IEEE Transactions on Industrial Informatics*, IEEE.

Zeng, N., Qiu, H., Wang, Z., Liu, W., Zhang, H. and Li, Y. (2018), "A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease", *Neurocomputing*, Vol. 320, pp. 195–202, doi: https://doi.org/10.1016/j.neucom.2018.09.001.

Zhang, J., Jiang, Q. and Chang, F. (2018), "Fault diagnosis method based on MFCC fusion and SVM", *2018 IEEE International Conference on Information and Automation, ICIA 2018*, IEEE, No. August, pp. 1617–1622, doi: 10.1109/ICInfA.2018.8812417.

Zhang, S., Member, S., Zhang, S. and Member, S. (2020), "Deep Learning Algorithms for Bearing Fault Diagnostics – A Comprehensive Review".

Zhu, J., Nostrand, T., Spiegel, C. and Morton, B. (2014), "Survey of condition indicators for condition monitoring systems", *PHM 2014 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014*, pp. 635–647.