INSPECTION OF CORROSION ON THE UNDERSIDE OF RAILS BASED ON AN ACOUSTIC TECHNIQUE

by

JINGYUAN YANG

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Engineering

The University of Birmingham

2023

ABSTRACT

The railway is a common global transportation mechanism for people's daily life, work, and travel. As society has continued to develop, the volume of railway transportation has increased rapidly, making the railway an essential infrastructure in transportation systems. Ensuring the safety of railway operations is crucial for maintaining regular railway services.

Currently, condition monitoring approaches are being rapidly improved to automatically monitor regular railway operations. These approaches include ultrasonic, acoustic, vibration, electromagnetic, laser, and thermographic ones. Automatic railway inspection techniques have the advantage of being able to detect various types of faults using different approaches. However, each technique has its limitations when it comes to identifying corrosion on the underside of the rail effectively.

This PhD thesis focuses on the inspection of corrosion on the underside of the rail using acoustic techniques. To achieve this objective, it introduces methodologies for simulation, feature extraction, and fault classification.

Firstly, inspection of corrosion on the underside of the rail requires sufficient acoustic data for the analysis. However, it is difficult to obtain them from the real world. Hence, a multi-physical Finite Element Model (FEM) is designed to simulate the generation of an acoustic signal caused by the impact of a hammer on the rail. The thesis explains the method for coupling two different simulated physical fields and describes the validation of the simulation model through laboratory tests. A dataset of simulated signals is generated using the validated model.

And then, in order to extract features from the simulated signals obtained from the simulation model and laboratory tests, the thesis compares the effectiveness of different signal decomposition methods for analysis. The most effective decomposition method was enhanced through a proposed method for automatically choosing the appropriate initial parameters. By evaluating the decomposition results, the effectiveness of the proposed method was proved.

Furthermore, the thesis develops a novel integrated classification model to classify the different features of healthy and faulty rails. The results of the classification demonstrate that corrosion on the underside of the rail can be identified through the developed integrated 1-D CNN model based on acoustic impact response signals. Additionally, according to the classification, the simulation model is validated for identifying corrosion on the underside of the rail in real-world scenarios.

This thesis makes three main contributions:

1) The development of an FEM-based simulation model that creates signals that are difficult to obtain in the real world. This simulation model can identify corrosion on the underside of the rail in real-world scenarios.

2) The proposal of a signal decomposition method that effectively extracts features from the impact response signal.

3) The introduction of a classification method that enables the identification of types and levels of corrosion on the underside of the rail simultaneously.

ACKNOWLEDGEMENTS

I express my deepest appreciation to the Birmingham Centre for Railway Research and Education (BCRRE) at the University of Birmingham for providing me with a conducive research environment and the resources necessary to undertake my PhD. Moreover, I am indebted to following fellows for their contribution to my PhD work:

Firstly, I would like to extend my sincere gratitude to my supervisor, Prof. Edward Stewart, for his exceptional guidance and encouragement throughout my PhD. His extensive knowledge and insightful feedback have been instrumental in shaping my research ideas and refining my work. His unwavering dedication to my success has been a constant source of inspiration.

I am also grateful to Dr. Mani Entezami for sharing his precious data and valuable algorithm with me.

I am indebted to my colleagues at BCRRE for their kind assistance and support, particularly Mr. Adnan Zentani, who went above and beyond in helping me prepare for my experiments, including moving those heavy rails.

I am deeply grateful to my parents for giving birth to me and providing me with love and support. Their belief in my abilities has been a constant source of motivation for me, and I could not have completed this thesis without their guidance and support.

I am grateful to my office friends for their encouragement during my PhD programme, especially Mr. Junhui Huang, who played an instrumental role in helping me build the machine learning model. Their camaraderie and friendship have made my time in the programme all the more enjoyable, and their support has been invaluable during the more challenging moments.

Finally, I would like to thank my friends who have played a significant role in my PhD journey. Especially, I am deeply grateful for the unwavering support and enduring friendship of Jiawen Liu and Liyang Han, whose constant presence in my life has been a source of comfort and strength throughout my PhD career.

PUBLICATIONS

During my PhD career, I published three journal papers.

The contents of Section 5.2 have been reviewed and published in the journal IET Signal Processing (‘Decomposition methods for impact-based fault detection algorithms in railway inspection applications’).

The contents of Section 5.3 have been reviewed and published in the journal MDPI Applied Science (‘An improved VMD method for use with acoustic impact response signals to detect corrosion at the underside of railway tracks’).

The contents of Section 6.4 have been reviewed and accepted in the journal MDPI Applied Science (‘1-D CNN Based Detection and Localisation of Defective Droppers in Railway Catenary’)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1-D CNN	One-Dimensional Convolutional Neural Network
AC	Alternating Current
ACFM	Alternating Current Field Measurement
ADMM	Alternative Direction Method of Multipliers
AE	Acoustic Emission
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
DfT	Department for Transport
DTW	Dynamic Time Warping
ECG	Electrocardiogram
ECT	Eddy Current Testing
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
FEM	Finite Element Model
FFT	Fast Fourier Transform
FRF	Frequency Response Function
HHT	Hilbert–Huang Transform
HT	Hilbert Transformation
ICF	Initial Centre Frequency
IMF	Intrinsic Mode Function

IRJ	Insulated Rail Joint
LDA	Linear Discriminant Analysis
LMD	Local Mean Decomposition
MFL	Magnetic Flux Leakage
MLP	Multilayer Perceptron
MSD	Mean Value of the Standard Deviation
NADA	Noise Assisted Data Analysis
NI	National Instruments
NiN	Network in Network
PF	Product Function
PSO	Particle Swarm Optimisation
RBF	Radial Basis Function
RCF	Rolling Contact Fatigue
ReLU	Rectified Linear Function
SD	Standard Deviation
SDOF	Single Degree of Freedom
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TCRP	Transit Cooperative Research Program
TTCI	Transportation Technology Centre, Inc.
VGG	Visual Geometry Group
VMD	Variational Mode Decomposition
WT	Wavelet Transform

CHAPTER 1 INTRODUCTION

1.1 Background

The railway is one of the common global transportation mechanisms for people's living, working, and travelling. With the development of society, the volume of railway transportation has increased rapidly. Based on the statistics from the UK Government's Department for Transport (DfT) [1], railway usage in the UK is growing rapidly and has almost doubled in the past 20 years; from 1983 to before the Covid situation (2019), passenger journeys on the railway have grown from 56 to 270.6 million, and the number of miles travelled by vehicles has increased from 3.9 to 23.4 million. Rail transportation accounts for 8% of passenger transportation and 7% of product transportation in the world, and these proportions are still increasing steadily [2].

The railway is the safest land transportation in the world. One of the reasons for this is the efficient safety maintenance and management of the closed railway systems which contain various sub-systems such as control systems, communication, signalling, infrastructure, rolling stock, information technology, etc. The railway is a significant infrastructure in railway systems. The safety of railway operation is important to ensure regular railway operation.

Traditionally, railway inspection has been main implemented through manual inspection. For example, to identify missing components, an inspector used to check the components by walking along the track. Another example is the measurement of the gauge (spacing of the rails) using devices such as a route scan. The process is shown in Figure 1.1.

Figure 1.1 Measuring gauge width [3]

This manual inspection can be inefficient and subjective. It is challenging to meet the requirements for development of the railway network, and also incompatible with the rapid expansion of expectations on railway maintenance. Currently, condition monitoring approaches, which include ultrasonic, acoustic, vibration, electromagnetic, laser, and thermographic approaches, are being rapidly improved to automatically monitor during regular railway operation. These techniques, integrated with various physical sensors, generate digital signals or images that can represent the condition of the railway infrastructure. By processing these data, it is possible to identify the fault information such as the different types or levels of damage. After collecting enough features from faulty samples to generate a dataset, it is possible to predict the defect condition of the rail.

In conclusion, while current inspection techniques are often adequate for maintaining the basic operation of the railway, they do not provide sufficient detail for efficient railway asset management. Automatic inspection techniques can be used to improve the evaluation of the health condition of the railway faster, more accurately, and more efficiently, which ensures the regular operation of the railway, reduces maintenance, and protects trackworker safety.

1.2 Rail damage and rail corrosion

Varying categories of rail damage exist in the exterior or interior of the rails. Currently, automatic railway inspection techniques have the advantage that different approaches are efficient in detecting different types of faults. Regarding the potential faults on a railway, rolling contact fatigue (RCF) is one of the most serious types of damage. The Hatfield incident shown in Figure 1.2 occurred because of an RCF crack [4].

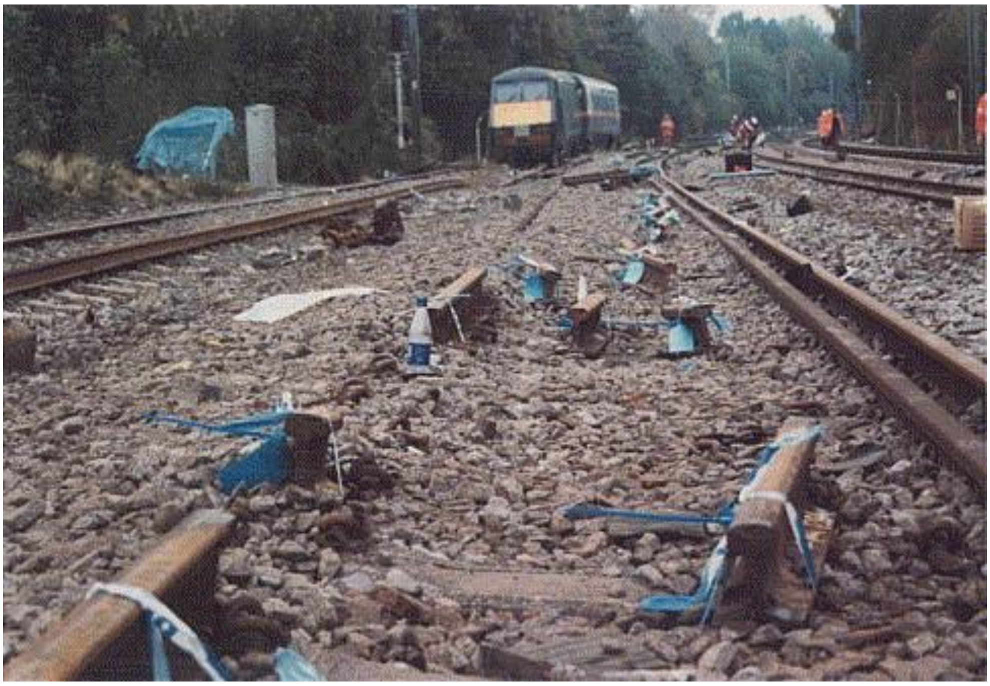


Figure 1.2 Derailment in Hatfield (2000) [4]

Besides RCF cracks, plastic flow [5], bolt hole cracks [6], and corrosion [7] also carry high risks for railway safety. Rail corrosion is a common fault occurring in plain-line railway tracks and has recently attracted much attention [8]. Rail corrosion will increase the risk of the rail fracture and derail. In order to protect the passenger, the relevant company will spend more money and time for the maintenance. It leads more train delay to the passengers [9] and economic loss of company [10].

It has been concluded that rail corrosion is caused by steel oxidation and stress corrosion. Steel oxidation is the most common situation occurring in a whole plain line, especially at positions with high rainfall. Stress corrosion usually occurs where stress exists (especially in the wheel–rail contact area). Micro-cracks will extend and change into obvious cracks as the stress extends and concentrates. The risk of rail corrosion leads to loss of rail material or even rail fracture.

Corrosion of the underside of the rail is commonly caused by oxidation. It generally has a large area on the underside of the rail. As corrosion is hidden on the bottom of the rail, it is difficult to detect. This potentially increases the risk of operational safety, but if detection systems could be developed significant cost savings and safety improvements could be made.

A corrosion detection system that could inspect the underside of the rail and highlight corrosion before incidents occurred could be used to reduce maintenance costs, optimise the railway operation through improved maintenance scheduling, and protect the safety of passengers and trackworkers. Such a system would first require an investigation into the environment (weather forecasts, historical rainfall data, accident reports, etc) to identify areas where the corrosion risk was significant and thus monitoring systems should be targeted. A device based on a non-destructive testing method for corrosion identification and quantification would then be deployed at the wayside in these areas. Such a device would use structural health monitoring techniques (signal processing, feature extraction, fault classification, machine learning, etc.) to analyse the information obtained from monitoring the rail in order to identify potential corrosion on the underside of the rail. This information could be communicated to a maintenance management centre, or directly reported to trackworkers deploying such a system.

To realise the concept summarised above, it is important to identify an efficient technique for the inspection of the underside of the rail.

1.3 Research gap

Current automatic railway inspection techniques have the advantages that different approaches are able to detect different types of faults. For example, ultrasonic inspection, the most common approach, is used for the inspection of deep surface-breaking and internal defects [11]. But it is not sensitive to small faults (RCF smaller than 4 mm [11]) and can also miss faults at the rail foot, especially corrosion defects. Ultrasonic inspection also requires direct access to the track, which constrains when it can be deployed and potentially puts trackworkers at risk. Visual inspection technology can be deployed remotely, and has advantages in detecting obvious surface faults (RCF faults, squats [12]). Information about the damage can be found directly by a camera and the technique involves no contact with the tracks. The limitation of visual inspection is the difficulty in exploring faults inside of the rail or at the bottom of the rail. It is hard to investigate corrosion on the underside of the rail (underside corrosion) by visual inspection because the damage cannot be observed by a camera or laser directly. Electromagnetic methods (including eddy current, magnetic flux leakage, etc) are suitable for detecting small cracks on the rail surface or near the surface. But it is difficult to detect deep faults which are inside of the rail, due to the skin effect [13]. If the underside corrosion is too large (corrosion area) or far from the measurement surface (e.g. corrosion occurring on the bottom of the rail when the measurement is made at the top), it is difficult to detect it using this technique. As with ultrasonic inspection, these techniques also require direct access to the track.

Considering all these elements, a non-contact, non-invasive, acoustic technique that doesn't require direct access to the track or a direct observation of the underside of the rail may be advantageous. By utilising acoustic sensors and monitoring equipment from a safe distance, inspections can be conducted without impacting rail operations. Secondly, acoustic inspection systems are quick to set up and deploy. This means that rail operators can quickly respond to emerging issues or conduct routine inspections without extensive planning and preparation, which can reduce the train delay and maintenance costs. Thirdly, acoustic inspection equipment is typically portable and modular. This mobility allows the maintainer to easily move the equipment to different sites as needed. Whether inspecting different sections of a rail network or addressing specific trouble spots, the flexibility of acoustic systems makes them highly versatile. Furthermore, many corrosion faults exist in wet conditions. Rail inspection devices located in these locations might be damaged due to the environment. Acoustic equipment avoids these risks of damage to the sensors or the system. It ensures that inspections can continue in locations that might be inaccessible to other sensor technologies. Additionally, acoustic approaches do not rely on physical contact with the rail infrastructure. They collect data remotely using sound waves, which means there is no risk of damaging the sensors or the rail components, unlike traditional methods involving physical contact sensors or instrumentation that might be damaged over time. Furthermore, in some rail inspection scenarios, placing sensors directly on the tracks might interfere with the operation of other sensor systems or cause damage to the sensors themselves. Acoustic approaches do not suffer from these constraints. They can be used at wayside without creating any interference with other rail infrastructure sensors or operations. This is particularly relevant for areas sensitive to impact site and other forces that could disrupt sensor readings.

Additionally, sufficient faulty samples are needed for research into fault inspection and prediction success. In the process of fault inspection, the fault size should be measured first, and useful information about the fault extracted through suitable inspection techniques. This is then used for visualisation from which it is possible to distinguish differences in the features of different types of faults. However, due to the challenging position of corrosion on the underside of the rail, it is difficult to measure the size and depth of the corrosion in the real world. In order to measure the extent of underside corrosion, the rail should ideally be taken away from the plain line, which would be hard to do in a deployed application.

Furthermore, after obtaining a sufficient number of samples of rail corrosion on the underside of the rail, it becomes challenging to classify these samples accurately, considering the various corrosion types and severity levels. Traditional methods may struggle to handle this task efficiently, as manual classification of each sample might waste time and be prone to errors. Machine learning presents itself as a highly effective technique for the classification. It offers several advantages, including automation, scalability, speed, consistency, adaptability to changing conditions, and the ability to handle multiclass classification. It is, therefore, a valuable tool for addressing the classification of corrosion at the underside of the rail.

In conclusion, the research gaps in this thesis includes three aspects. Firstly, there is a limitation for identifying the corrosion at the underside of the rail. Secondly, there is difficult to obtain sufficient faulty samples in a practical context, where obtaining direct access to the track is challenging. Thirdly, how to automatically identify and classify different corrosions at the underside of the rail.

1.4 Project concept

The concept of this project is to combine the advantages of the acoustic technique with proposed methods of inspecting the underside of the rail to automatically identify corrosion in that area. This would support objectives of reducing maintenance costs, enhancing normal railway operations, and ensuring passenger and trackworker safety. The strategy is that microphones are deployed at high-risk areas along the railway that are prone to corrosion. These areas could be determined through an analysis of environmental information, including weather forecasts and historical rainfall data.

Once deployed, the microphones would be used to collect acoustic signals generated during wheel-rail impacts at impact sites such as crossing noses and IRJs, allowing for non-destructive testing and real-time health condition monitoring including the detection and quantification of corrosion at the underside of the rail.

When the corrosion exists at the underside of the rail, the pattern of the acoustic signal will change. Advanced signal processing techniques can be applied to extract structural information from these acoustic signals, and then to distinguish between a healthy rail and a rail affected by corrosion. Additionally, by applying machine learning to evaluate the shapes and types of corrosion, such a system could offer a comprehensive evaluation of the health condition of the rail. These efforts would then be fed back to rail maintainers, suggesting which specific rail sections would require replacement.

1.5 Scope and terminology

The protect proposed in the concept description represents a significant research endeavour. However, based on the concept, a research scope can be defined in order to explore the fundamental principles required for the development of such a system.

The aim of the research presented in this thesis is to investigate the possibility of using acoustic techniques to identify corrosion at the underside of the rail. The work presented has explored this concept in a laboratory environment. It is difficult to address the use of the wheel-rail impact as an impulse source in this environment, and so simulated impacts are used.

Rail faults include those both external and internal to the rail. This research focuses on external faults, i.e. those detectable by observing the surface of the rail. External faults are generated due to four main reasons (manufacturing, improper use, RCF, and corrosion). Based on the investigation by Tan [14], corrosion of the underside of the rail is difficult to detect. Hence, this research focuses on corrosion of the underside of the rail, which is called underside corrosion in this thesis.

The simulated wheel/rail impact signals are called vibration/acoustic impact response signals in this thesis. Considering the signal processing and feature extraction, this research focuses on signal decomposition methods, which is a common approach used for processing impact response signals [15]. A classifier (Convolutional Neural Networks (CNNs)) based on machine learning methods is applied for categorising impact response signals with different features. CNNs present contains advantages when compared to other machine learning methods, especially in tasks involving image and spatial data. Their capacity to autonomously acquire hierarchical features, offer spatial invariance, and employ weight sharing and local connectivity sets them apart. CNNs also excel in real-time processing, transfer learning, and consistently achieve state-of-the-art performance in various computer vision applications. Their versatility, scalability, and generalisation capabilities make them a compelling choice for a wide range of tasks. Considering the research of classifying the fault on the rail based on the one-dimensional signals, a one-dimensional Convolutional Neural Networks (1-D CNNs) is the selected classifier and is applied for classifying one-dimensional signals in this research.

1.6 Hypothesis and objectives

The central hypothesis of the thesis is that acoustic inspection of impact response signals can be employed as an effective and automated method for detecting corrosion on the underside of the rail. To explore this hypothesis, the following questions need to be considered:

1) Which inspection methods are appropriate for inspection of underside corrosion, and what are their limitations?

2) Does an impact response signal contain information that can be used to identify underside corrosion?

3) Given that it is difficult to obtain real-world defect examples, is it possible to use physical simulations to generate data of sufficient validity that can be used in the development of fault detection algorithms?

4) Which signal processing method is most suitable for the selected inspection method?

5) Is it, therefore, possible to classify and quantify underside corrosion using impulse response signals?

The first two questions can be answered from a review of the literature. From a review of current research on railway corrosion, the advantages and limitations of the current inspection method can be understood and then a potential inspection method can be selected. Additionally, a review of signal processing techniques for the selected detection method can determine a suitable signal processing approach to extract the corrosion features.

The third question can be answered from a validated simulation model to create sufficient simulated acoustic impact response signals. The model is validated by comparing the similarity of simulated signals with the signals obtained from laboratory tests.

The fourth question can be answered through a comparison of methods selected from the literature review (Chapter 2). The details of the selected methods and how they can be adapted specifically for use with impact response signals are also described in Chapter 5.

The last question can be explored through the evaluation of laboratory test results obtained using feature extraction and classification methods, which are described in Chapters 5 and 6.

1.7 Research structure

This research first introduces the potential types of rail faults and their corresponding inspection techniques. In order to identify the difference between healthy and faulty rails, the algorithms for extracting the information about rail faults are introduced. A classification method is introduced to categorise the different types and different damage levels of these faults. Due to it being difficult to obtain sufficient data from the real world to support the research, a method using a validated simulation model to generate simulated data is introduced. the details of above introduction are addressed in Chapter 2.

Additionally, to realise the above objectives, the methodologies of simulation, feature extraction, and fault classification approaches are introduced. To simulate a situation that generates an acoustic signal by the impact of a hammer on the rail, a multi-physical Finite Element Model (FEM) is designed. The method for coupling two different simulated physical fields is introduced and then the simulation model validated through a laboratory test (Lab experiment A). A dataset of simulated signals is then generated by the validated model, and a method to extract the features from simulated signals is researched. Moreover, the classification method (machine learning-based method) and its optimisation algorithm are introduced to classify these different features of healthy and faulty rails.

Considering the FEM model, in order to verify the accuracy of the simulation, a further laboratory experiment is completed (Lab experiment B). The experimental results are compared with signals generated from the simulation model. Different geometries of corrosion (different types, different corrosion levels) are designed following the maintenance standard (the maximum area/depth of faults that need to be replaced) [16]. These are used for generating the acoustic signals by the validated FEM model. The acoustic data are then grouped and used as a dataset for further research. The details of building and validating the FEM model is introduced in Chapter 4.

The simulated signals need to be processed to extract the features for the identification. Therefore, a suitable signal processing method is explored for the acoustic impact response signal. Based on the selected suitable method, the algorithm is improved to denoise it and obtain better effects of the faulty features. These works are introduced in Chapter 5.

After extracting useful features from the acoustic impact response signals using the improved algorithm, a classification method is applied to assign different corrosion categories. The simulated data are used for training the classification model and optimisation methods are used to improve the effectiveness of the classification. Finally, the experimental results are used to demonstrate the overall classification system accuracy. The details of the classification system are introduced in Chapter 6.

Furthermore, a summary of the contributions made by, and the novelty of, the work which has been completed in this thesis. The limitation of the this research are given. And future work, based on the results and conclusions obtained in this thesis, is also discussed. The conclusion of the research is introduced in Chapter 7.

The flow chart of this research is presented in Figure 1.3.

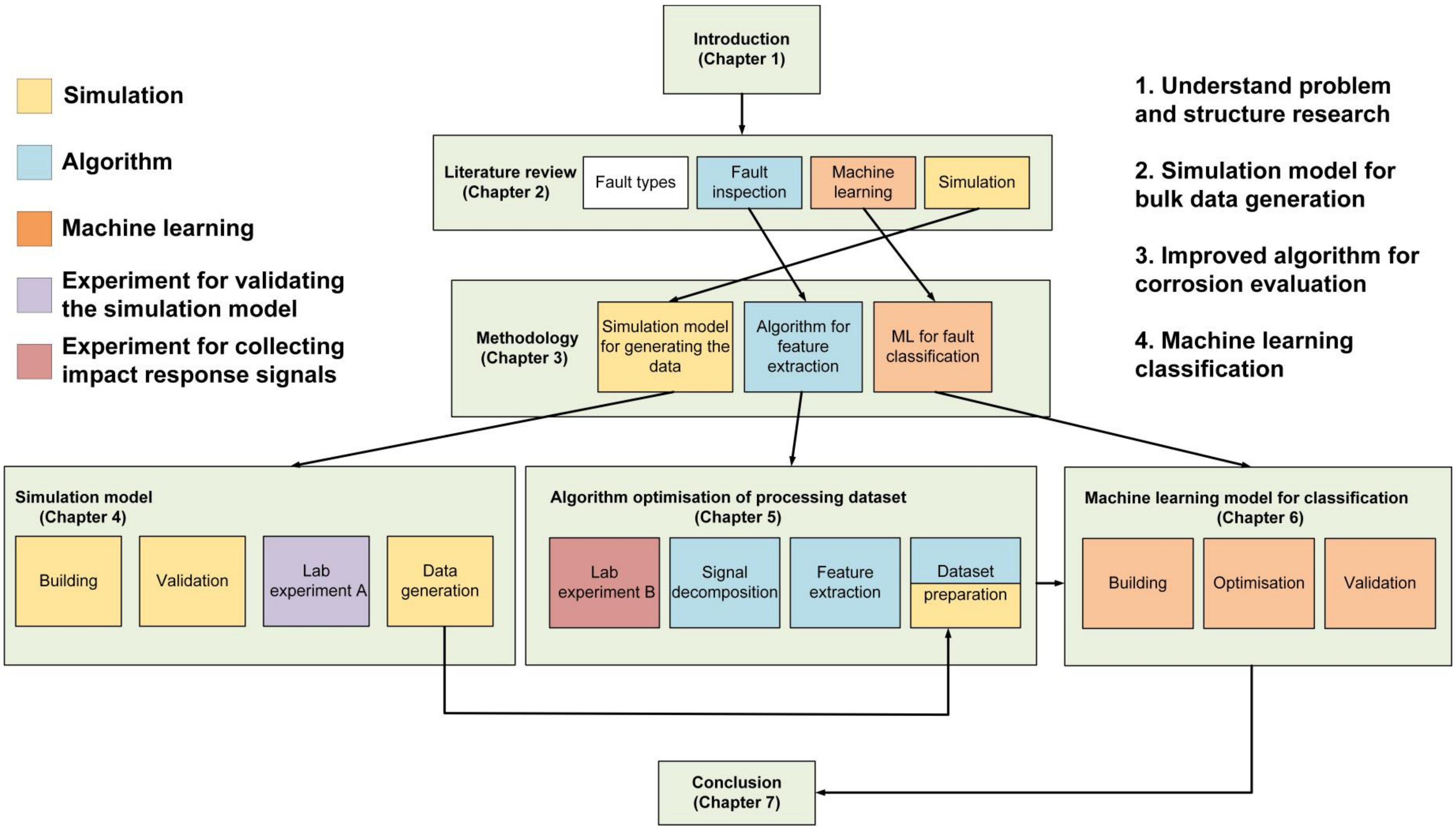


Figure 1.3 Flow chart of the research

CHAPTER 2 LITERATURE REVIEW

Considering the research gap discussed in the introduction, a review of the investigation for an optimal method to solve this gap is addressed in this chapter. Firstly, general types of railway damages are listed. Specifically, the formation of rail damage and the damage categories are introduced first. In terms of railway corrosion, which is the research target of this thesis, the different types of corrosion are introduced. Secondly, different approaches to inspection of the railway are presented. The advantages and disadvantages of each approach for inspecting its corresponding type of railway damage are discussed. And then, from the discussion of railway inspection methods, it is understood that vibration signals and acoustic signals could potentially be applied for the inspection of underside corrosion of the rail. Hence, applications of processing vibration signals and acoustic signals are introduced. Additionally, the research status of processing the impact response signal, the potential signal that will be used for inspection, is addressed. Thirdly, the Finite Element Method (FEM) for simulation in the railway domain is introduced. The modelling techniques for the FEM are presented. Natural frequency is one of the important parameters for the identification of rail conditions. Hence, the current FEM techniques for simulating the natural frequency are addressed. And then, the relevant research on the signals selected for inspection of the railway, vibration and acoustic signals, is introduced. Finally, considering fault classification, the current machine learning-based methods for the railway domain are introduced. Firstly, the common classification methods are introduced, and then the potential classification method applied in this thesis is addressed to illustrate its advantages and the improvements which need to be completed. In order to present the structure of the literature review, a flow chart was drawn as shown in Figure 2.1.

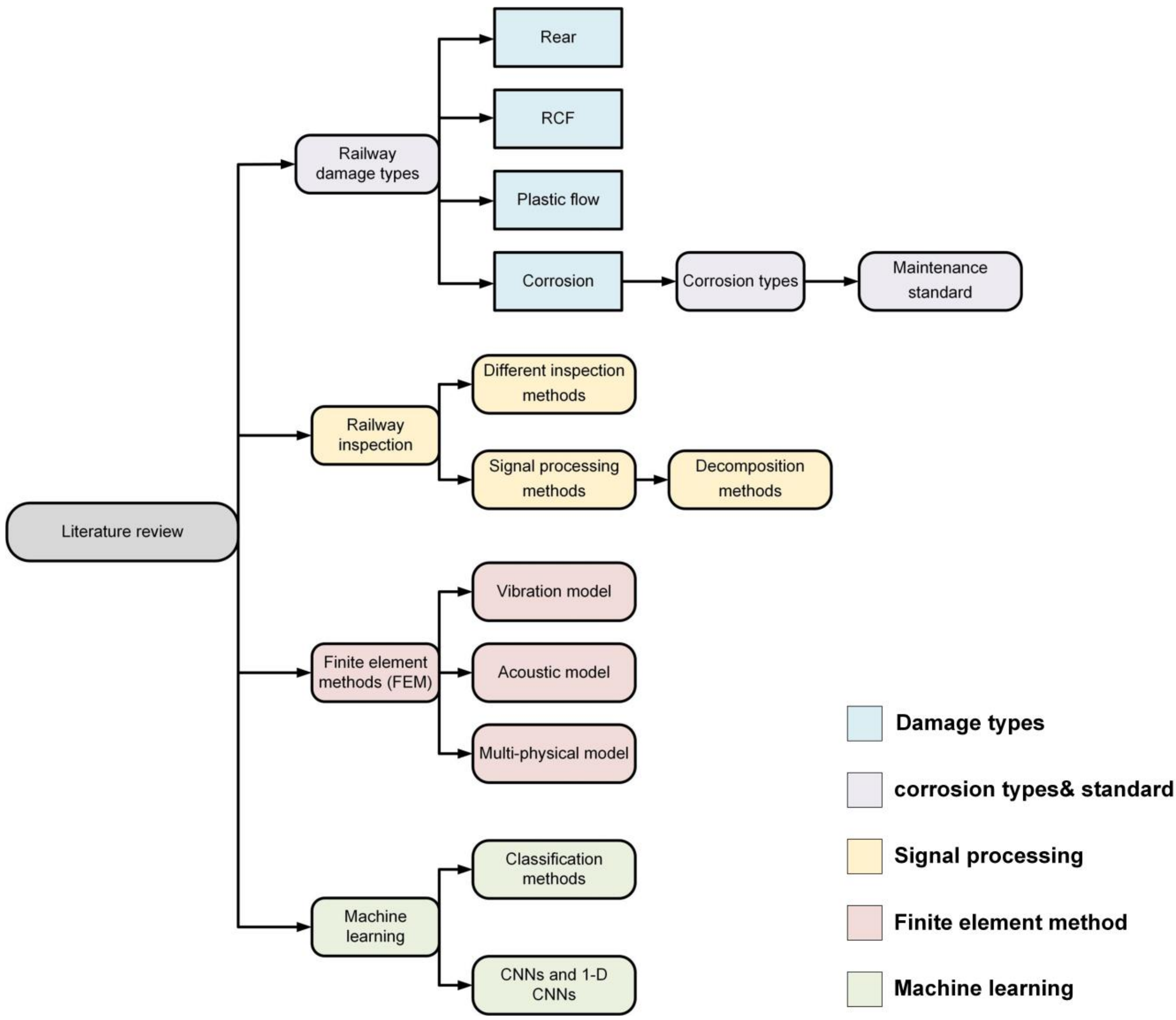


Figure 2.1 Flow chart of the literature review

2.1 Overview of rail damage

This section discusses the main types of rail damage. In order to understand the principles of their formation, it is important to first understand the mechanisms behind it. Studies analysing the classic reasons for fault generation on the rail through force analysis are reviewed, and different types of damage on the railway are classified. Additionally, the most common types of rail damage such as wear, rolling contact fatigue, plastic flow, and corrosion are analysed. Then, different types of rail corrosion are introduced. And in order to understand the fault level of the rail corrosion (what size of fault should be detected and replaced), the standard of maintenance on the rail corrosion is introduced.

2.1.1 Main types of rail damage

Railway tracks are subject to a variety of types of damage that can lead to serious consequences, including accidents and derailments. To prevent such incidents, it is crucial to understand the principle of the formation of damage on the rail. The damage formation mechanism is based on force analysis, and several researchers have provided an overview of rail faults and their consequences. This section discusses the most common types of damage, including wear, Rolling Contact Fatigue (RCF), plastic flow, and corrosion. The section explores each type of damage, its causes, and its consequences, and provides examples and images to help readers understand the issues. Understanding the different types of damage that can occur on railway tracks is essential for developing effective maintenance and monitoring strategies that can prevent accidents and ensure safe and efficient train operation.

To understand the principle of damage formation, the formation mechanism should be understood first. The damage formation mechanism is based on force analysis on the rail. Cannon et al. [10] published an overview of rail faults and their consequences on railway development. It analyses the classic reasons for fault generation on the rail by the loading by force analysis. Under this loading and impacting of vehicles, cracks and deformation are formed. Lesley et al. [17] classified the failure of different components on the railway, pointing out the common types of cracks and deformation on the track. Tian et al. [18] summarised the different detection technologies for different cracks, which enhances the understanding of the types of crack that most need to be detected or monitored on the rail. The classic types of rail damage are listed as follows.

Wear is one of the most common types of damage on the rail. The rail wears due to the passage of wheels over the rail head, eventually leading to material loss. Wear includes headwear (corrugation wear) and side wear, which are shown in Figure 2.2. Head wear occurs at rail head areas along the whole railway line and side wear generally occurs in wheel–rail contact areas.

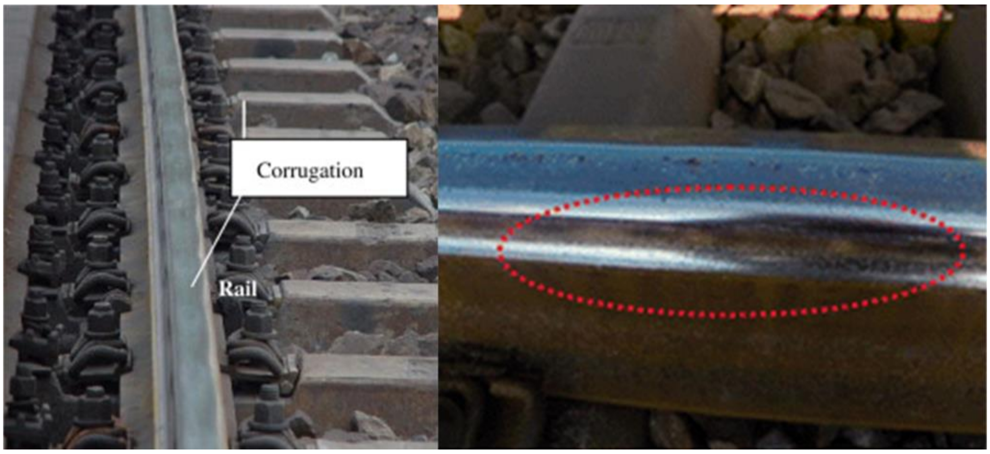


Figure 2.2 Examples of wear faults [19],[20]

RCF is the one of the most common faults in all types of the railway system [21], and is the main reason for the maintenance and replacement of the plain rail line. An example of RCF is shown in Figure 2.3.

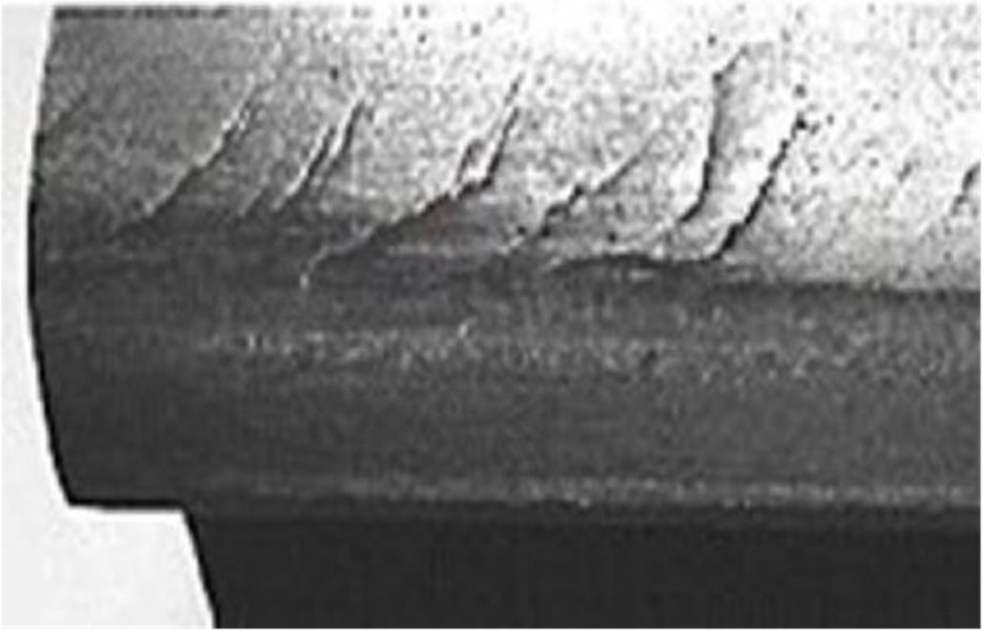


Figure 2.3 Example of RCF fault [22]

Cracks formed by RCF are due to high stress on the surface. RCF defects potentially occur from head checks, gauge corner cracks, and squats. Specifically, RCF cracks are generated in an area where the components occur in rolling contact or sliding contact [23]. The combination of high normal and tangential stresses between the rail and wheel is the main reason leading to RCF cracks. In the beginning, the rail–wheel contact generates a microscopic crack. This crack will propagate along the rail surface to the rail near the surface (around 10°) with a depth of approximately several millimetres. A formed RCF crack will increase the stress around this crack, which leads to further breaks of the rail. An outstanding example of an accident due to RCF is the derailment at Hatfield in 2000, in which 39 people were injured and 4 died. The economic loss exceeded over £1 million. A common RCF crack pattern is shown in Figure 2.4.

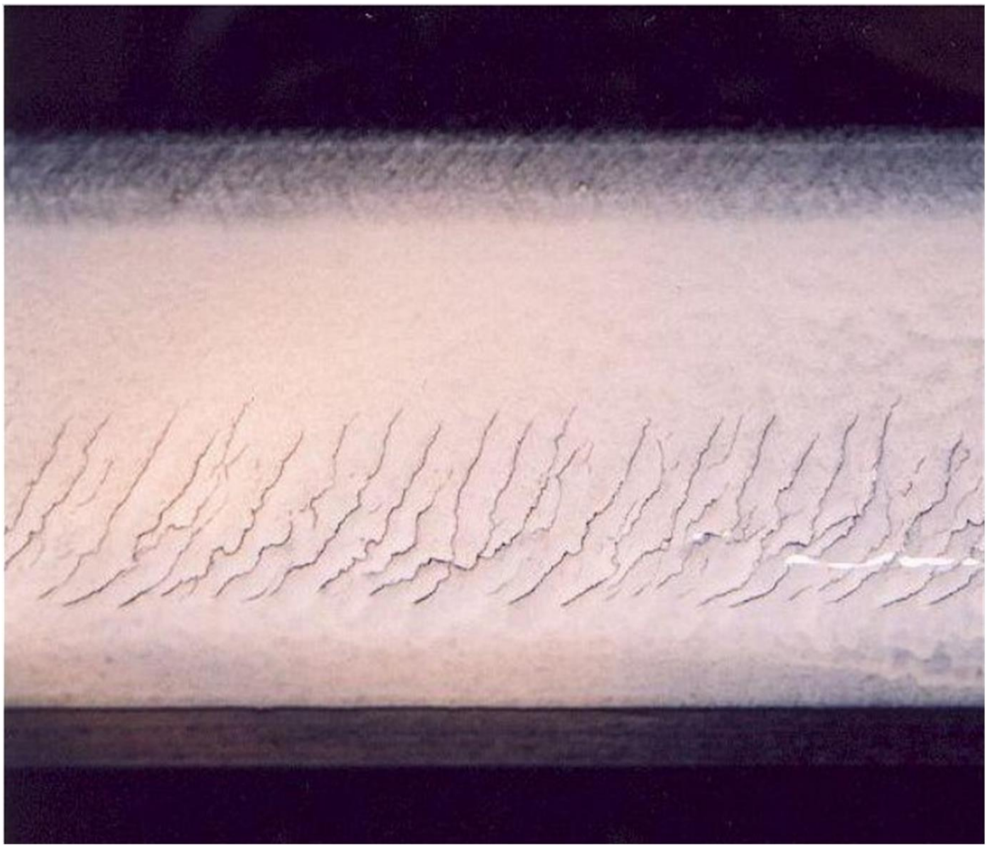


Figure 2.4 Example of RCF cracks [20]

Plastic flow generally exists in the rail–wheel contact area which is very small. With the massive weight of a vehicle concentrating on this area, it can generate large stress on the rail. Steel subjected to this stress can start to plastically deform and flow away from the highly stressed area, which is called plastic flow. The consequence of plastic flow is that lips of the material will exist on the rail surface (shown in Figure 2.5), which will decrease the friction between the wheel and rail, resulting in derailment.

Figure 2.5 Example of plastic flow [24]

Corrosion is the most common fault in rails. A rail is one of the few steel products that is usually used in an open environment without any extra corrosion protection. Hence, corrosion is one of the serious issues in all rail damage questions.

Corrosion can exist in different locations on the rail. The formation of corrosion is generally based on two reasons. Firstly, corrosion is generated due to exposure to the environment and oxidation by acidic species (generally from sunlight, relative humidity, temperature, atmosphere, and salt deposition); the rail material will be damaged and peel off, resulting in fracture of the rail. It is difficult to find and monitor corrosion. An example of environmental corrosion is shown in Figure 2.6.

Figure 2.6 Environmental corrosion on a rail [25]

Another reason is stress corrosion, which has been a popular topic of research in recent years. The main consequence of stress corrosion is cracks, which commonly exist on the rail surface.

Corrosion generally exists near the rail base and rail foot [7]. Corrosion at the rail foot or rail base leads to thinning of the rail foot and instability of the rail (displacement, increasing stress). Figure 2.7 shows different types of rail corrosion. Figure 2.7a illustrates significant corrosion loss at the rail base, primarily attributed to stray current. In Figure 2.7b and Figure 2.7c, we observe corrosion resulting from salt precipitation, likely originating from localised corrosion, which eventually led to substantial metal loss and the shedding of rust debris. Figure 2.7d demonstrates severe corrosion affecting the base and foot, with the upper portion faring relatively better. Meanwhile, Figure 2.7e presents a rail foot heavily corroded with a thick rust layer. Although corrosion failures of rails have not been directly linked to major accidents such as derailments, it is widely acknowledged that engineering failures typically arise from multiple contributing factors, with corrosion playing a significant role as a precursor.

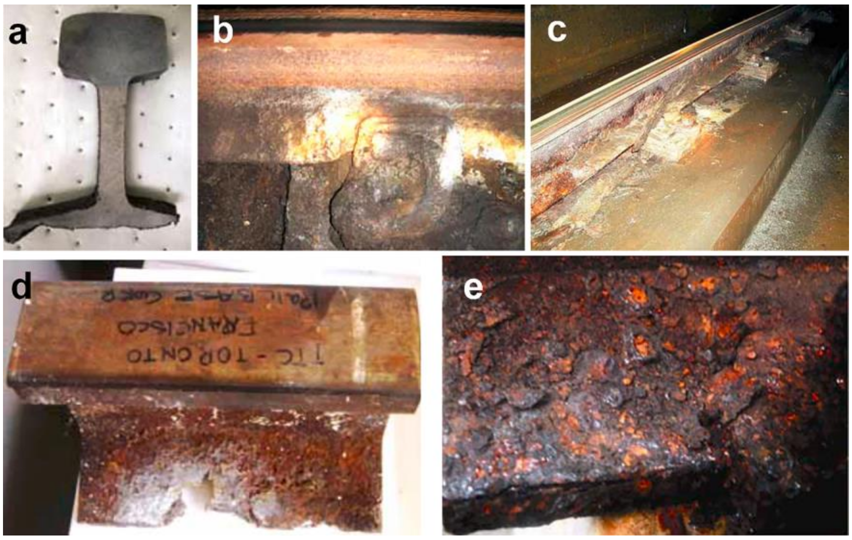


Figure 2.7 Different types of rail corrosion [7]

Corrosion will not lead to significant loss of the rail. However, some types of corrosion exist at locations where it is invisible. Compared with other types of rail damage, rail corrosion is easier to ignore and harder to detect on time, which is the major problem related to corrosion that has not been the subject of enough concern. Hence, it is necessary to attract more attention to the corrosion problem.

2.1.2 Main types of rail corrosion

The types of rail corrosion can be separated into general corrosion and localised corrosion. General corrosion generally exists on the large surface of a rail and initialises due to pitting corrosion. These pits on the rail surface quickly coalesce and form a large area of material loss. However, general corrosion won’t lead to crucial structural damage. General maintenance and cleaning of the rail surface can easily protect the rail’s health.

Localised corrosion exists from the rail surface to the inside of the rail. An example of localised corrosion is crevice corrosion [26], which is shown in Figure 2.8.

Figure 2.8 Crevice corrosion [26]

Crevice corrosion usually shows at the rail foot or the rail base. Serious crevice corrosion will decrease the depth of the rail foot.

Stress corrosion is another important type of corrosion on the rail. Stress corrosion is generated due to the expansion of localised corrosion. Stress corrosion can significantly influence the steel’s strain due to repeat loading [27]. A crack will be formed by the interaction between the strain and the stress corrosion. An example of stress corrosion is shown in Figure 2.9 [28].

Figure 2.9 Stress corrosion on a rail [28]

Besides the above two localised types of mechanical corrosion, electrical corrosion is another type that needs to be considered. Electrical corrosion can have a significant impact on the railways [29]. When electrical current flows through a track circuit, it can cause corrosion on the rail. This corrosion can lead to a loss of track integrity and increase the risk of derailment. In addition, electrical corrosion can also cause damage to signalling equipment and affect the overall safety of the railway system. Therefore, it is important to regularly monitor and address any electrical corrosion on the railways.

2.1.3 The standard of maintenance on rail corrosion

The standard of maintenance is a very important factor in understanding what size of fault should be detected and replaced, and is also helpful for maintenance of the railway. In this section, the standard of maintenance on rail corrosion is introduced. Based on the standard, we can understand the size of the fault that will be created for the simulation and experiment which are used in the following chapters.

The standard of railway maintenance in Great Britain is based on the government document *Permanent Way Design and Maintenance Standards*. This standard includes the structural size of all railway components. The requirement for maintenance or replacement can also be searched for in Issue 5 of this standard [30]. Based on this standard, the size of the rail profile can be understood, which is shown in Figure 2.10 (the rail brand is 56E1).

However, in terms of the maintenance of rail base corrosion, this standard does not mention the requirement for replacement. Considering the standard of corrosion maintenance and replacement, the Transportation Technology Centre, Inc. (TTCI), which is funded by the Transit Cooperative Research Program (TCRP) in America, studied and issued a standard for corrosion on the rail base. From the research at TTCI, the metal loss due to corrosion in the vertical direction should not be more than 1/4 depth of the rail bottom. If the corrosion is deeper, the rail should be replaced [14]. For example, in terms of the 56E1 rail, the depth of corrosion at rail bottom should not be over 2.8mm.

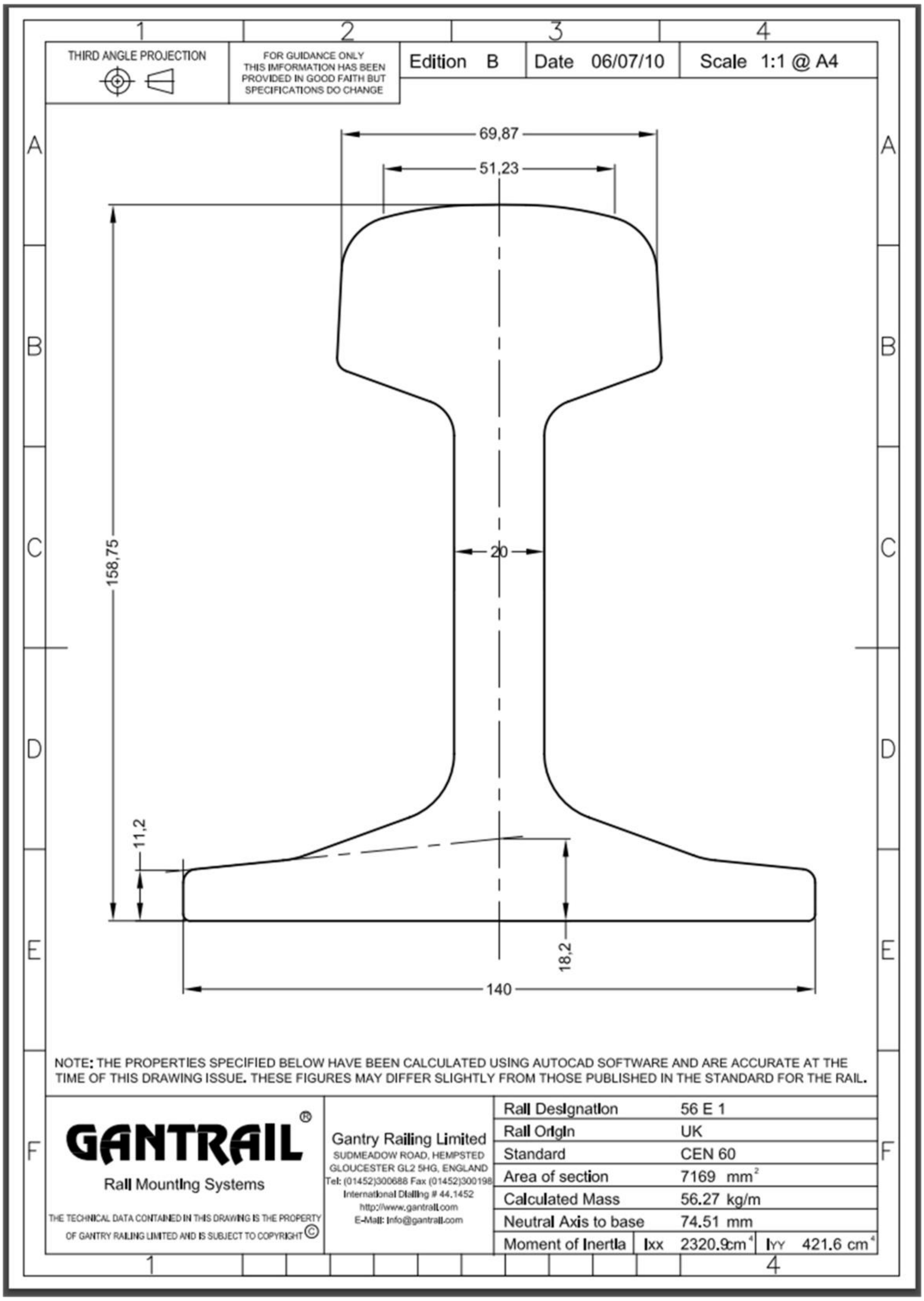


Figure 2.10 Rail profile (56E1) [31].

In conclusion, based on the maintenance standard as shown in Figure 2.10, the corrosion at the bottom of the rail base is too thick (11.2 mm) and is invisible to detection by current inspection methods. Additionally, the current maintenance method is to replace the whole track experimentally. The real condition of the corrosion at the bottom of the rail base is unknown before replacing the track. Hence, we can ask whether there are suitable methods that can identify the corrosion condition on the bottom of the rail. To explore the answer to this hypothesis, it is necessary to understand the current techniques for railway inspection.

2.2 Overview of railway inspection

Railway inspection is one of the methods of protecting the railway's normal operation. The various inspection technologies have been significantly improved in this century. In this section, firstly, the common railway inspection methods are introduced. The advantages and disadvantages of each inspection method are addressed. From reviewing these inspection methods, the shortcomings of each inspection related to rail corrosion can be understood and a potential method for the inspection of underside corrosion can be selected. Secondly, signal processing techniques are generally used for extracting the feature from the signal. Hence, the research status of different signal processing methods in railway inspection is introduced. Finally, as introduced in Chapter 1, the impact response signal potentially can be used for identifying corrosion at the rail base; the relevant research focusing on processing this kind of signal is addressed. This section addresses the potential techniques and methods for inspecting underside corrosion of the rail.

2.2.1 Review of approaches to railway inspection

2.2.1.1 Introduction of inspection techniques for track damage

Nowadays, in terms of the different types of faults which are mentioned in the above sections, the current methods can effectively complete the condition monitoring and fault inspection. Each technique can focus on the specific damage outside or inside of the rail. Regarding the different principles of physics, inspection methods on the railway can be separated into magnetic, electric, acoustic, vibration, and thermographic methods. This section introduces these techniques and their corresponding detected targets.

1) Ultrasonic

Ultrasonic inspection is a technique commonly used for inspection of the rail. In traditional ultrasonic inspection, the probe is settled on the rail surface. Due to the gap existing between the probe and the rail surface, an ultrasonic couplant is used to help the wave propagation between the probe and the detected material. Nowadays, guided ultrasonic wave testing is a recognised method used for railway inspection, which is shown Figure 2.11.

Figure 2.11 Guided ultrasonic wave [32]

The interaction between the guided ultrasonic wave and rail damage can change the physical conditions such as wave transmission, wave reflection, mode conversion, energy dissipation and subtle wave features on the rail [33]. By analysing these changes, the damage parameters (location, size, damage level) of the rail structure can be evaluated. But the guided ultrasonic wave only has significant advantages for surface or near-surface faults such as RCF cracks [34], vertical split cracks [35], and gauge side flaking [36]. For example, Wu et al. [37] used guided ultrasonic waves to detect the railhead surface in the switching area. From the proposed signal processing and feature recognition, faults in this area can be efficiently detected with good accuracy. Mariani et al. [38] proposed an advanced guided ultrasonic wave method to detect cracks near the surface of the rail head. The frequency analysis method is considered, and the features of the cracks can be obviously diagnosed. However, currently, the guided ultrasonic wave method is not sensitive enough to be effective for detecting rail damage deep inside or at the bottom of the rail. The accuracy of the rail detection is too low for application in the field.

2) Acoustic emission

Acoustic Emission (AE) is a phenomenon in which an elastic wave is generated by releasing energy in early material deformation. One of the earliest studies on AE was conducted by E. O. Hall in 1953 [39], who investigated the AE signals generated during tensile testing of metals. Since then, numerous studies have been conducted to develop and improve AE techniques for different applications such as high-pressure vessels, bridges, rail sleepers, etc [40]. In terms of railway inspection, the AE technique is a method that can monitor variance of the rail structure, especially crack deformation on the rail, and has been widely used in recent years. For example, Murav'ev and Baiteryakov [41] considered the rail on a bridge to analyse the stress–strain conditions when a vehicle passes through the bridge. Considering track inspection with AE technology, Zhang et al. [42] investigated the feasibility of using AE for detecting and diagnosing rail defects such as cracks and wear. The authors found that AE is effective in detecting the onset and growth of cracks and can provide valuable information for the maintenance and inspection of railway systems. Generally, the AE technique has the potential to realise real-time monitoring on the railway. However, it can only detect a crack when it has just been generated. It is difficult used for rail damage that has been already formed. Additionally, there is a low risk of corrosion at the beginning of damage formation. Hence, AE technology might not be suitable for this kind of damage.

3) Visual technology

Visual inspection is a common method of rail inspection. It is a fast, accurate, and economical method integrating techniques of data acquisition, monitoring, and image processing. The devices (camera, light source, auxiliary sensor) used for visual inspection are always settled on a vehicle. In the process of visual inspection during the vehicle's operation, a light source illuminates the rail components and a camera captures the image from the track. Using advanced image processing and feature extraction techniques, the health condition of the rail can be monitored. The process of visual inspection is shown in Figure 2.12.

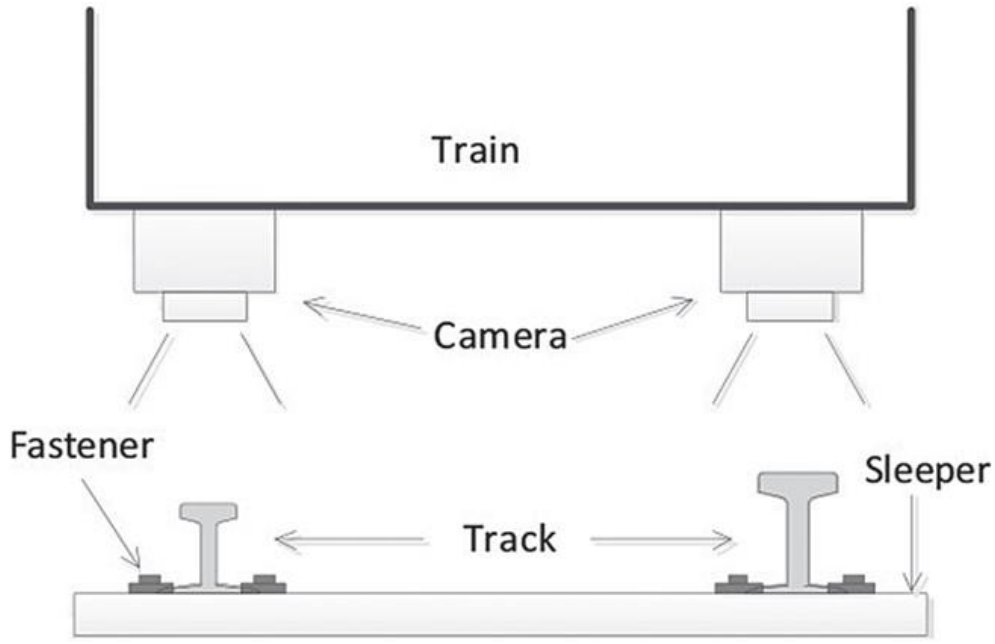


Figure 2.12 The process of visual inspection [43]

Visual inspection focuses on the texture or cracks on the rail surface [43]. For example, Deutschl et al. [44] classified texture defects on the rail surface based on a spectral image-difference algorithm. Mandriota et al. [45],[46] evaluated defects on the rail surface based on classifying the images captured by a laser camera, using a Gabor filtering method for denoising and a K-nearest classifier for image classification. Ye et al. [47] proposed a novel 3-D perceptual system to detect surface defects in railway tracks based on a 2-D laser sensor. Compared with the traditional 2-D inspection method, the proposed method can detect a wider surface. Defects on the rail surface can be considered more completely.

The current visual inspection method is widely used for detecting surface defects on the railway with decent effectiveness. However, a limitation of visual inspection is that it can only evaluate surface defects. There is limited research currently considering the inspection of fault inside of the rail or at the bottom of the rail.

4) Electromagnetic methods

Electromagnetic methods have become increasingly popular in railway inspection due to their ability to detect faults without physical contact with the rail [48]. These methods apply electromagnetic waves to detect variations in the rail's electrical and magnetic properties, which can be indicative of potential faults.

One advantage of electromagnetic methods is their ability to detect faults in real time, allowing for prompt maintenance and reducing the risk of accidents [13]. Additionally, electromagnetic methods can be used to inspect rails without requiring the rail to be taken out of service, minimising downtime and the costs associated with maintenance.

However, there are also several disadvantages to electromagnetic methods: they are highly dependent on the physical properties of the rail, and any variations in these properties can affect the accuracy of the inspection [49]. Additionally, electromagnetic methods require a skilled operator to interpret the inspection data accurately. Finally, they may not be able to detect all types of faults, as some faults may not cause significant changes in the rail's electrical or magnetic properties.

Several specific electromagnetic techniques used in railway inspection have been identified. These include Magnetic Flux Leakage (MFL), Eddy Current Testing (ECT), and Alternating Current Field Measurement (ACFM). Each of these techniques has its advantages and disadvantages, depending on the specific application. But all electromagnetic methods have a disadvantage, the skin effect, a phenomenon that occurs in conductors carrying Alternating Current (AC). It refers to the tendency of the current to flow primarily near the surface of the conductor, rather than evenly through its entire cross-section. This occurs because the magnetic field created by the current induces an opposing current in the conductor, which counteracts the flow of current near the centre of the conductor. This phenomenon will lead to the electromagnetic method only being able to detect damage on or near the surface. Underside corrosion is too deep to be detected by electromagnetic methods.

5) Vibration

Vibration is a common phenomenon existing in the world. In terms of the application of vibration techniques to railway inspection, a number of works have been completed. A vibration signal can be defined as a periodical signal (from the rotational components such as bogies [50], gearboxes [51], and bearings [52]) or a non-periodical signal (impact response signal [53], generated by rail roughness [54]). Considering rail inspection by vibration signals, the hammer impact is a common method used for generating the vibration signal [55]. An example of an hammer impact is shown in Figure 2.13

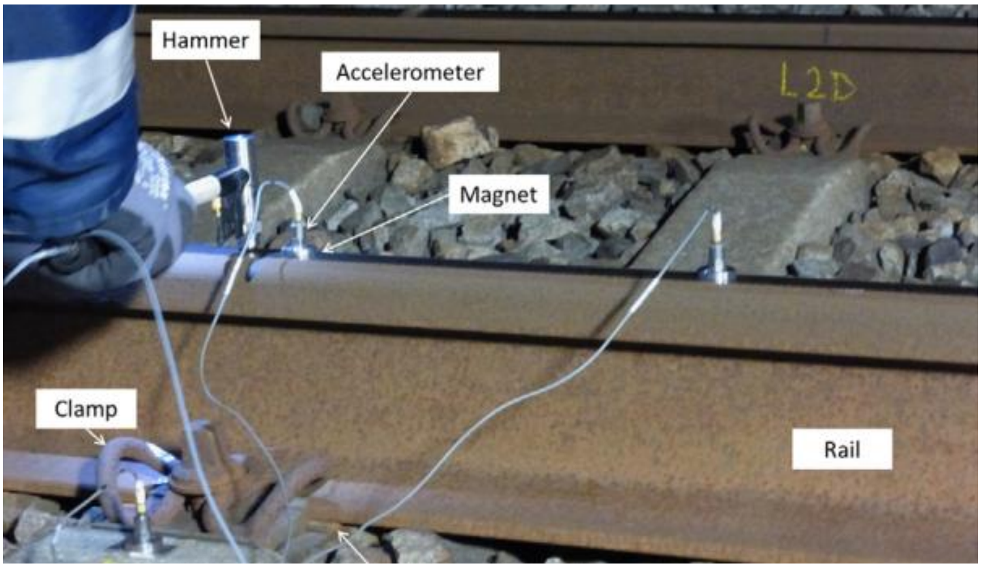


Figure 2.13 Vibration test [56]

To analyse the vibration signal, at present, Wavelet Transform (WT) analysis and the signal decomposition method are the two most popular methods of combined time–frequency analysis widely used. Jena et al. [57] applied a method combining wavelet analysis and empirical mode decomposition for fault identification, which has obvious effects on gear damage. But due to the limited sample detected, it is hard not only to localise the defects but also to detect different types of damage. Espinosa et al.[58] and Yan and Gao [59] used a Hilbert–Huang transform to diagnose stator faults and mechanical faults. The above analysis of the vibration signal is based on rotational machines, which can create the vibration signal by themselves. However, the rail is a different component for which the vibration source needs to be generated through another device (hammer, wheel–rail contact, etc.). The hammer impact test is the most common method for rail inspection through the vibration signal. The vibration signal is generated through the impulse given from the hammer to the rail. This signal is called the impact response signal [60]. Currently, analysis of the impact response signal is done by the Frequency Response Function (FRF). For example, Oregui et al. [56],[61] studied the feasibility of using an FRF to identify the characteristic frequencies of track damage in vibration data obtained using the hammer impact method. Kaewunruen et al. also considered the inspection of different components of the track (such as IRJs, which are track discontinuities) by using an impact hammer and vibration measurements [62],[63],[64].

However, regarding the application of the vibration signal to rail inspection, there are some problems that need to be solved. Firstly, the vibration signal is inevitably influenced by various noises. In terms of the impact response signal, the main reason for the noise is the different impact positions and sensor positions. Additionally, the track is put on a rail base. Different bases are made of different materials (concrete, wood, etc.). Hence, the vibration impact response signal generated from the hammer on the different bases will be different.

6) Acoustic listening method

Acoustic listening technology was widely used in the early automobile manufacturing industry. For example, in 1983, Sun detected cracks in the front axle and I-beam of automobiles by hammer impact [11]. In 1984, the Ford Motor Company in the United States applied acoustic listening technology on the crankshaft and achieved good results [65]. In applying this method to railway fault inspection, the principle is listening to the sounds generated from the rail–wheel impact or the vibration of a component on the vehicle. The history of the acoustic method in railway inspection can be traced to the Industrial Revolution. Due to the limited acoustic collection devices in that period, an engineer would detect a fault by listening to the sound when they impacted the railway with a metal hammer.

With the development of non-destructive technology, the acoustic method became a popular automatic technique which has been widely used in railway inspection. At the 13th World

Non-Destructive Testing Conference held in Sao Paulo in 1992, Raj [66] pointed out that acoustic defect testing is not only an assistant tool for visual technology but also a basic non-destructive testing technology. Since the start of the 21st century, with the development of computer technology, the application of acoustic listening technology has been further researched [67],[68],[69],[70].

Although acoustic listening inspection has been used in railway inspection, it still has some limitations which are waiting for enhancement. Most of the current research based on acoustic inspection is in the laboratory stage. One of the reasons for this is that there are varying types of noise generated during railway operation (wheel rolling, vehicle components, Doppler effect, etc.). The useful information from the acoustic signal will be masked by these noises. In recent years, a popular research topic is how to extract these useful patterns of the acoustic signal using suitable signal processing techniques. The current practical acoustic inspection on the railway mostly is used for the rotational machines on a vehicle, such as bearings [71], gearboxes [72], bogies [73], etc. There is limited research on the use of acoustic methods for track inspection.

The detection characteristics of the above technologies are summarised in Table 2.1.

Table 2.1 Detection characteristics of different rail inspection technologies

Method	Basic inspection theory	Available detection objects	Characteristics
Guided wave	Propagation of the guided wave is reflected back and forth along a parallel boundary, and the guided wave energy is scattered or absorbed because of any damage existing.	Deep cracks in the head rail; Surface & near-surface cracks at web and rail foot	Fast speed; Structural reachability; Hard to measure quantitively
AE	The signal is collected from an elastic wave which is generated by releasing energy in the early solid deformation.	Rail breaks; Wheel burns; Squats; Wet spots; Worn rail profiles	Poor for detecting internal cracks; Can only detect in predefined areas; Complex signal processing technology

Visual inspection	Refers to the inspection of the workpiece under test by the eyes or with the help of visual aids such as a charge-coupled device camera.	Surface damage including cracks, squats, corrosions, and deformation	High detection speed; High accuracy, good practicability; Non-contact; Cannot detect internal cracks
Electromagnetic methods	Analysis of the different physical phenomena which come from a change in the electromagnetic field due to the structure of objects.	Surface and near-surface cracks	High accuracy; Mature technology; Hard to detect deeper failures

Vibration	The rail surface is impacted with a small hammer or a specialised instrument, and then the accelerated signal is collected from a specific sensor.	Internal and surface damage to the track	Easy to get the signal; Multi-position detection from the same signal; Complex signal processing technology
Acoustic listening	Impact the rail surface with a small hammer or a specialised instrument, and then listen with a microphone or sensor for the sound waves propagating through the air from the rail.	Surface and subsurface cracks	Non-contact; Easy to get the signal; Realise multi-position detection from the same signal; Complex signal processing technology

In view of the problems identified in the above literature review, it can be summarised that current rail inspection techniques achieve good surface and near-surface crack detection performance. However, for inspection of the web rail or footrail as well as the rail bottom, they have limited ability to detect failures on the surface or inside the rail. In addition, current inspection technologies with a single function can only detect parts of the rail, which means it is impossible to complete a full test for a whole track at one time.

2.2.1.2 Inspection techniques for track corrosion

The inspection of corrosion is a popular topic, which has been widely used on metal machines such as pipelines [74], oil storage tanks [75], aircraft [76], etc. Considering the inspection of corrosion on inside surfaces (a similar situation to inspection of the corrosion at the bottom of a rail base), ultrasonic and electromagnetic inspection are two common methods that have been applied for testing. The ultrasonic method is better for online inspection but the noise is a serious problem that covers the real corrosion information. The electromagnetic method is effective for evaluating the corrosion level. The thickness of the pipeline (or tank) is able to overcome the skin effect. However, whatever the pipeline or tank, the thickness of these two components is thinner than the bottom of the rail base, which means the above techniques might not be suitable for inspection of the corrosion at the bottom of a rail base.

Rail corrosion, as one of the common faults occurring in plain-line railway tracks, has attracted notable attention in recent years [77]. Hernández et al. [25] evaluated the corrosion risk by the chemical simulation of corrosion formation and propagation at the rail base. 3-D visual inspection and visualisation are commonly used in corrosion inspection processes

[47],[78],[79],[80]. However, most inspection focuses on the accessible surfaces of the rail. Considering corrosion of the underside of the rail, there are limitations in the current inspection techniques. For example, it is difficult to detect corrosion through 3-D visual inspection where a mapping laser and optical scanner cannot be directed towards the fault. The ultrasonic method is not sensitive to small faults (RCF smaller than 4 mm) and can also miss the faults at the rail foot, especially corrosion defects. Electromagnetic methods can be limited in extracting information from the underside of the rail where their signals are unable to pass through the whole rail body due to the skin effect [81] and other attenuation.

In summary, the current techniques have limitations for the inspection of the underside of the rail. The most important reason for this is the location of the fault as it is below the rail and not accessible. Of the current methods, vibration/acoustic methods analysing the signal generated by an impact (hammer or wheel–rail impact) are used for structural inspection with significant effectiveness [82],[83]. They have also been used in railway inspection at the research stage [16],[55],[84]. However, the signal generated from the impact contains environmental noise which can mask the health information existing in the signal.

The environmental noise generated from the impact can have various sources, and the specific noise levels may vary depending on factors like train speed, track conditions, and the surrounding environment. The environmental noise generated from the impact includes [85], [86], [87]:

Train Wheels on Tracks: The rolling contact between train wheels and the tracks generates noise, often referred to as wheel-rail noise. This noise can be particularly significant at curves and switches.

Train Engines: Diesel locomotives and electric trains produce engine noise, including the sound of the engines themselves and exhaust noise.

Train Horns and Whistles: Trains use horns and whistles for safety at road crossings and in specific situations. These can be loud and disruptive, especially in residential areas.

Brake Noise: The application of brakes, especially air brakes, can create screeching or squealing noises, which can be particularly noticeable as a train slows down or comes to a halt.

Infrastructure Noise: Noise can also originate from railway infrastructure like switches, signals, and overhead wires.

Noise from Station Activities: Noise generated at railway stations, such as announcements, platform activities, and the arrival/departure of trains, can contribute to environmental noise.

Maintenance Activities: Maintenance work on tracks, trains, and railway equipment can result in noise disturbances, especially during nighttime maintenance operations.

Ventilation Systems: Modern tunnels and underground railways may have ventilation systems that generate noise.

In this research, considering the various environmental noise from the impact with different frequencies, it is difficult to added different types of noise in one signal. Hence, the Gaussian white noise which could randomly simulated different noises was added for the analysis. And then, there is a need to understand current signal processing techniques in order to select a suitable method for the vibration/acoustic impact response signal.

2.2.2 Application of signal processing to track inspection

2.2.2.1 Applications of signal processing for track corrosion

As mentioned in Section 2.2.1.1, a one-dimensional signal (guide wave, AE, vibration, or acoustic) is commonly used for track inspection. The related application of signal processing to track inspection has also been widely explored for denoising redundant information and extracting the faulty features from the original signal. The principle of a signal processing method is generally based on analysis of the time domain (mean value, standard deviation, range, kurtosis, root mean squares, crest factor, skewness, etc.) [88],[89], the frequency domain (spectrogram, power spectral density, frequencies, energy ratio, entropy, damping ratio, Hilbert–Huang transform, etc.) [90], and the spatial domain, and time–frequency joint analysis (short-time Fourier transform, time–frequency distribution, Cohen's class, etc.) [91],[92].

Considering the technique of signal processing for track inspection, frequency analysis is mostly used. The Hilbert–Huang Transform (HHT) [93] is one of the techniques which has been widely used in processing signals (generally vibration/acoustic signals) for railway fault inspection. Based on the progress of the Hilbert transform, Huang proposed a method using the Hilbert transform as an adaptive signal processing method to extract the features (especially time–frequency energy) from stationary, non-stationary, or transient signals. The first step of the HHT (the core of the HHT) is to decompose the time series signal into a collection of band-limited quasi-stationary functions, which are called Intrinsic Mode Functions (IMFs). The second step is to analyse these IMFs by Hilbert spectral analysis.

The method for decomposing time series signals is called signal decomposition, which is widely and effectively used for signal processing. For example, the HHT and WT are generally used to analyse the IMFs obtained from the decomposition method.

Signal decomposition techniques are widely used in processing many different types of signals and are particularly common in applications such as rotational machine monitoring (roller bearings) [94],[95] and medical examination (pulse, electroencephalogram, electrocardiogram, etc.) [96],[97],[98]. Compared to traditional Fast Fourier Transform (FFT) and WT methods, mode decomposition can identify and categorise minor faults more accurately [99]. Common mode decomposition techniques include Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), Local Mean Decomposition (LMD), and Variational Mode Decomposition (VMD).

EMD was first introduced by Huang et al. [93] and has been used to consider many rotating machine types [100],[101]. It is a recursive method for analysing non-linear and non-stationary signals that adaptively decomposes any signals into their oscillatory components which are called IMFs. EMD identifies characteristic fluctuations in the time series representation of data. It represents signals as extensions of basic functions that are derived directly from the signal itself without predefined mother wavelets [102]. However, the results of EMD rely on accurately identifying extremities in the signal which is susceptible to noise. Distortion of the IMFs may also arise through errors introduced during interpolation and Hilbert Transformation (HT) [103].

To overcome these limitations of EMD, advanced versions such as EEMD and LMD have been developed [104],[105]. EEMD is an advanced signal processing method which considers the influence of background noise and aims to reduce it by applying a population averaging process to multiple tests [106]. The technique is classified as a Noise Assisted Data Analysis (NADA) method as it also incorporates noise data which are then included in average-based processing to develop the final result. EEMD has been applied in a number of domains. Wang et al. [107] and Amirat et al. [108] combined it with an HT technique to identify early-stage faults in the bearings of rotating machines. The feasibility of using EEMD on impact response signals has also been demonstrated by Sun et al. [109]. However, they did not conduct comparative research with other decomposition methods and so were unable to prove that EEMD is the best decomposition method for impact response signals.

LMD is a technique which identifies the envelope and frequency modulation of a signal. It decomposes signals into representations of the amplitude, the frequency, and a product function [110]. LMD is widely used in vibration signal analysis of rotating machines. Xu et al. [111] and Wang et al. [112] considered its use for fault diagnosis of roller bearings and gearboxes. In comparison to EMD and EEMD, however, LMD performs poorly when a signal contains a significant level of noise [113].

VMD is a self-adaptive signal decomposition method proposed in 2014 [114], which has been widely adopted in inspection applications, particularly for rotational machines. Wang et al. [115] proposed an improved VMD method, combining it with a Deep Convolutional Neural Network (DCNN) for detecting rolling bearing element faults, which enhances the visibility of the fault feature and therefore improves the accuracy of diagnosis. Yi et al. [116] considered the feasibility of extracting fault features from turbine bearings through a combination of VMD and Particle Swarm Optimisation (PSO). This application has the advantage of being able to decompose more complicated signals compared to conventional EMD. An et al. [117] proposed an application combining VMD and permutation entropy to identify different bearing fault types with high classification accuracy. Miao et al. [118] considered the application of the inspection encoder signals from gearboxes. The signals were decomposed using an advanced VMD method combined with a sparse representation to extract fault information. Miao et al.'s proposed method both suppresses noise and enhances fault impulses. Considering these works, VMD is often applied to signals emanating from rotational machines. However, the features of impact response signals are somewhat different. Considering the feasibility of decomposing impact response signals using VMD, Yang et al. [16] demonstrated that VMD is able to decompose impact response signals from railway tracks with high classification accuracy.

2.2.2.2 Comparison of different signal processing methods

The advantages and disadvantages of decomposition methods introduced in the previous section vary depending on the specific signal being analysed, and it is not appropriate to draw conclusions without focusing on the same category of signal. Therefore, in this section, a review comparing different decomposition methods is demonstrated based on various types of signals.

Wang et al. [113] studied the effectiveness of EMD and EEMD of seismic signals, using the time–frequency spectrum as an evaluation approach. The study showed that EEMD is able to avoid mode mixing and has better performance in real geology than does EMD. Shrivastava and Singh [119] compared the effectiveness of EMD and EEMD of chatter signals and found that EEMD has advantages based on spectrogram results. Chen et al.

[120] analysed the effectiveness of EMD and LMD of chatter signals and concluded that EMD has an advantage due to a smaller average error, while LMD has shorter irritation times for decomposition and a weaker end effect. Thameur et al. [121] compared LMD and EMD for early detection of gear defects and evaluated the effectiveness of early fault detection using kurtosis as an index on IMFs decomposed by different approaches. The study showed that the characteristics were more obvious when using LMD. Wardana [122] conducted a comparative study between EMD and VMD to detect oscillation in control loops and analysed the significance, regularity, and sparseness of characters decomposed by VMD and EMD. The study showed that the characters decomposed by VMD were more significant than those decomposed by EMD in terms of detection effectiveness. Hadiyoso et al. [123] evaluated the effectiveness of EMD, EEMD, and VMD for the extraction of respiration waves from plethysmography waves and found that EEMD has the advantage of a smaller average error when comparing the IMF decomposed by these approaches with a real respiration wave.

2.2.3 Techniques for processing the impact response signal

Impact response signals are different from rotational signals: impact response signals are impulse signals with gradually decreasing energy, while rotational signals have a periodic peak in the time domain. Hence, kurtosis-based methods are generally not considered suitable for processing impact response signals because of their dependence on the distribution of peaks within a signal. Additionally, as they are non-periodic signals, it is generally difficult to calculate the centre frequency of impact response signals. There is limited research focusing on the use of impact response signals to detect track faults. But the property of the impact response signal is an impulse signal. Hence, it might be possible to find some progress from other research domains that use a similar impulse signal for inspection. An electrocardiogram (ECG) is a physiological impulse signal showing the activity of the heart. It is widely used for the diagnosis of cardiac abnormalities. In view of the signal processing methods used for ECG signals, signal decomposition is one of the common techniques used to analyse ECG signals [124]. For example, Zhang and Wei [125] proposed an advanced EMD technique to denoise ECG signals, which has effective adaptability and wide utilisation. VMD is another technique that has been used for processing ECG signals. Singh and Pradhan [126] applied VMD to an ECG signal. Due to the principle of VMD being based on the central frequency, the purpose of this approach is to decompose the ECG signal into narrow-band IMFs with different frequency ranges. By applying discrete WT to analyse the target IMF which contains the ideal frequency range, the health condition of the heart can be diagnosed. Compared with the existing ECG denoising techniques, the proposed method offers a significant improvement. Besides the ECG signal, the collision signal is another impulse signal which is similar to the impact response signal. For example, Sun et al. [109] analysed the acoustic signal generated by corn kernels impacting a plate. EEMD was used for decomposing this acoustic signal. Damaged kernels could be diagnosed through frequency analysis.

In conclusion, the review in this section indicates that the application of the signal decomposition is potentially beneficial for processing vibration/acoustic impact response signals.

2.3 Finite element method of simulating railway conditions

The simulation model as a numerical technique has been widely used to solve problems in the domains of engineering and mathematical physics. The FEM is one of the simulation methods used to calculate differential equations based on different types of discretisation to obtain approximate solutions. These differential equations can perform in plenty of physical fields. In an FEM model, the interest field is separated into different types of elements. Given sets of boundaries and initial conditions, the FEM model can approximate the differential equations and provide a numerical solution.

In terms of railway simulation, the FEM has been widely used for railway component detection [127]. A common example is the simulation of the condition of rail–wheel contact.

In this project, an FEM model will be used to simulate the condition of generating an validated impact response signal through the hammer impact. The following sections review the four parts of current research based on the simulation of the natural frequency, the simulation of creating vibration impact response signals, the simulation of creating signals in the acoustic field, and validation of the FEM model.

2.3.1 FEM techniques for simulating the natural frequency of structures

Before simulating the physical conditions by using the FEM model, the physical properties of the simulated target should be understood. The natural frequency, which is also named eigenfrequency, represents the frequency of a system/object tending to oscillate without any driving and damping force [128]. The natural frequency of a system is dependent on its own properties and any external influence is irrelevant. In order to obtain the natural frequency of a target, the hammer impact test is a common method that has been widely used [129]. For example, Shen et al. [130] proposed a method for monitoring crossing noses by evaluating the natural frequency obtained with a hammer impact test. A fault at the crossing nose can be detected through several modes of natural frequency. Kaewunruen et al. [62],[63],[64] considered the inspection of different components of the track and its supporting concrete, using impact hammer and natural frequency measurements. In their research, the damage was identified and quantified in a variety of components such as IRJs and concrete.

For evaluating the natural frequency of a target, two approaches have currently been researched. Firstly, focusing on a system which is not difficult to illustrate by mathematic solutions, the natural frequency of the system can be evaluated by solving the mathematic solutions directly. For example, a Single Degree of Freedom (SDOF) system, which is one of the common types of systems, can be expressed by a second-order differential equation. The natural frequency of the SDOF system can be evaluated by obtaining the material parameters and solving this second-order differential equation. However, most of the systems in the world have multiple degrees of freedom, and their natural frequency cannot be evaluated by solving the mathematic equation.

The second method is to use an FEM model to evaluate the natural frequency of a system. The current simulation research is able to maturely evaluate the natural frequency of a system. For example, Popp [131] explored a method for crack detection by natural frequency. The first 10 natural frequencies are obtained through a numerical FEM model to analyse the fault condition of basalt machine parts. He et al. [132] analysed the health condition of a cantilever beam through the natural frequency. The results were obtained by Ansys. This simulation model can predict and provide a reliable natural frequency value of the cantilever beam.

2.3.2 FEM techniques for the simulation of signals created in a vibration field

FEM models have been widely used in creating vibration signals, especially for simulating the vibration signal from a periodical machine [133],[134],[135]. These works supply a number of experiences to simulate the conditions of vibration from the machine. However, the initial condition of the periodical machine is periodical vibration, which is different from the initial condition of a hammer impact (an impulse force).

Considering the analysis of impulse signals which are impacted by other things (impact hammer, rail–wheel contact), Pletz et al. [136],[137] simulated and evaluated the condition of rail damage due to the wheel–rail interaction at the head of the crossing nose based on a 3-D FEM model. Schupp et al. [138] researched the dynamic response of wheel–rail interaction at the crossing nose. This research described the physical conditions when the rail is impacted, and also supplied the physical mechanism of the formation of the impact response signal. Iglesias and López [139] built an FEM model using Abaqus to estimate the Rayleigh damping parameters of aerospace materials through the hammer impact test. The vibration impact response signal obtained by the FEM model was compared with the experimental results to validate the accuracy of the FEM model. These studies can be referred to as the basis for building a correct vibration model.

2.3.3 FEM techniques for the simulation of signals created in an acoustic field

The principle of generating simulated acoustic signals by the FEM model has been researched since the middle of the last century [140],[141],[142]. Kopuz and Lalor [143] explored the condition of an interior acoustic field based on the FEM. The sound generated due to the vibration is predicted by the FEM model. In their research, the numerical function of acoustic propagation (which is well known as the Helmholtz equation) is linked and applied to the simulation model and then the discretisation method for solving the numerical model is discussed. Additionally, analysis of the acoustic method is completed by the FRF method. Their research indicates that it is possible to create an acoustic field which is generated through vibration. With improvements in computation and the simulation model, there has been more research focusing on analysis of the acoustic field through the FEM model [144],[145],[146]. Most of the recent research in this area focuses on the simulation of an acoustic field generated by AE. Considering the simulation of an acoustic field generated by the impact, Wang et al. [147] simulated the acoustic field which is generated by the hammer impact of a bolted joint on a beam. Metaphysical fields including an acoustic field and solid mechanics field were built for the simulation. A point in the simulation model was settled as the impact source and a sine pulse was applied to simulate the force of impact. The amplitude and period of the impact pulse were settled following the experimental data. Considering that the condition of this simulation was not relevant to the time, the simulation was solved in the frequency domain. And the frequency of acoustic pressure and the FRF were obtained as the results of this simulation for the analysis. These results were validated by comparison with the experiment results. However, this research has some limitations that need to be addressed. Firstly, the impact source is settled as a point, which is different from the hammer impact in the test. The profile of the hammer should be a square instead of a point. Secondly, the force given for the simulation model is a sine signal, which is different from the impact signal in the experiment. Most importantly, the result solved from this simulation is in the frequency domain. The impact response signal was still not obtained. Hence, in conclusion, there is sufficient research focusing on generating acoustic fields and obtaining acoustic signals but research on the simulation for obtaining the impact response signals is limited.

2.4 Machine learning-based method for railway fault classification

After extracting faulty features from a signal, fault classification will be used to distinguish the different types of faults. Machine learning is an intelligent application that has grown rapidly based on data analysis and computing. There are four different types of machine learning algorithm: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [148].

Generally, the research area of machine learning includes classification analysis [149], regression [150], data clustering [151], association rule learning [152], reinforcement learning techniques [153], and feature engineering and dimensionality reduction [154],[155],[156]. Classification is the mapping of a function from input variables to output variables. The variables are labelled in the different categories. For example, in fault inspection, classifying faults as ‘healthy’ or ‘unhealthy’ can be a problem. There are three types of classification problems which are shown as follows.

1) Binary classification is a simple classification task to classify two categories such as 'true and false' or 'yes and no' [157]. These two categories have a relationship between normal and abnormal.

2) Multiclass classification refers to classification tasks that have more than two categories. Different from binary classification, all categories are individual without a relationship. For example, railway fault inspection considers the three fault categories 'RCF crack', 'corrosion', and 'plastic flow'.

3) Multi-label classification is an important consideration and one of the examples is relevant to several categories. For example, considering railway fault inspection, cracks can be presented under the categories of 'RCF cracks', 'corrosion cracks', 'bolt cracks', etc.

Currently, classification methods have been maturely researched in the machine learning and data science literature. Common networks for classification are included in the next section.

2.4.1 Common classification methods

In this section, common networks for fault classification are introduced. The Naive Bayes method is based on Bayes theorem, with the assumption that each pair of features is independent [158]. The advantage of the Naive Bayes method is that it is able to use a smaller dataset to estimate the necessary hyper-parameters and obtain quick classification results [159]. Linear Discriminant Analysis (LDA) is a linear classifier based on fitting class conditional densities to generate the data. This method maps the given data into a lower dimensional space to simplify the complexity of the model and reduce the computational cost. An LDA model is suitable for data with the same covariance matrix [159]. Logistic regression is a statistical model used to solve questions of classification [160]. Logistic regression can be applied to the regression questions and classification questions. But it is not commonly applied for dealing with classification questions. A Support Vector Machine (SVM) is another machine learning technique for classification, which can also be used for regression. The principle of SVM is to create a single hyper-plane or set of hyper-planes. SVM is more efficient for classifying a dataset with high dimensional spaces. In terms of the different mathematical functions (kernel function) such as linear, polynomial, and Radial Basis Function (RBF), the performance of the SVM is different. But the performance of SVM is not efficient for a dataset with significant noise or overlapping target classes [161]. A decision tree is a non-parametric supervised learning method, which is widely used for both regression and classification. The famous algorithms of the decision tree include ID3 [162], C4.5 [163], Chaid [164], and CART [165]. The performance of a decision tree is good for dealing with a dataset with redundant attributes. However, small changes in the dataset will change the overall shape of the decision tree and any unsuitable data will badly influence its construction.

2.4.2 CNN and 1-D CNN models

Deep learning is the most popular method used for classification recently. The deep learning method is based on big data study and has been widely used for computer vision, speech recognition and pattern recognition, etc. [166]. The Convolutional Neural Network (CNN) is a common deep learning method. CNN has been widely used for image processing [167], image recognition [168], image classification [169], etc. The advantage of a CNN is the great computational burden and decent ability to automatically diagnose the sample's features. Nowadays, there are many mature deep learning models based on CNN such as AlexNet [170], ZFNet [171], NiN (Network In Network) [172] Visual Geometry Group (VGG) [173], Inception [174], Xception [175], and ResNet [176]. CNN models have previously been applied for dealing with two-dimensional data (images, video, etc.). Considering a one-dimensional signal, an improved CNN named 1-D CNN is proposed for the analysis. 1-D CNN uses the one-dimensional arrays as the kernel, which is the difference between 1-D CNN and 2-D CNN. But in terms of the convolutional layers and Multilayer Perceptron (MLP) layers, it is similar to 2-D CNN. 1-D CNN has been widely used for time series signals. A good example of the application of 1-D CNN on time series signals is that in bio-signals [177],[178],[179]. And considering the signals generated from the railway, 1-D CNN has been used for the classification of the structural fault on the railway system with the significant effectiveness [[180],[181]]. And also according to these researches, it proves that 1-D CNN is able to classify the one-dimensional signals. Hence, 1-D CNN will be used for the classification in this thesis.

2.5 Conclusion

In conclusion, a literature review was conducted to answer the research questions proposed in Chapter 1. A detailed review of current research which could correspond to the work in further chapters was completed.

Specifically, this chapter starts with an introduction to the various types of rail damage. From this review, it is understood that rail corrosion is one of the types of damage that needs to be given more concern. A review of the corrosion types was carried out, and the relevant maintenance standards are introduced.

Secondly, the review of approaches to railway inspection emphasises the importance of techniques for inspecting rail damage, as well as the application of signal processing to railway inspection. Specifically, several techniques for railway inspection were identified, such as visual inspection, ultrasonic testing, vibration, and acoustic listening. The advantages and disadvantages of each technique are addressed. And then, in order to identify faults based on the inspection methods, signal processing methods are introduced. Furthermore, the current methods for processing impact response signals are reviewed.

Thirdly, in order to build a simulation model for simulating impact response signals, the research on FEM simulation models is introduced. Specifically, the review identified several FEM techniques for simulating different aspects of railway systems, such as the natural frequency of structures, creating signals in a vibration field, and creating signals in an acoustic field. These simulations can help in predicting potential faults and enhancing the design of railway systems.

Finally, machine learning-based methods for railway fault classification have been examined, considering the development of common classification methods such as decision trees and SVM. The advantages and disadvantages of each approach were reviewed. Additionally, considering the deep learning method, the current research on relevant deep learning methods (CNN and 1-D CNN models) was reviewed.

Overall, the literature review presents a wide range of techniques and methods for monitoring and maintaining the safety and efficiency of railway systems. By incorporating proactive measures for preventing rail damage, enhancing railway inspection processes, and utilising advanced simulation and classification techniques, railway systems can be improved to provide better safety, efficiency, and reliability.

CHAPTER 3 INITIAL METHODOLOGY AND KEY TECHNIQUES

This chapter introduces the methods which are used in the thesis and discussed in depth in the following chapters. Firstly, considering the modelling and simulation, the method of coupling a multi-physical field is introduced. This method was used to simulate the process of forming an impact response signal and observing its propagation throughout the simulation. These validated simulations provided the volume of data required to develop algorithms for underside corrosion fault detection. Secondly, the method of selecting the appropriate signal processing algorithm for processing the impact response signal is introduced. Finally, an application of machine learning is introduced, to classify corrosion types of the rail base, as well as the levels of that corrosion.

3.1 Acoustic–vibration dynamic coupling simulation model

This simulation model aims to simulate acoustic impact response signals which are generated using a hammer. The model is constructed by two physical fields. The vibration field is used to generate the impact response signal by the hammer and the acoustic field is used to simulate the process of transferring the vibration to the acoustic signal.

In order to understand the relationship between an impulse, the rail structure, and the resulting impact response signal, the system has been considered both theoretically and practically. Initially, a damped Single Degree of Freedom (SDOF) system response function is used to represent the impact and to evaluate the response.

Impact response signals occurring on the railway are generally generated by manual hammer impact or percussion events at the wheel/rail interface. This physical phenomenon can be presented as shown in Equation 3.1 [182].

$$M\ddot{x} + C\dot{x} + Kx = P(t) \quad 3.1$$

where *K* represents the rail rigidity matrix, *M* is the mass matrix, *C* is the damped coefficient matrix, x, $\dot{x}$, and $\ddot{x}$ are displacement, velocity, and acceleration, respectively, and $P(t)$ is the impact excitation. If Φ is the regularisation master model matrix of Equation 3.1, given as $x = \Phi y$, adding Φy to Equation 3.1 and multiplying through by Φ^T gives Equation 3.2.

$$M_r\ddot{y} + C_r\dot{y} + K_r y = \Phi^T P(t) = F(t) \quad 3.2$$

where M_r is the model mass matrix, C_r is the model damped coefficient matrix, K_r is the model rigidity matrix, and $F(t)$ is the mode impact excitation. Equation 3.2 consists of *n* independent linear differential equations with second-order constant coefficients. Each linear differential equation is presented as shown in Equation 3.3.

$$\ddot{y}_i + 2\sigma_i\dot{y} + \omega_i y = F_i(t)/m_i (i = 1,2\ldots\ldots n) \quad 3.3$$

where σ_i, ω_i, and m_i are an attenuation coefficient, the damping ratio, and the mass of the *i*th component respectively.

Assuming the initial condition is $x(0) = \dot{x}(0) = 0$, and combining it with Equation 3.1, the SDOF damped impulse model is therefore as shown in Equation 3.4.

$$\begin{cases} M\ddot{x} + C\dot{x} + Kx = P(t) \\ x(0) = \dot{x}(0) = 0 \end{cases} \quad 3.4$$

Under these boundary conditions, Equation 3.3 can be solved using Duhamel's integral transferred to the frequency domain. Its solution is given by Equation 3.5.

$$y_i(t) = \frac{1}{m_i \omega_{di}} \int_0^t F_i(t) e^{-\sigma_i(t-\tau)} \, sin(\, \omega_{di}(t-\tau)) d\tau \qquad 3.5$$

where ω_{di} is the i^{th} order eigenvalue of the characteristic equation and dependent on the target structure. In this application, the impact response signal is therefore dependent on the different rail structures being considered.

Simplifying Equation 3.5, applying the initial conditions from Equation 3.4, and integrating, the SDOF damped system response function can be given as shown in Equation 3.6.

$$x(t) = \frac{1}{m\omega_d} e^{-nt} \, sin(\, \omega_d t) \qquad 3.6$$

where ω_d is the total eigenvalue which summarises all ω_{di}.

Ideally, an impact response signal would be constructed by superposing multiple instances of Equation 3.6 with different parameters. And Equation 3.6 indicates that the features within an impact response signal are not dependent on the intensity of the impulse.

The impact sound is generated by air molecules that are around the rail. The molecules, interfered with by rail vibration, will change the air density to generate a sound wave radiating around the rail. The principle of coupling the vibration and acoustic fields is used to find the velocity of molecules at the interface of the vibration and acoustic fields.

Considering the wave propagation in a single dimension (*x*-axis), the relationship between the acoustic pressure and the velocity of air molecules can be given in Equation 3.7.

$$\rho \frac{dv}{dt} = -\frac{\partial p}{\partial x} \quad 3.7$$

where ρ is the air density, p is acoustic pressure, and v is the velocity of air molecules. And the relationship between the density and the velocity of air molecules can be given by Equation 3.8.

$$-\frac{\partial}{\partial x}(pv) = \frac{d\rho}{dt} \quad 3.8$$

Considering the air molecules have mass, the pressure of air molecules will be changed corresponding to the interference of the acoustic field. While the acoustic field is changed (interfering with the vibration), the relationship between the pressure and density of air molecules can be presented in Equation 3.9.

$$dP = c^2 d\rho \quad 3.9$$

where P is the pressure of air molecules and c^2 is the acoustic velocity.

The above equation considers sound propagation in a single dimension. In terms of three dimensions, Equation 3.7 can be expanded as per Equation 3.10, and Equation 3.9 can be expanded as shown in Equation 3.11.

$$\rho \frac{dv}{dt} = -gradp \quad 3.10$$

$$\nabla^2 p = \frac{1}{c_0} \frac{\partial^2 v}{\partial t^2} \quad 3.11$$

where $gradp$ is the gradient of the acoustic pressure and ∇^2 is the Laplace operator which can be expanded as $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

Then, based on Equation 3.10 and Equation 3.11, the process of acoustic propagation through the air due to the vibration can be given as shown in Equation 3.12.

$$\frac{\partial^2 p}{\partial r^2}+\frac{2}{r}\frac{\partial p}{\partial r}=\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} \tag{3.12}$$

where r is the distance of wave propagation. By solving Equation 3.12, the acoustic pressure can be presented as in Equation 3.13:

$$p=\frac{A}{r}e^{j(\omega t-kr)}+\frac{B}{r}e^{j(\omega t+kr)} \tag{3.13}$$

In this equation, *A* and *B* are undetermined constants, ω is the angular frequency, and *k* is the propagation constant. The first term of Equation 3.13 represents the acoustic pressure generated from the impact position propagating to the outside field. The second term represents the acoustic pressure reflected from the simulation boundary. Assuming that there is no reflection from the outside barriers (i.e., the second term equals 0), Equation 3.13 can be simplified to give Equation 3.14:

$$p=\frac{A}{r}e^{j(\omega t-kr)} \tag{3.14}$$

Based on Equation 3.8 and Equation 3.10, the relationship between the vibration speed and density can be rewritten as Equation 3.15.

$$\frac{\partial p}{\partial r}=-\rho_0\frac{\partial v_r}{\partial t} \tag{3.15}$$

where v_r is the vibration speed in three dimensions and ρ_0 is the density of the air field. Then, combining Equation 3.14 and Equation 3.15, the acoustic speed can be solved as in Equation 3.16.

$$v_r = \frac{1}{j\omega\rho_0}\frac{\partial p}{\partial r} = \frac{A}{r\rho_0 c_0}\left(1 + \frac{1}{jkr}\right)e^{j(\omega t - kr)} \quad 3.16$$

The vibration velocity in the frequency domain is given as Equation 3.17.

$$u = u_a e^{j(\omega t - kr_0)} \quad 3.17$$

where u_a is the amplitude of the vibration velocity in the frequency domain and kr_0 is the initial phase of the vibration velocity. Due to the velocity of the acoustic molecules equalling the velocity of the vibration velocity, the boundary condition can be given as Equation 3.18.

$$v_r|r = r_0 = u \quad 3.18$$

Combining Equation 3.17 and Equation 3.18, the constant *A* can be calculated as shown in Equation 3.19.

$$A = \frac{\rho_0 c_0 r_0^2}{1 + (kr_0)^2}(kr_0 + j)u_a = |A|e^{j\theta} \quad 3.19$$

Assuming the initial phase kr_0 is zero, $\theta \approx \frac{\pi}{2}$, the acoustic pressure can be written as Equation 3.20.

$$p \approx j\frac{k\rho_0 c_0}{4\pi r}Q_0 e^{j(\omega t - kr)} \quad 3.20$$

where $Q_0 = 4\pi r_0^2 u_a$ represents the amplitude of acoustic speed, i.e., acoustic intensity. Once u_a is obtained from the vibration model based on Equation 3.6, the acoustic pressure can be calculated using Equation 3.20. In the laboratory experiment, the acoustic signal can be detected through the microphone. Hence, the level of acoustic pressure generated from the simulation model can represent the acoustic signal. By analysing the acoustic pressure, it is possible to obtain information of the rail structure.

In conclusion, the acoustic–vibration dynamic coupling simulation model presented in this section aims to simulate the generation of impact response signals in a railway system through manual hammer impact. The model consists of the vibration field and acoustic field. Considering the vibration model, the vibration is generated from the hammer impact. Using the coupling method, the vibration source is transferred to the acoustic model as the initial condition. The acoustic signal is generated by this vibration source and propagates to the acoustic field. Then, the acoustic signal can be obtained from the acoustic field, which is used to analyse the information of the rail structure. Hence, in order to extract the information from the acoustic signal, the potential methods for processing the impact response signal are introduced in the next section.

3.2 Signal decomposition method

After obtaining the signal from the simulation model, a suitable method needs to be found to extract the features from different faulty types of signals. As introduced in Chapter 2, compared to traditional Fast Fourier Transform (FFT) and wavelet transform methods, mode decomposition methods have the advantages of identifying and categorising minor faults more accurately [99]. Hence, in this section, the key decomposition methods including Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), Local Mean Decomposition (LMD), and Variational Mode Decomposition (VMD) are introduced.

Based on the discussion in the literature review, EMD, EEMD, VMD, and LMD are the four classic decomposition methods that can potentially be considered to decompose the impact response signal. The basic theories of these four decomposition methods are summarised as follows.

3.2.1 Basic theory of EMD

EMD is a recursive method for analysing non-linear and non-stationary signals that adaptively decomposes any signals into their oscillatory components which are called Intrinsic Mode Functions (IMFs).

The general process of EMD is shown as follows:

1) All of the local maxima/minima of a signal are connected by a pair of cubic splines to form upper and lower envelopes.

2) The mean values of pairs (upper and lower) along the splines are calculated as m_1.

3) The first component is obtained through these mean values subtracted from the original data $f(t)$ at corresponding *x*-axis locations, which is called sifting [183] and is represented by Equation 3.21.

$$h_1[(n)] = f(t)[n] - m_1 \qquad 3.21$$

   $h_1[n]$ should satisfy two conditions. Firstly, the number of extrema and zero crossings must be equal or have a maximum difference of less than 1. Secondly, the average value of the envelope should be zero at any given point.

4) If $h_1[f(t)]$ satisfies the above conditions, it is defined as the first IMF $f_1[n]$.

5) If $h_1[f(t)]$ does not satisfy the above conditions, it is treated as a signal to be sifted through step 1) and step 2). Hence, the second component can be obtained as Equation 3.22.

$$h_2[n] = f(t)[n] - m_2 \tag{3.22}$$

where m_2 is the mean value of $h_1[n]$. If $h_2[n]$ does not satisfy the above conditions, steps 3) and 4) are repeated.

The above conditions are performed through an index against the Standard Deviation (SD) which is calculated from two consecutive sifting results ($h_{i-1}[n]$ and $h_i[n]$), which is given as Equation 3.23.

$$SD = \sum_{n=0}^{N} |h_{i-1}[n] - h_i[n]| \tag{3.23}$$

When the SD value is in a predefined range, the sifting is stopped.

The residue $r_1[n]$ is obtained by subtracting the first IMF $f_1[n]$ from the original data $f(t)[n]$, which is given as Equation 3.24.

$$r_1[n] = f(t)[n] - f_1[n] \tag{3.24}$$

$r_1[n]$ is then subjected to the same sifting process, which is given as Equation 3.25.

$$r_j[n] = r_{j-1}[n] - f_j[n] \tag{3.25}$$

The sifting process is stopped while $r_j[n]$ is a constant or monotonic function. All the residues can be presented as $r_j[n]$. And the total residue R is computed by summing all residues. The EMD is then stopped. The number of repetitions is given as N. The decomposition of the original signal can be present as Equation 3.26.

$$f(t) = \sum_{i=1}^{N} f_i[n] + R \tag{3.26}$$

3.2.2 Basic theory of EEMD

EEMD is an improved signal processing method which considers the influence of background noise and aims to reduce this noise by applying a population averaging process to multiple tests [106]. The technique is defined as a Noise Assisted Data Analysis (NADA) method as it also incorporates noise data which are then applied in average-based processing to obtain the final result.

EEMD is a method based on the EMD algorithm, which solves the problem of mode distortion in EMD (mode distortion refers to the phenomenon where a signal is altered such that the shape or characteristics of its modes or frequency components are changed). The principle of EEMD uses an ensemble of white noise-added signal replicas, which are decomposed into IMFs using the EMD method. The EEMD method then calculates the mean of the IMFs obtained from each of the white noise-added signal replicas to obtain the final IMF, and the added noise can be eliminated.

The details of EEMD are given as follows:

The new signal $Y_n(t)$ is generated by adding the white noise signal wgn_n to the original signal $f(t)$ and is given by Equation 3.27.

$$Y_n(t) = f(t) + wgn_n \tag{3.27}$$

where n is the number of trials.

1) $Y_n(t)$ is then decomposed as a set of IMFs and a residual by EMD, which is represented as Equation 3.28.

$$Y_n(t) = \sum_{i=1}^{N} IMF_i + R \tag{3.28}$$

2) After n trials by repeating steps 1) and 2), the final i^{th} (assuming $Y_n(t)$ is decomposed into i IMFs) IMF of EEMD is computed by averaging the total i^{th} IMFs decomposed by n time series. The process of decomposing IMFs through EEMD is given as Equation 3.29.

$$IMF_i = \frac{1}{n}\sum_{i=1}^{n} IMF_i \tag{3.29}$$

The results decomposed by EEMD depend on the decomposition mode (n) and the amplitude of the noise.

3.2.3 Basic theory of VMD

The principle of the VMD algorithm is to decompose the components of a signal into different adaptive bands for the development of a generalised Wiener filter. The input signal is decomposed into limited bandwidth modes, u_k, with each sub-signal tightly focused on its own central frequency. The bandwidth of each mode can be evaluated by solving the constrained variational problem shown in Equation 3.30.

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \alpha_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] exp^{-j\omega kt} \quad \right\|_2^2 \right\}$$

3.30

$$s.t. \sum_k u_k = f$$

where δ is the Dirac distribution, u_k is the set of all modes, and ω_k is the centre frequency of each mode.

In order to remove the constraints on Equation 3.30, the constrained variational problem can be solved considering a Lagrangian multiplier $\lambda(t)$ and a quadratic penalty term α, which is given in Equation 3.31.

$$L(\{u_k\},\{\omega_k\},\lambda) =$$

$$\alpha \sum_k \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] exp^{-j\omega kt} \quad \right\|_2^2 \right\}$$

3.31

$$+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle$$

In order to obtain an optimal solution of the constrained variational model, the Alternative Direction Method of Multipliers (ADMM) is applied for solving Equation 3.31. Hence, the mode u_k and centre frequency ω_k can be updated as shown in Equation 3.32 and Equation 3.33, respectively.

$$u_k^{n+1} \leftarrow \underset{u_k}{argmin} L(\{u_{i<k}^{n+1}\},\{u_{i\geq k}^{n+1}\},\{\omega_i^{n+1}\},\lambda^n) \qquad 3.32$$

$$\omega_k^{n+1} \leftarrow \underset{u_k}{argmin} L(\{u_i^{n+1}\}, \{u_{i<k}^{n+1}\}, \{\omega_{i\geq k}^{n+1}\}, \lambda^n) \tag{3.33}$$

Decomposition to identify all sub-signals (IMFs) can be achieved through continuous iteration as presented in Equation 3.34.

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{3.34}$$

3.2.4 Basic theory of LMD

LMD is based on obtaining a Product Function (PF) from the original signal. The principle is separated into two steps. Firstly, the original signal is smoothed by extracting local means and estimating the envelope. Secondly, the smoothed signal is subtracted from the original signal. And then the envelope estimation [184] is used for amplitude demodulating the result.

The LMD decomposition process is shown as follows:

1) The mean value of the maximum and minimum points of each half-wave oscillation of the original signal is computed. The mean value of each two successive extrema (m_i) is given as Equation 3.35.

$$m_i = (n_i + n_{i-1})/2 \tag{3.35}$$

where n_i and n_{i-1} are two adjacent external data points. The local means m_i can be plotted as many straight lines. And then the approach of moving averaging is used to smooth these lines to obtain a smoothed local mean function l_{ij}.

2) The local magnitude of two adjacent extrema (n_i and n_{i-1}) is computed by Equation 3.36.

$$a_i = |n_i - n_{i-1})/2| \qquad 3.36$$

By plotting all local magnitudes as step 1), the estimated envelope a_{ij} can be calculated (j is the demodulated time).

3) Subtract the local mean function l_{ij} from the original signal $f(t)$, as presented in Equation 3.37.

$$h_{ij} = f(t) - l_{ij} \qquad 3.37$$

4) Then the demodulated components s_{ij} are computed by dividing h_{ij} by the estimated envelope a_{ij}, which is represented in Equation 3.38.

$$s_{ij} = \frac{f(t) - l_{ij}}{a_{ij}} = \frac{h_{ij}}{a_{ij}} \qquad 3.38$$

5) Multiplying all estimated envelopes a_{ij} from different demodulations, the integrated envelope is calculated by Equation 3.39.

$$a_i = a_{i1}a_{i2} \cdots a_{ij} = \prod_{j=1}^{n} a_{ij} \qquad 3.39$$

where the ultimate objective is $\lim_{n\to\infty} a_{ij} = 1$.

The PF can then be calculated through Equation 3.40.

$$PF_i = a_i s_{ij} \tag{3.40}$$

6) The first PF is subtracted from the original signal to obtain a new signal u_1; and u_1 is treated as the new original signal to subtract the second PF. This process is repeated until u_1 contains no more oscillations or becomes a constant. The process is presented in Equation 3.41.

$$\begin{cases} u_1 = f(t) - PF_1 \\ u_2 = u_1 - PF_2 \\ \vdots \\ u_k = u_{k-1} - PF_k \end{cases} \tag{3.41}$$

Finally, the signal is decomposed as k PFs and a residual, which is presented as Equation 3.42.

$$f(t) = \sum_{i=1}^{k} PF_i + u_k \tag{3.42}$$

The above four decomposition methods will be used to process the impact response signal. Their capability and effectiveness will be analysed in Chapter 5. After conducting the analysis, the decomposition method which demonstrates the greatest effectiveness in decomposing impact response signals will be optimised for feature extraction. The features obtained from various impact response signals using the selected decomposition method will then be used for fault classification in order to identify the different rail conditions.

3.3 Machine learning-based techniques

As introduced in Chapter 2, deep learning is the most popular method used for classification recently. Deep learning methods are based on big data study and have been widely used for computer vision, speech recognition, pattern recognition, etc. [166]. CNN is one of the common deep learning methods used for classification. In this thesis, a 1-D CNN will be used for classification of the acoustic impact response signals. A CNN consists of an input, convolution layer, pooling layer, fully connected layer, and output layer, as shown in Figure 3.1.

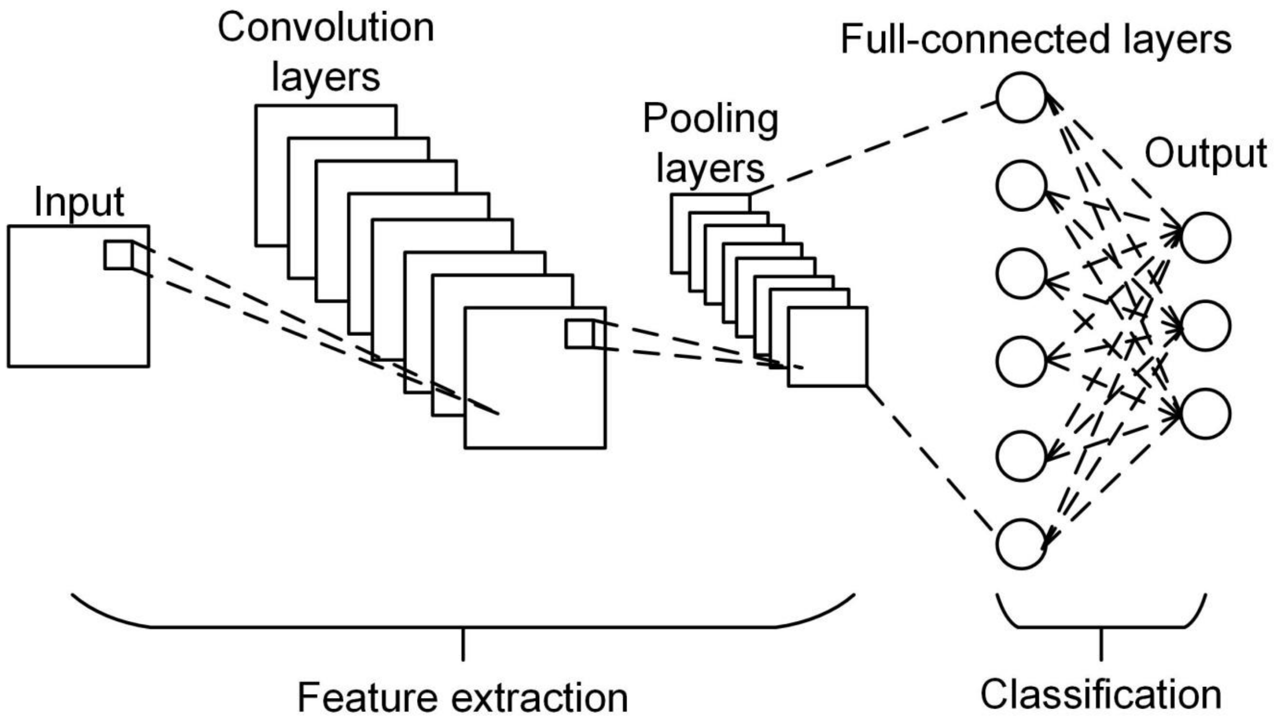


Figure 3.1 Structure of a CNN

The core of a CNN is the convolution layer. Each convolution layer contains a lot of convolution kernels which are used for processing the input data and extracting the features. In a 1-D CNN, the data from the previous convolution layer $l-1$ (output) to the l layer (input) are demonstrated as Equation 3.43, where x_k^l is a neuron of the input layer; b_k^l is the scalar bias of the k^{th} neuron of the i^{th} input layer; s_i^{l-1} is the i^{th} neuron from the $l-$

1 output layer; and ω_{ik}^{l-1} is the kernel from the i^{th} neuron of the $(l-1)^{\text{th}}$ output layer to the k^{th} neuron of the l^{th} input layer.

$$x_k^l = \sum_{i=1}^{N_{l-1}} conv1D\left(\omega_{ik}^{l-1}, s_i^{l-1}\right) + b_k^l \tag{3.43}$$

The neuron of intermediate output is defined as y_k^l, which can be obtained from input x_k^l. It can be presented as Equation 3.44, where s_k^l is a neuron of the output layer; $\downarrow ss$ is the down-sampling operation with the scalar factor (ss).

$$\begin{gathered} y_k^l = f\left(x_k^l\right) \\ s_k^l = y_k^l \downarrow ss \end{gathered} \tag{3.44}$$

Then the output from convolution layers is transferred to the pooling layers for further processing. The function of the pooling layer is to compact the number of parameters of the neural network. Max pooling and mean pooling are two common pooling methods, and max pooling is more popular applied for the classification as it introduces slight translation and distortion invariance, accelerates convergence, and enhances generalisation capabilities [185]. Therefore, in this thesis, the max-pooling method is selected, which can be given by Equation 3.45.

$$\begin{gathered} g_k^l = f\left(s_k^l\right) \\ z_k^l = \beta_j^l g_k^l \downarrow ss \end{gathered} \tag{3.45}$$

where β_j^l is the scalar bias.

Then the features extracted from the convolutional layers and pooling layers are integrated to the fully connected layer and presented as Equation 3.46. The output p^l represents a feature which could be relevant to one of labels from the dataset.

$$p^l = f(w^l x^{l-1} + b^l) \tag{3.46}$$

where w^l is the weight of layer l.

In machine learning, hyper-parameters are parameters that are not learned during the training process but are set manually by the user prior to training. These parameters determine the behaviour and performance of the machine learning model. Hyper-parameters can have a significant impact on the model's accuracy and generalisation ability. For example, the learning rate is a hyper-parameter that controls the step size of the gradient descent algorithm during training. It is a crucial parameter to optimise the performance of a model, as too low a learning rate can cause slow convergence and too high a learning rate can cause the model to diverge or overshoot the optimal solution. The learning rate determines how much to adjust the weights and biases of the model in response to the error (or loss) function calculated during training. A small learning rate means the model learns slowly, while a high learning rate means the model learns quickly.

The activation function is another hyper-parameter. It is a function in a neural network to decide if a neuron should be activated or not. The most common activation functions are defined as three categories: ridge functions, radial functions, and fold functions [186]. The common activation functions are Sigmoid, Tanh, and ReLU (Rectified Linear Function). The definition of these three functions are given in Equations 3.47, 3.48, and 3.49.

$$S(x) = \frac{1}{1 + e^{-x}} \qquad 3.47$$

$$S(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad 3.48$$

$$S(x) = \max(0, x) \qquad 3.49$$

The graphs of these three activation functions are given in Figure 3.2.

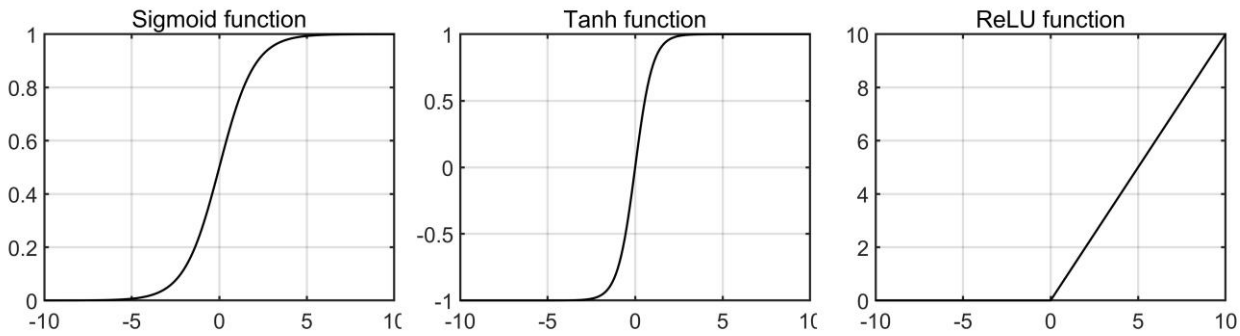


Figure 3.2 Plots of three activation functions

Different activation functions possess distinct advantages and disadvantages. For example, the Sigmoid and Tanh functions can contribute to the occurrence of the vanishing gradient problem, which refers to the phenomenon of extremely small or zero gradients of the loss function with respect to the weights in the lower layers during the training of deep neural networks. And ReLU can only learn on examples for which their activation is not zero. Hence, in this thesis, the three activation functions are all applied for learning and the best will be selected by evaluation of machine learning performance.

Besides the above hyper-parameters, there are others such as hidden layers, number of epochs, dropout rate, batch size, initialisation of weights, etc [187]. These parameters determine the behaviour and performance of the machine learning model. And these parameters need to be settled before the classification. Therefore, based on the classified dataset in this thesis, a method considering automatically selected hyper-parameters will be introduced in Chapter 6 to obtain the best performance of the classification from the 1-D CNN model.

In conclusion, this section introduces the basic theory of the CNN. The specific hyper-parameters are also introduced. The theory of 1-D CNN will be used for creating the classification model in Chapter 6. A method for automatically selecting hyper-parameters will also be introduced in Chapter 6. Additionally, in order to obtain the best rail base corrosion classification accuracy, a 1-D CNN model with a special structure is proposed and will be introduced in Chapter 6.

CHAPTER 4 FEM MODEL TO SIMULATE IMPACT RESPONSE SIGNALS

This chapter introduces a method for creating different types of simulated impact acoustic response signals using an FEM simulation model. Firstly, a simulated impact response signal representing the condition of a hammer impact on the rail was generated based on a Single Degree of Freedom (SDOF) system which is introduced in Chapter 3. By changing the parameters of this system, the patterns of different rail structures were investigated. However, the signal generated from the SDOF system is not able to reflect the rail conditions in the real world. Hence, secondly, a 3-D FEM simulation model was built. The process of building the simulation model is presented. The Frequency Response Function (FRF) model was first built up to validate whether the geometry of the rail in the simulation model is similar to that of a real rail. And then, as mentioned in Chapter 3, a simulation model was built for generating the simulation signal. The acoustic signal was generated through vibration and propagation in an acoustic field. There are, therefore, three steps of model building that will be introduced in this section: the vibration model, the acoustic model, and the coupled multi-physical model, respectively. In order to validate the accuracy of the built simulation model, some comparisons were made with the results obtained in laboratory experiments. After validation, a simulation model with different rail geometries (i.e., the different rail fault types) was used to generate simulated signals. These simulated signals were then used to create a simulated dataset for further analysis and machine learning applications.

4.1 SDOF system

This section introduces the use of an approach more commonly associated with control systems in the simulation of rail conditions. The rail was modelled as a physical system by feeding the basic material parameters of the rail into the system. The system can be described mathematically by a set of differential equations that represent the motion of the mass element. By applying an impulse signal as the input, the output of the system can be obtained, which represents the impact response signal of the rail. And then, by changing the basic material parameters of the rail, signals representing different rail conditions can be obtained and analysed. In summary, this section explores the difference of the impact response signals obtained from different rail conditions through a simulation model. And the possibility of applying this simulation model to create simulated signals to represent the signals from real conditions is discussed.

Each physical model can be represented as a system which contains a unique physical formula. The system will not be changed when the input and output are changed. In the railway domain, there are many conditions that can be illustrated by a physical system. For example, the catenary system can be represented as a system with two degrees of freedom [188]; the secondary suspension systems of rail vehicles can be represented by a physical system [73]. Considering the rail, a damped SDOF system is used to represent the impact and to evaluate the response. The parameters of the SDOF system, including the mass, damping coefficient, and rigidity, can be adjusted to reflect the characteristics of different rail structures. In this way, the simulated impact response signal can be generated to represent the rail condition.

As mentioned in Chapter 3, the impact response signals occurring on the railway generally are generated by a manual hammer impact or percussion events at the wheel/rail interface. The SDOF system consists of mass (*M*), damping coefficient (*C*), and rigidity (*K*). In terms of the structure of the rail, the values of these three parameters are given based on the work of Zhao [189] (damping coefficient will be represented by coefficient ratio), as shown in Table 4.1.

Table 4.1 Rail structure parameters

Geometric size and parameters	Value
Rigidity ($KN \cdot mm^{-1}$)	41
Mass (kg)	5000
Coefficient ratio	0.02

An impulse signal is then added to the SDOF system to obtain the impact response signal. A step signal is used for the input signal to simulate the condition of the hammer impact. A simulated signal is then obtained through the given input signal and the parameters of the SDOF system, as shown in Figure 4.1.

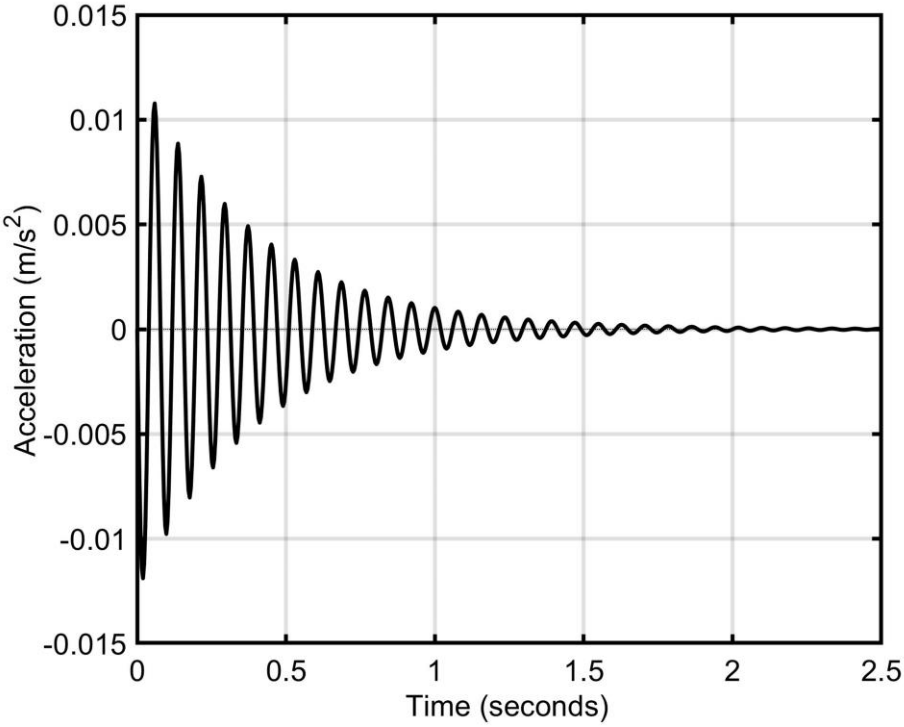


Figure 4.1 Simulated signal obtained from SDOF system

Then, the parameters of the SDOF system are changed within a range in order to present the patterns of the signal generated by different rail structures. As the rigidity and damping coefficient won't change when rail material is lost, the variance of the parameter changes only considers a change in the mass. Considering the level of rail damage (material loss), the variance of mass is given as 90% and 95%.

A Bode plot of the SDOF systems and frequency analysis of the signal are then applied to illustrate the difference between healthy and damaged rails. Firstly, the Bode plot represents the property of the designed SDOF system. The method of obtaining the Bode figure is based on the FRF. The FRF is obtained by calculating the frequencies from the input data and output data. The frequencies of input and output data are obtained through a Fourier transform. The Fourier transform of the output is given as $V(f)$ and that of the output is given as $U(f)$. And then the FRF $H(f)$ is defined through the frequency of the output divided by the frequency of the input. The details of obtaining the Bode plot are given in Equation 4.1.

$$H(f) = \frac{V(f)}{U(f)} \tag{4.1}$$

The Bode plots of SDOF systems corresponding to healthy and damage rails, respectively, are shown in Figure 4.2.

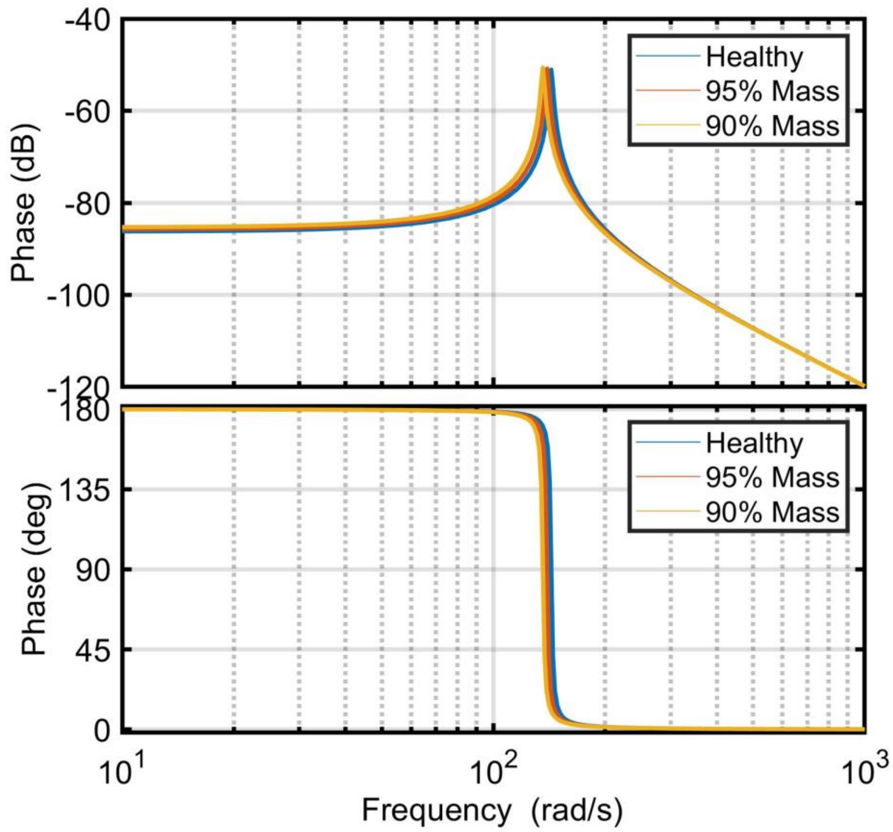


Figure 4.2 Bode plots of SDOF systems corresponding to different rail conditions

By comparing the output signals from different SDOF systems, the differences in the natural frequency of the rail can be detected and used for rail fault diagnosis. The change in the output signal from different SDOF systems (different rail conditions) can be observed in Figure 4.3.

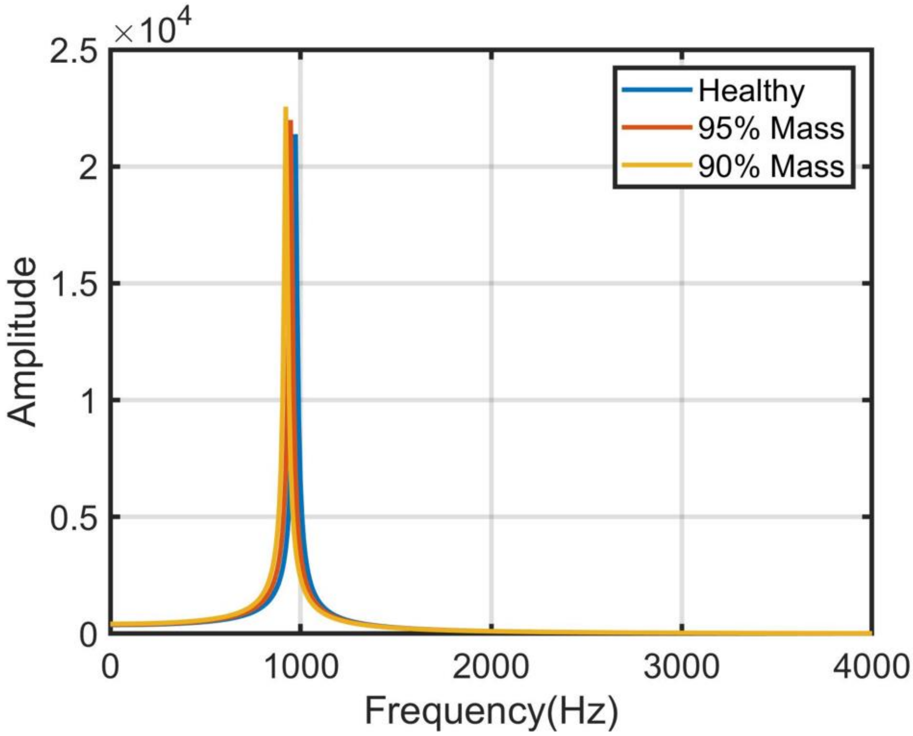


Figure 4.3 Frequency of simulated signals from SDOF systems

From Figure 4.3, it can be seen that the frequencies from different simulated signals are different. Frequency increases gradually when the mass increases. Hence, by analysing the frequency properties from the impact response signal, different rail conditions can be identified.

However, the above simulation is a simple one for investigating the patterns of different rail conditions (i.e. healthy and faulty rails). This simple solution can be used to rapidly explore the effects of varying parameters corresponding to the health of the rail. However, it is just a simplified representation of a real rail system and there is a limited correlation between simulation and experimental results. Furthermore, considering the current technique method, it is too difficult to enhance this system to obtain a signal which is similar enough to a real impact response signal (rail structure is a multi-dimension system, which is much more complex than an SDOF system). Hence, a detailed simulation model relating to more realistic conditions of a hammer impact on the rail is required. The signal obtained from this model can be related to the signal which is generated in real experiments by a hammer impact. Therefore, in order to obtain a more accurate simulation model, a 3-D FEM simulation model has been developed.

4.2 Model introduction

This section introduces the process of building a coupled multi-physical model. The simulation is based on the finite element software COMSOL Multiphysics. This model considers all the important physical phenomena and interactions that occur during a hammer impact on the rail, such as the vibration of the rail, the propagation of acoustic waves in the surrounding environment, and the coupling of these physical phenomena. Specifically, firstly, the simulation to obtain the FRF of the rail is completed. The natural frequency (i.e., eigenfrequency), which represents the information of the simulated rail, is obtained. This eigenfrequency will be compared with that obtained from the experiments, to ensure the simulated rail is the same as the rail in the experiments. Secondly, from the impact response signal generated through the hammer impact, a model simulating the process of the hammer impacting the rail to obtain the vibration signal (also called the vibration field) is introduced. Thirdly, the process of generating an acoustic signal through the vibration mentioned in the previous section (also called the acoustic field) is presented. Finally, the process of building a coupled multi-physical model (coupling the vibration field and the acoustic field) is introduced. This coupled simulation model will be used to generate the simulated acoustic impact response signals.

4.2.1 FRF model

Before obtaining simulated signals through the simulation model, it was necessary to understand whether the simulated target is the same as the real world. Hence, a pre-simulation was completed in order to obtain the eigenfrequency of the simulated rail.

The method of calculating the eigenfrequency was based on the FRF [190]. In COMSOL Multiphysics, there is a mature algorithm to calculate the eigenfrequency [191],[192]. In order to obtain the eigenfrequency of the rail, the process of building the FRF model is presented as follows.

In order to obtain the FRF of the rail, the rail section was considered in isolation, i.e. without any other rail components added. Hence, only a rail based on the standard of 56E1 rail [30] was created for the analysis. And boundary condition were given as six degrees of freedom for simulating suspending the rail for obtaining the FRF. The rail parameters were given considering the standard of the rail materials [193]. The parameters of the rail materials are shown in Table 4.2 and the geometry of the simulated rail is shown in Figure 4.4.

Table 4.2 Geometric size and parameters of the rail

Geometric size and parameters	Value
Length (mm)	2000

Young's modulus (Pa)	2e8
Poisson ratio (0.29)	44.5
Density $(kg \cdot m^{-3})$	7870

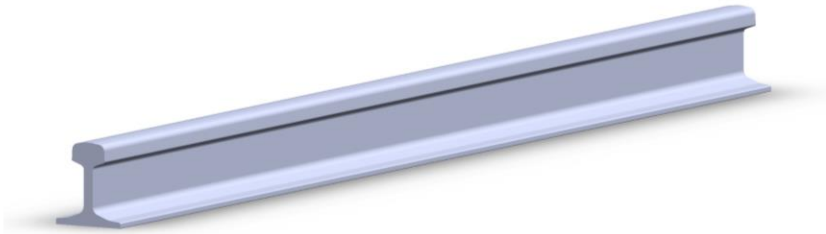


Figure 4.4 Geometry of the simulated rail

The eigenfrequency results are shown in Figure 4.5. It can be seen that the eigenfrequency of the selected rail was 111.13 Hz. This result was validated by the FRF experiment in the laboratory, which is described in the following section.

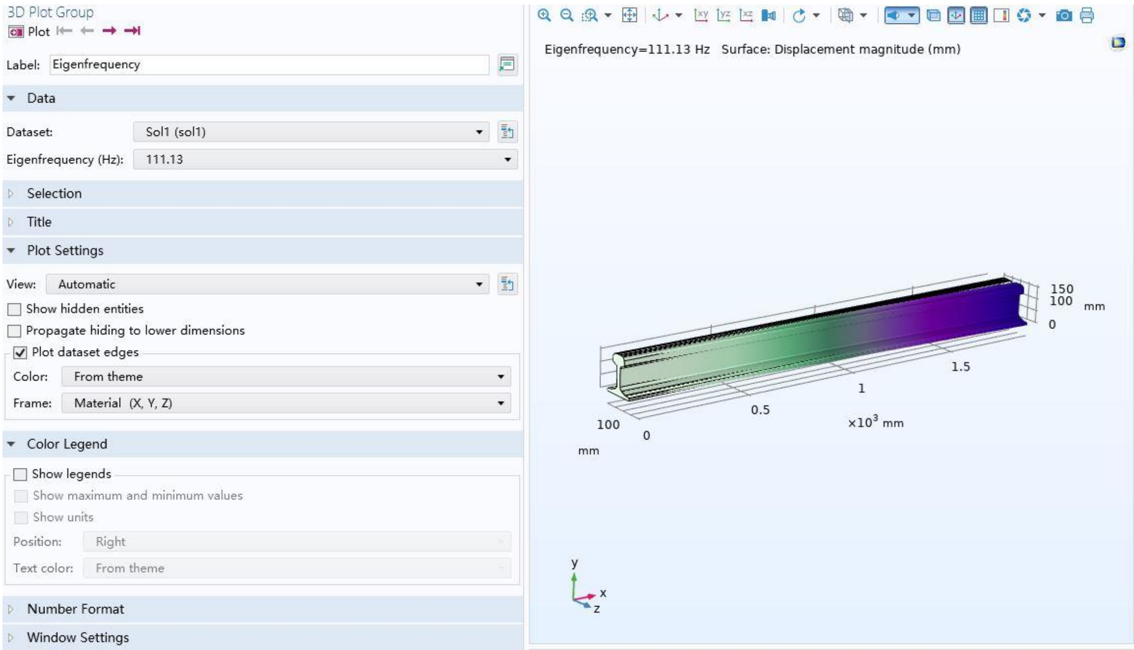


Figure 4.5 Eigenfrequencies of the rail

4.2.2 Vibration model

Based on the theory in Section 3.1, a vibration model was built up. The model contains the rail which was used for FRF simulation in Section 4.2.1. Additionally, a base was set at the bottom of the rail and the site of the hammer impact was chosen to be at the top of the rail and close to the rail end. The impulse can be simplified as a step signal. The impulse time was chosen based on the current analysis and experiments [194]. As the force of the impulse would not affect the frequency of the signal (as mentioned in the methodology), a value of 160 N was chosen for the simulation. The impulse time was set as 0.02 s; the simulated impulse is shown in Figure 4.6.

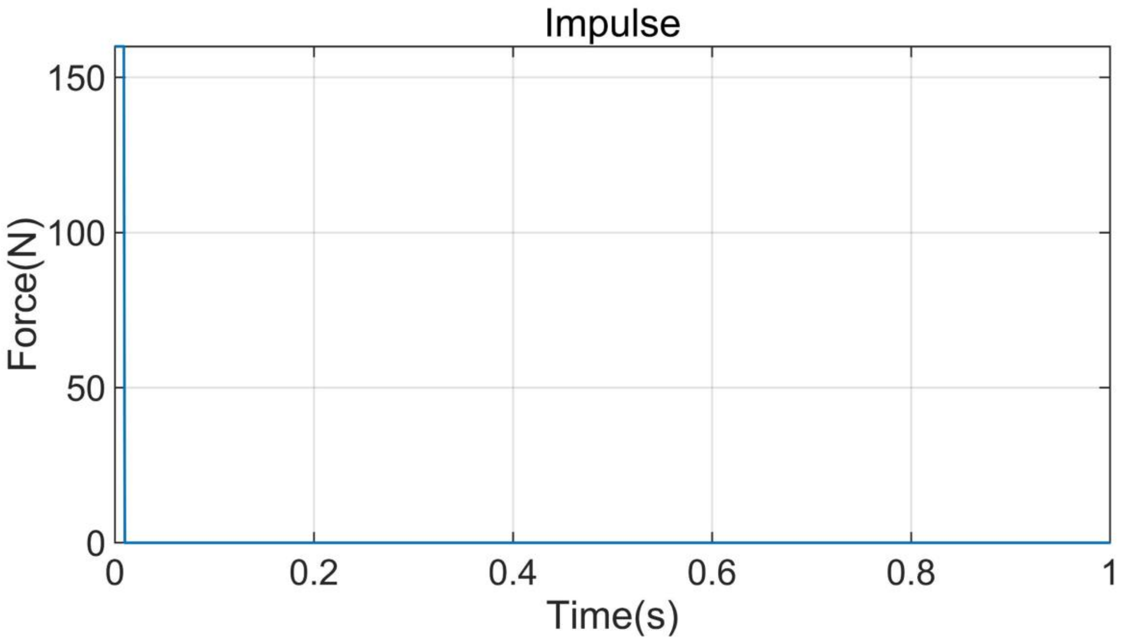


Figure 4.6 Hammer impulse

Before the simulation, the boundary conditions of the vibration model should be set. To simulate the hammer impact test in the real world, the bottom surface of the rail in the simulation was fixed. Additionally, the mesh size was set manually. The maximum mesh size (h_{max}) should be lower than one and a half times the minimum wavelength (λ_{min}). The formula for calculating mesh size is given in Equation 4.2.

$$h_{max} = \frac{\lambda_{min}}{1.5} = \frac{c}{1.5 \cdot f_{max}} \qquad 4.2$$

where c is the wave speed and f_{max} is the maximum solved frequency.

The calculation time and step are similar to those in the experiments. The parameters for initialising the vibration model are shown in Table 4.3.

Table 4.3 Parameters for initialising the vibration model

Parameter		Value
Dynamic impulse (*kg·m/s*)		160
Number of mesh elements	Number of vertex elements	Number of mesh elements
	Number of edge elements	2289
	Number of boundary elements	5282
	Number of elements	9422
Calculation step (s)		1e−5
Calculation time (s)		1

The result of the simulation model is presented as the displacement, which is shown in Figure 4.7.

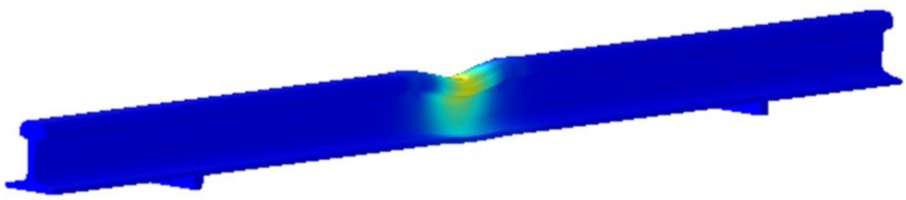


Figure 4.7 Displacement result of the simulation model

Figure 4.7 illustrates the condition of the rail after being impacted by the hammer.

In view of the Figure 4.7, the different of colours at the middle of the rail corresponds to the stress on the rail when the hammer was impacting on the rail. And the deformation of the rail is also shown in the Figure 4.7. However, the variation of the deformation is hard to visualise correspond to the stress variation. Hence, in order to demonstrate the formation of the rail after the impact, the deformation is multiplied to allow it to be visible in the Figure 4.7. The formation is not the real condition of the rail. It is the exaggerated condition to demonstrate the displacement of the rail surface after impacting. Meanwhile, vibration was generated due to this impact. The vibration signal collected at 100 mm along the head of the rail is shown in Figure 4.8.

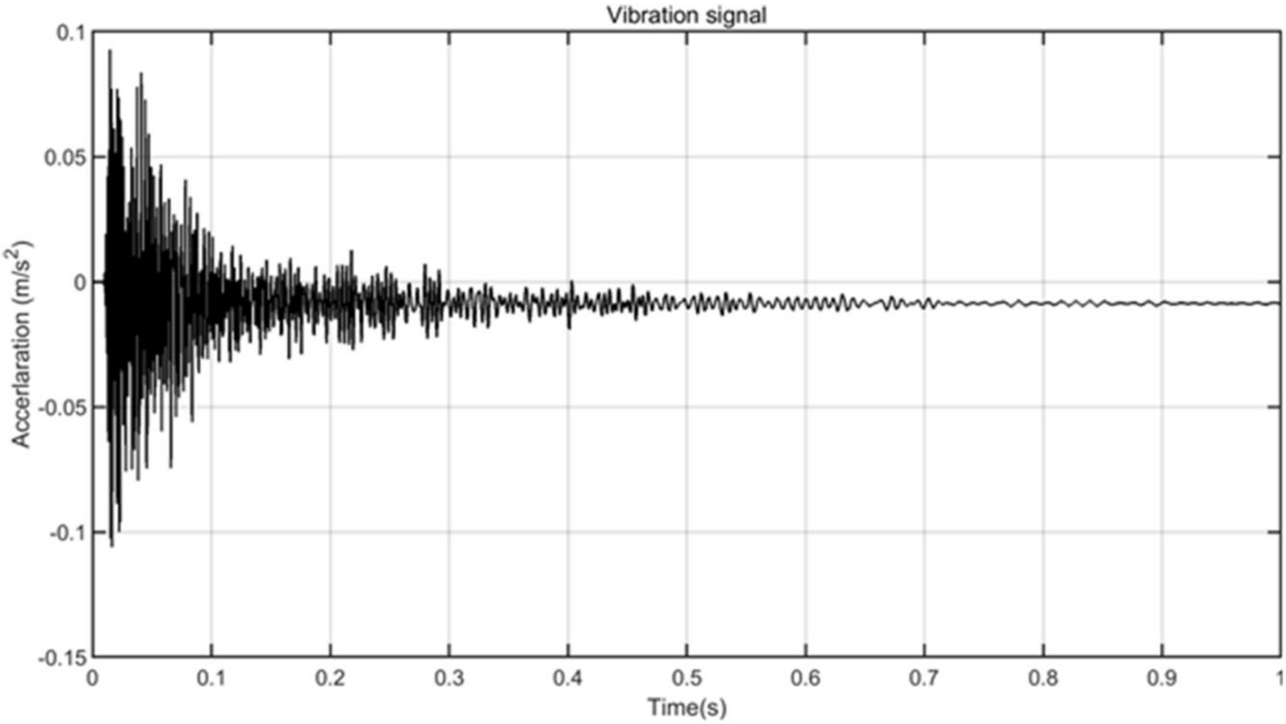


Figure 4.8 Simulated vibration signal

In order to validate the vibration model, the simulated vibration signal in Figure 4.8 was compared with a signal obtained from laboratory experiments, which will be introduced in the following section. Once the model was validated, the simulated signal could be used to simulate the acoustic signal, as described by the theory introduced in Section 3.1.

4.2.3 Acoustic model

The acoustic model was designed to propagate the sound generated from the vibration model. The theory of acoustic signal propagation in an acoustic field is introduced in Equation 3.16. As described in the methodology, the values of the vibration velocity and acoustic velocity are the same at the interface between the vibration and acoustic fields. Therefore, the vibration velocity can be used as the initial condition to initialise the acoustic model. Additionally, in order to reduce computational cost, a small acoustic field with boundaries designed to absorb sound from the rail is considered, and reflections are ignored due to the extremely large acoustic field. The parameters of the acoustic model are shown in Table 4.4.

Table 4.4 Parameters for initialising the vibration model

Parameter	Value
Speed of the sound (m/s)	340
Density of the sound (kg/m^3)	1.29
Acoustic field size (length × width × depth (mm^3))	2200 × 200 × 500
Acceleration (m/s^2)	Calculated by vibration model

The acoustic model created in COMSOL is shown in Figure 4.9.

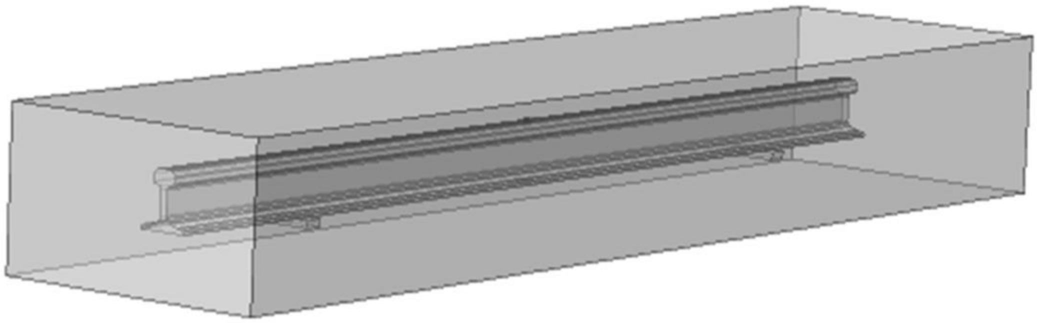


Figure 4.9 Acoustic model

4.2.4 Coupled multi-physical model

After building up the vibration model and the acoustic model, the next step is using the vibration (i.e., the acceleration) as the input of the acoustic model to obtain the acoustic signal based on Equation 3.17. It is notable that the input for the acoustic field in the coupled model is not the impulse signal. It should be vibration from the rail, which is different from the input for the independent simulation model (vibration/acoustic model). Hence, two different initial conditions should be set. The initial condition of the vibration field is an impulse input at the impact site, which is as same as the one in Section 4.2.2. And the initial condition of the acoustic field is the vibration source of the whole rail. The input position of the acoustic field was set as the whole of the rail surface.

The details of the coupled multi-physical model are given in Table 4.5.

Table 4.5 Parameters of the coupled multi-physical model

Physical field and simulation models		Parameters	Value
Vibration field	Material parameters	Young's modulus (*Pa*)	$2e^{12}$
		Poisson ratio (0.29)	44.5
		Density $(kg \cdot m^{-3})$	7870
		Dynamic impulse (*kg·m/s*)	160
	Initialise condition	Impulse time (s)	2.5e-3
Acoustic field		Speed of the sound (m/s)	340
		Density of the sound (kg/m^3)	1.29
		Acoustic field size (length × width × depth (mm^3))	3200 × 200 × 500
Parameters of the simulation model		Vertex elements	Parameters of the simulation model
		Edge elements	3384
		Boundary elements	9462
		Total elements	66792
		Calculation time (s)	0.2
		Sampling rate (Hz)	8000

4.3 Model validation

The preceding sections introduce the process of constructing the coupled multi-physical model. In this section, the model is validated through two aspects using laboratory tests. Firstly, the experimental target should be similar to the simulation model. To ensure that the rail in the simulation model was the same as the one in the experiment, a natural frequency test (FRF test) was conducted. The results could be validated against the simulated results in Section 4.2.1. Secondly, another experiment was designed to generate the vibration and acoustic impact response signals under the same conditions as the simulation model. The experimental signal was compared to the signal generated from the simulation model using statistical methods.

4.3.1 Model validation by FRF validation

In this section, an experiment (Lab experiment A) was designed to identify the natural frequency of the selected rail. The structural condition of the experimental is most similar to the simulation for the FRF calculation. A previous pre-experiment, completed with a 1 meter rail section, found significant variance of the natural frequency obtained from the rail caused by reflections from the end of the rail section. Hence, a longer rail needed to be used for this experiment. Further tests were carried out and it was established that a 3 meters rail section was sufficient to overcome the reflection problem. The experiment was set up as shown in Figure 4.10. The length of the whole rail is 3 meters. In order to evaluate the structural information of the rail through the FRF, 3 meters of rail was suspended by two plastic wires. This corresponds to the simulations used to evaluate FRF as described in Section 4.2.1. The distance between these two wires was 2 meters. National Instruments (NI) equipment was used for data acquisition. And accelerated sensor (Model 353B03) and microphone

(GRAS 46AE) was used for collecting the vibration signal and acoustic signal respectively. The acceleration sensor was fixed at the middle of the two wires. The impact hammer was used to generate the vibration from the rail. Meanwhile, the impact force was recorded. Additionally, the positions of the impact site were 100 mm, 200 mm and 300 mm from the acceleration sensor.

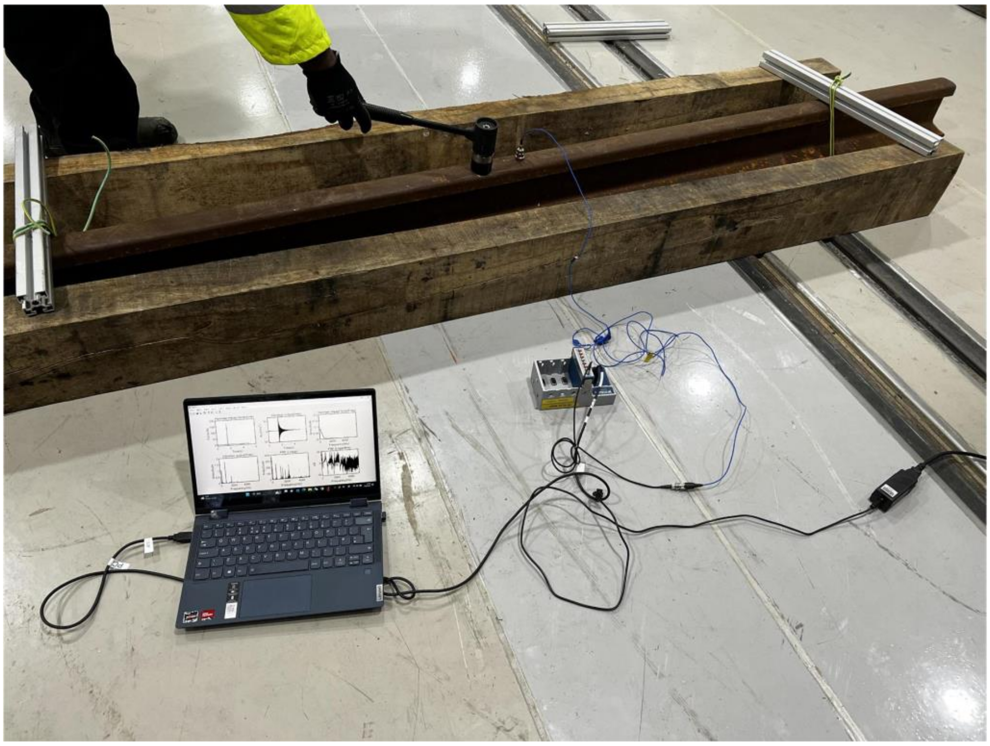


Figure 4.10 FRF test

The process of calculating the natural frequency is as follows. Firstly, the frequency of the vibration signal and impact force was computed by FFT, represented as $a(s)$ and $F(s)$, respectively. Based on the theory of the transfer function as shown in Equation 4.3, the natural frequency can be obtained by dividing the frequency of the vibration signal by impact force.

$$H(s) = \frac{a(s)}{F(s)} \qquad 4.3$$

The FRF experiment result is shown in Figure 4.11.

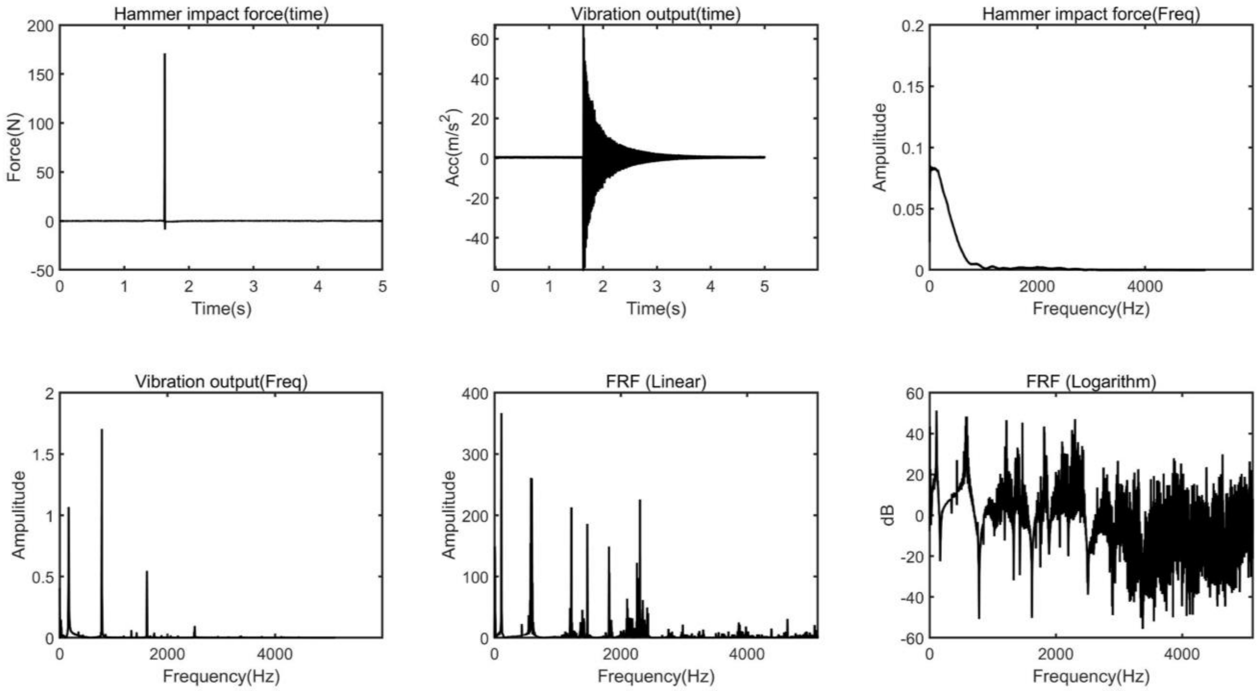


Figure 4.11 Results of FRF experiments

The six parts of Figure 4.11 represent the impact force by the hammer and vibration output, respectively in the time domain and frequency domains; and the natural frequency results are shown in the linear and logarithmic forms. An enlarged view of the impact force of the hammer in the time domain is shown in Figure 4.12. It can be seen that the span of the impulse time is 0.0025 s, which corresponds to the impulse that was given in the simulation model.

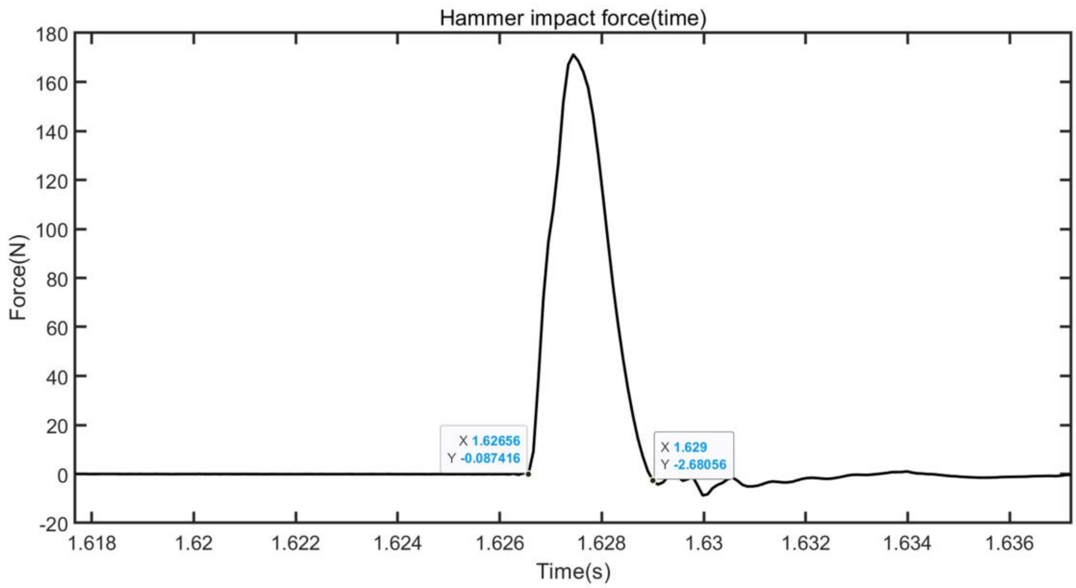


Figure 4.12 Impact force

The natural frequency can be identified in the final subfigure of Figure 4.11, which is enlarged in Figure 4.13. The first peak from whole frequency band represents the natural frequency.

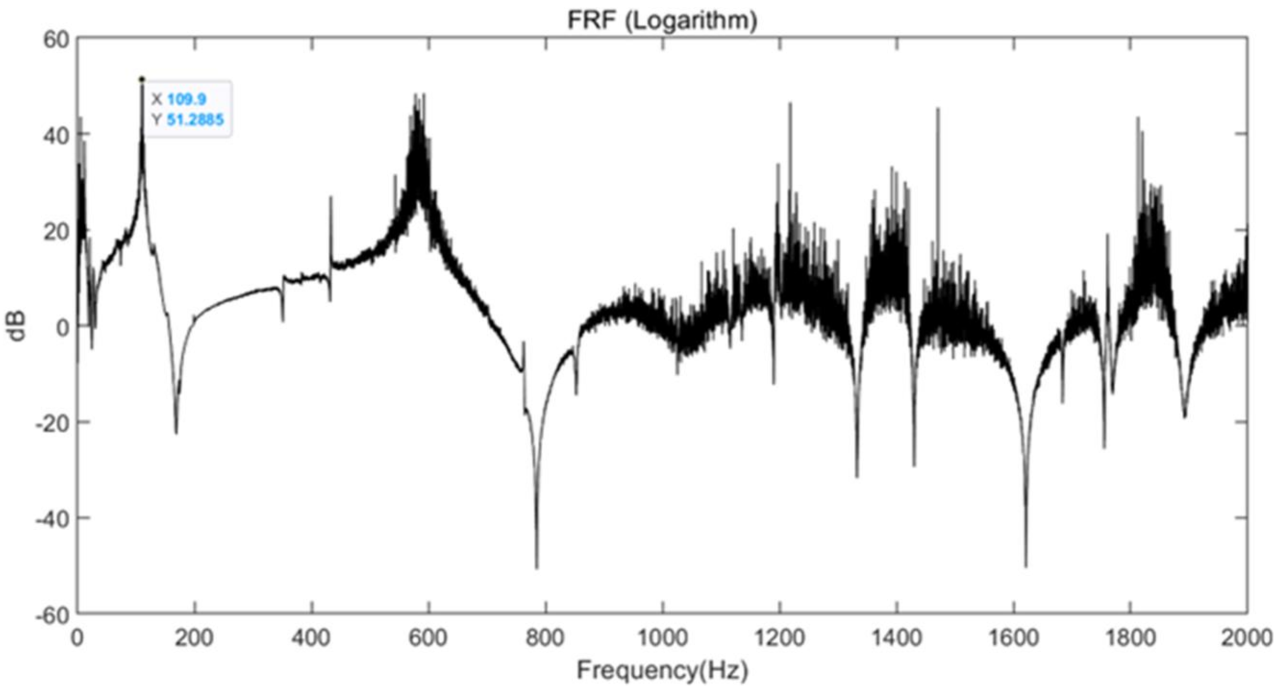


Figure 4.13 Natural frequency of the rail

If the rail is too short, its natural frequency will differ significantly from that of a real railway track due to reflections from the rail section ends. The frequency of an impact response signal, therefore, varies when the impact position changes relative to the end positions. Therefore, the length of the rail should be selected to ensure that it is long enough to accommodate an acceptable variation in frequency.

As mentioned in the laboratory experiment, a 3-metre rail was selected for the testing. The impact position was in the middle of the rail. An acceleration sensor was used at positions 100, 200, and 300 mm either side of the impact position. All natural frequency results from each side were analysed. A sketch of the laboratory experiment and impact position are given in Figure 4.14.

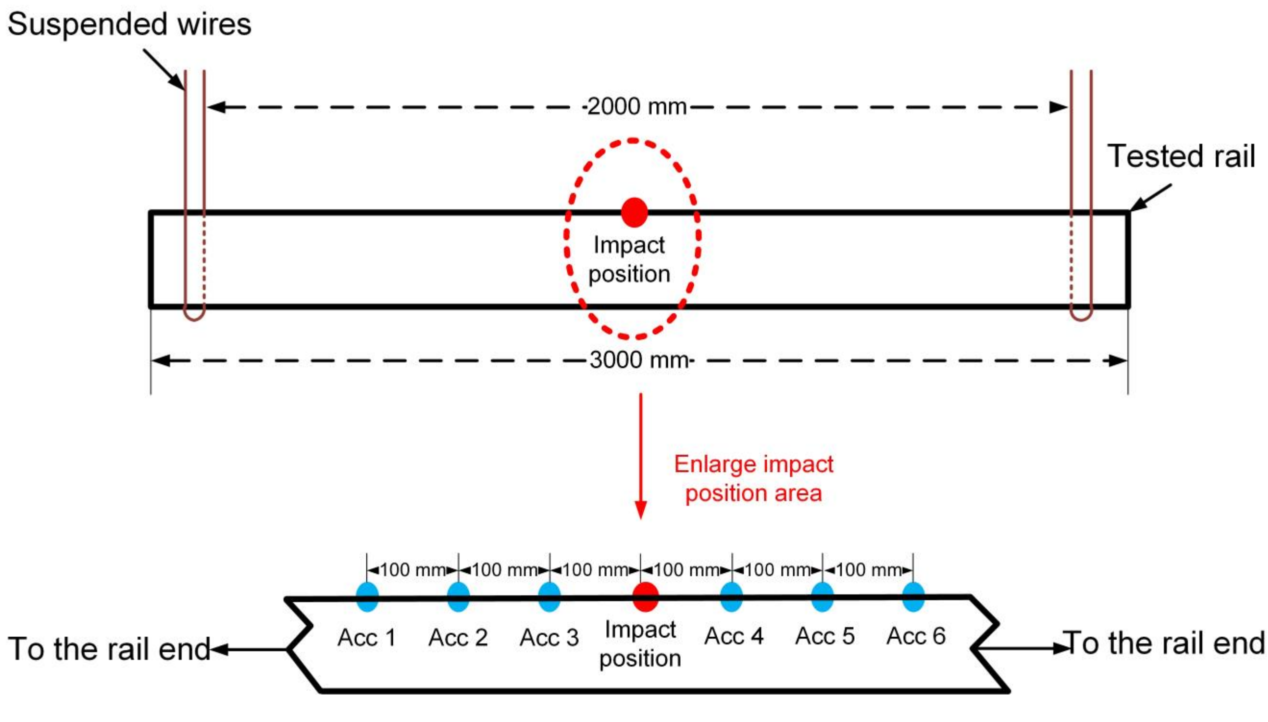


Figure 4.14 Impact position of the natural frequency experiment

The frequency range of the impact response signal generated from different positions was obtained as shown in Table 4.6.

Table 4.6 Natural frequency range obtained by impacting different positions

Position (left side)	Frequency range	Position (right side)	Frequency range
100 mm (Acc 3)	109.4–111.2 Hz	100 mm (Acc 4)	109.0–111.3 Hz
200 mm (Acc 2)	111.9–113.1 Hz	200 mm (Acc 5)	111.5–112.4 Hz
300 mm (Acc 1)	113.3–114.1 Hz	300 mm (Acc 6)	112.8–113.5 Hz

In Table 4.6, the results in the left two columns represent the variation in natural frequency when the rail position is impacted on the left side, while the right two columns represent the variation in natural frequency when the rail position is impacted on the right side. This allows direct comparisons of distances either side of the impact point. The frequency results in Table 4.6 demonstrate similar frequency ranges from both sides. The variance of natural frequency from both sides was lower than 5 Hz when the impact position changed from 100 mm to 300 mm (left: 4.7 Hz; right: 4.5 Hz). Additionally, the two suspension wires were moved further apart (2.5 meters), and the rail was tested again with the longer test section. It was found that the variance was still around 5 Hz. Hence, 5 Hz is an acceptable variance, and the length of the rail is appropriate for use in the research.

Comparing the results obtained from the experiment and simulation, the value of the natural frequency calculated from the simulation model (111.13 Hz) was in the frequency range of the tested rail, proving that the rail created in the simulation model was similar to the laboratory rail.

4.3.2 Model validation by comparison with experimental results

As mentioned in the previous section, the rail in the simulation model was related to the rail which was used in the experiment. After this validation, the simulation model also needs to be validated. Firstly, the simulated signal can be obtained as shown in Figure 4.15.

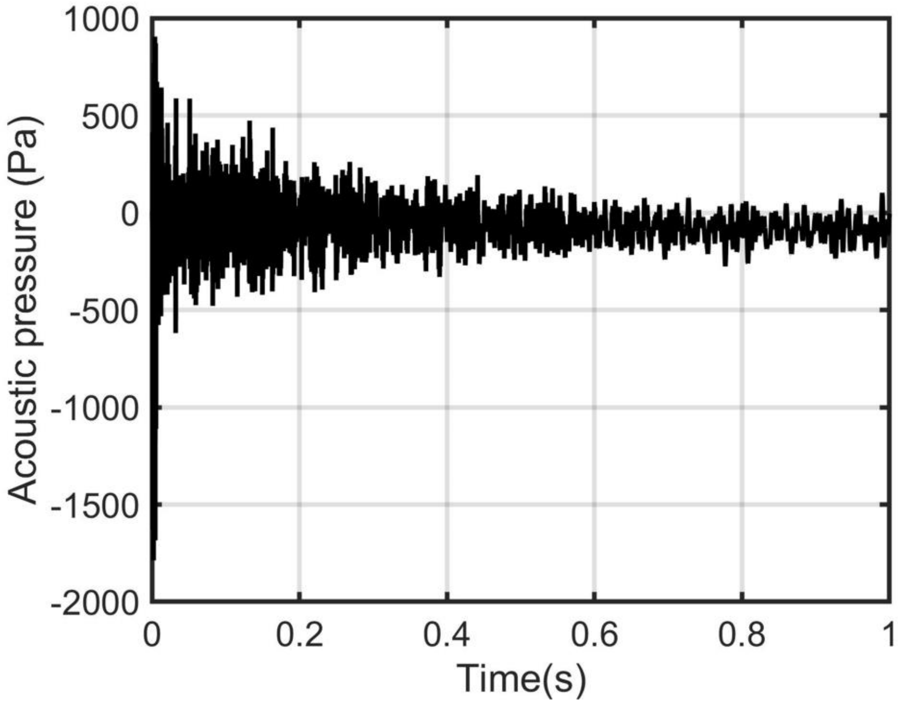


Figure 4.15 Simulated signal

The acoustic signals generated from rails with different faults were compared in terms of consistency with signals obtained from the simulation model and the experiment. The shape of the faults designed through the 3-D model was based on the faults made in the experiment. These faults were located on the underside of the rail. The faults made in the experiment are shown in Figure 4.16 and the size of the rail faults is given in Table 4.7. Approximations using square, rectangular, and circular fault shapes were designed due to it being easy to manufacture test samples for the corresponding experiments. Also using irregular faults in the simulation models would have dramatically increased the computational complexity and may potentially have introduced sources of error into the calculations. Hence, these three types of faults are designed for the analysis. Considering these three fault shapes, square and circular faults represent two types of corrosion existing with a small area. The rectangular fault represents corrosion with a large area, which was through the whole bottom of the bottom surface. And the three circular faults with different depths represent the different corrosion levels.

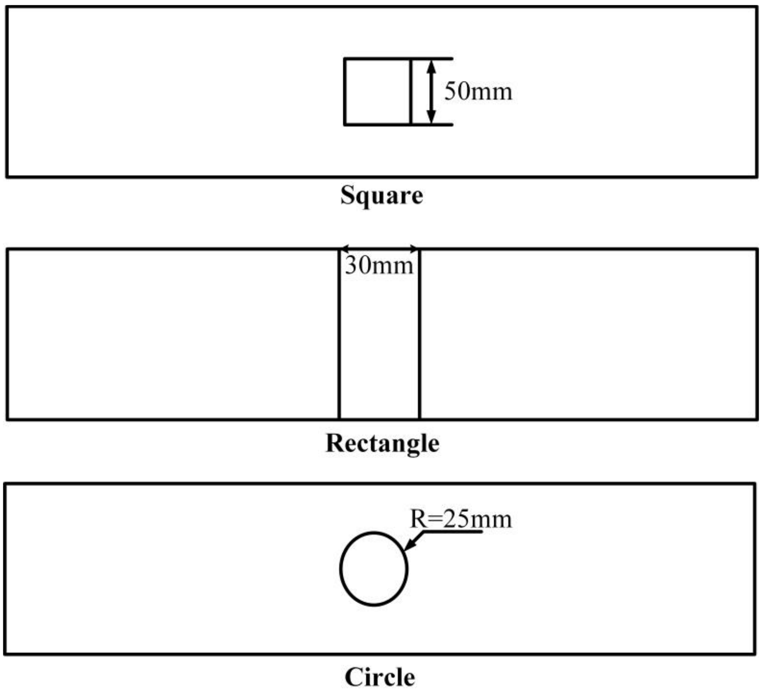


Figure 4.16 Shapes and sizes of faults

Table 4.7 Types and sizes of the faults

Fault type	Fault size	
Square fault	50 × 50 × 0.75 mm	
Rectangular fault	Through the whole bottom and with 0.75 depth	
		Circular fault
Circular fault	25 mm radius	0.75 mm depth
		3 mm depth

From the laboratory experiment, the experimental signals were generated as shown in Figure 4.17.

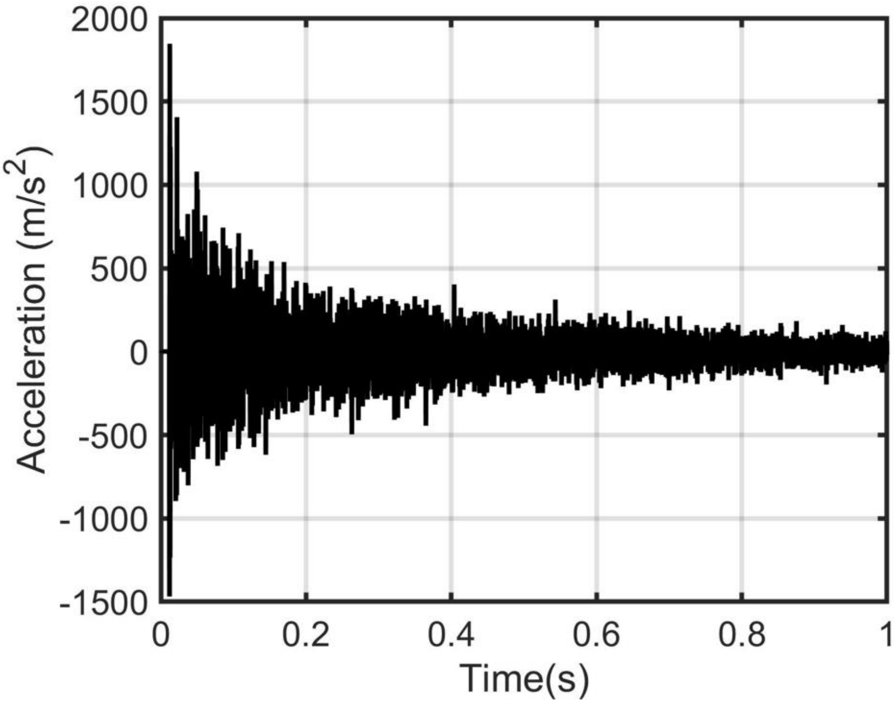


Figure 4.17 Experimental acoustic signal from a healthy rail

The output signal from the simulation model was validated for consistency. The simulated signal was compared to the signal from the experiment in both the time domain and the frequency domain.

Firstly, in the time domain, it was difficult to obtain similar signals from both the simulation and experiment due to the environmental noise. However, as both signals belonged to the impulse signal, their impulse energy trends could still be compared. Comparison of the two signals in the time domain is shown in Figure 4.18 and comparison of the two kinds of signals in the frequency domain is shown in Figure 4.19.

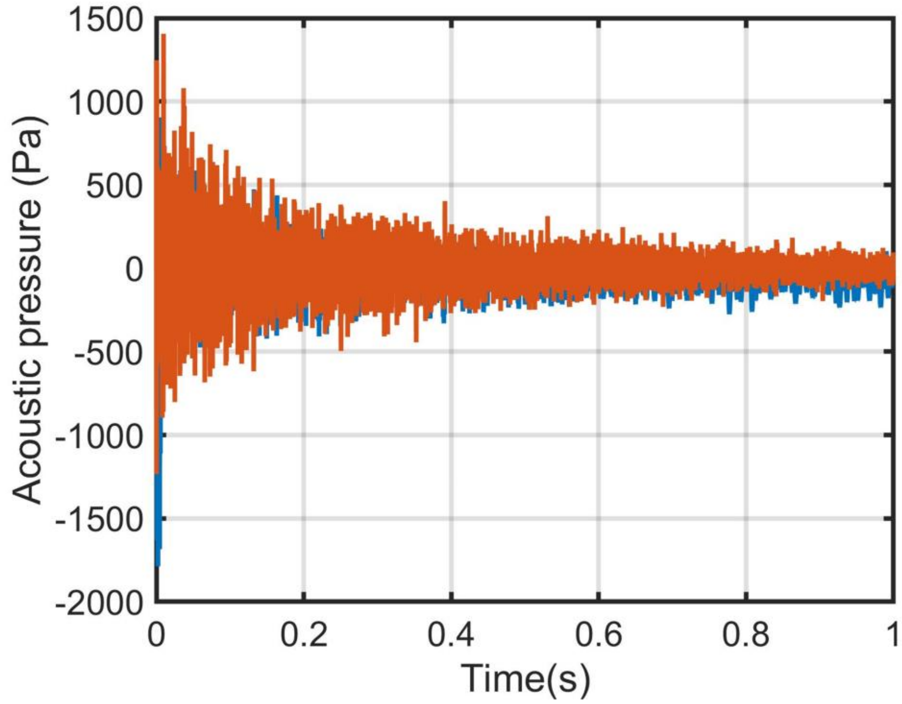


Figure 4.18 Comparison of the two kinds of signals in the time domain

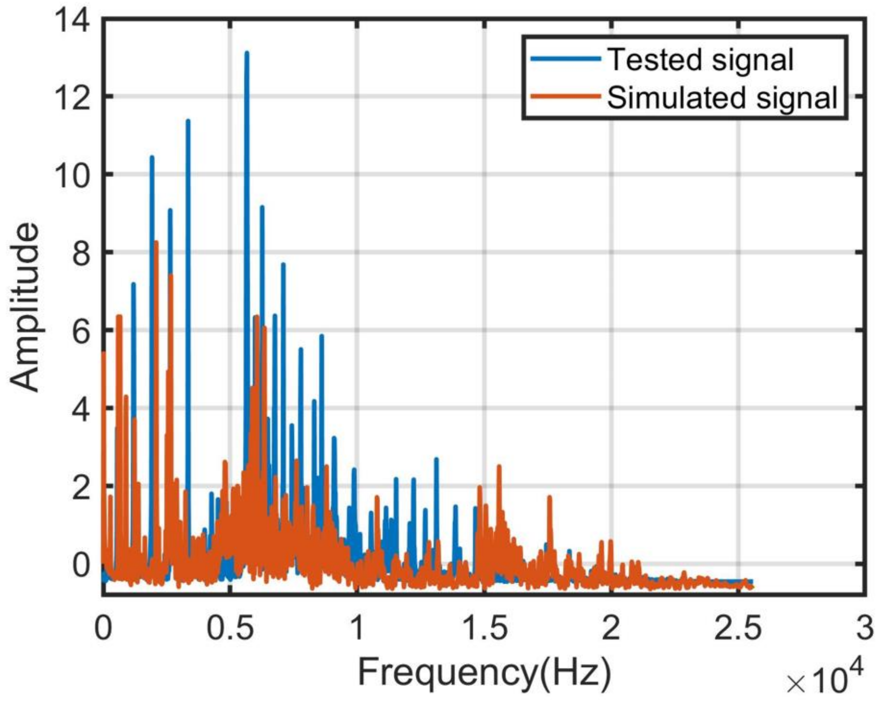


Figure 4.19 Comparison of the two kinds of signals in the frequency domain

Based on Figure 4.19, the frequencies obtained from the experimental signal and simulated signal are similar. The natural frequency results in Figure 4.13 indicate that the band of natural frequency of interest is between 0 and 5000 Hz. And Figure 4.19 shows that the frequency peaks of the two signals are very close in this range. However, it is not sufficient to evaluate the consistency of the two signals only by comparing their frequency peaks. To assess the consistency in the frequency domain, two methods can be used. Firstly, the variance of energy between the two signals should be consistent in the frequency domain. The increasing trend of frequency values should be similar to the frequency increases. To represent this trend, the energy of the signals in the frequency domain was summed, as shown in Equations 4.4 and 4.5; the trend of the increasing energy with frequency is shown in Figure 4.20.

$$e_i = \sum_{i=1}^{N} f(t) \qquad 4.4$$

$$E = [e_1, e_2, \dots, e_N] \qquad 4.5$$

where $f(t)$ is the signal in the frequency domain; N is the sampling frequency of the signal; e_i is the energy summing energies of the first i frequencies; and E is the set of the energies of e_i.

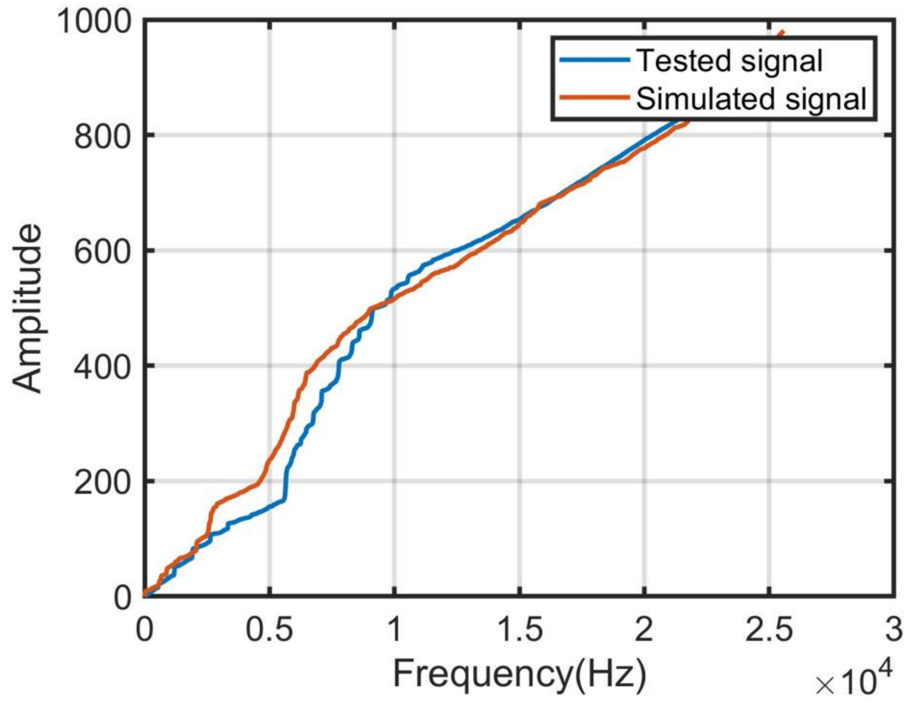


Figure 4.20 Trend of the energy with increasing frequency

The results from Figure 4.20 indicate that the trends of the two signals were similar. At frequencies lower than 2500 Hz, the trends of the two signals increased with a similar gradient. From 2500 to 9000 Hz (corresponding to the point where the two lines cross), the energy of the simulated signal was larger than that of the experimental signal. But the increased gradient in this region was similar. And from 9000 to 25,600 Hz, the energy of the simulated signal continued to be larger than that of the experimental signal. However, the gradients of the two lines also remained similar, indicating that the variance of energy was similar as the frequency increased.

Another way to assess the consistency of the two signals in the frequency domain is to use statistical indices. These indices provide a direct reflection of the similarity between the two signals. Some common statistical indices are similarity indices and correlation coefficients. The most common similarity indices include Euclidean distance [195], Manhattan distance [196], Chebyshev distance [197], Hamming distance [198], etc. However, these indices only provide a numerical value and there is no standard to determine the similarity of two signals. Common correlation coefficients include Pearson, Kendall, and Spearman. These three correlation coefficients have standards to measure the similarity of two signals. The values of these correlation coefficients range from −1 to 1. The closer a value is to −1 or 1, the more similar the two signals are. Conversely, the closer the value is to 0, the more dissimilar the two signals are. According to the standard definition [199], if the correlation coefficient is above 0.4, two signals can be considered strongly relevant. The three correlation coefficients of the two signals obtained from different rail conditions are provided in Table 4.8.

Table 4.8 Correlation coefficients comparing simulated and experimental signals

Rail condition	Pearson	Kendall	Spearman
Health	0.436	0.516	0.695
Square fault	0.470	0.638	0.628
Rectangular fault	0.530	0.587	0.710

Circular fault	0.461	0.512	0.674

The results in Table 4.8 indicate that the simulated signals and experimental signals obtained for different rail conditions are strongly relevant. The lowest value from the table was over 0.4.

Hence, from the discussion of the consistency between the simulated and experiment signals, it was concluded that the simulated signal could be used for further analysis. Additionally, the simulation model could be applied for creating a promising simulation dataset.

From the simulation carried out, the simulated signals were generated from different rail geometries. However, only a single signal could be generated from each rail geometry. Hence, in order to generate sufficient signals from the same rail geometry (healthy or different levels of faults), the properties of the rail were changed within an acceptable range to create more rails with the same conditions (healthy or faulty). Considering the three initial parameters of the rail, the Poisson ratio and density cannot be changed with the material loss. Hence, only the Young's modulus can be adjusted to obtain a set of simulated signals. The Young's modulus (E_S) of the rail was changed in a range between 99.5% E_S to 100% E_S. After the simulation, a dataset containing six categories of simulated signal was created. The simulated categories were as same as the sample types of the experiment as shown in Table 4.7. Each category contains 300 signals, so there are 1800 simulated signals in total in the simulated dataset.

4.4 Conclusion

In conclusion, this chapter discusses the creation of a simulation model to obtain acoustic impact response signals. Firstly, the possibility of generating simulated impact response signals through a physical theory (SDOF system) was explored and it was found that the obtained signals could not be similar to real signals. Secondly, an FEM simulation model was designed to create a simulated signal that matched the signal from the laboratory test. The process of building the FEM model was introduced, including the FRF model, vibration model, acoustic model, and coupled multi-physical model. The simulated acoustic signal was obtained through the FEM model. Thirdly, the validation process of the simulation model involved comparing the FRF from the simulation model with the one from the real rail in the laboratory experiments and comparing the simulated signal with the experimental signal in the frequency domain. The similarity between the two signals was evaluated based on the similarity of the frequency peaks, variance of energy, and correlation coefficients.

Once the accuracy of the simulation model was validated, it was used to generate simulated impact acoustic response signals for different rail geometries representing different rail fault types. The simulated signals were created by applying the impact force at different rail conditions and collecting the corresponding acoustic response signals. These simulated signals were then used to create a simulated dataset for further analysis and machine learning applications.

In the next chapter, the impact response signal will be processed by different signal processing methods and the features of the impact response signal will be extracted for further analysis through an appropriate method. Then, by using the algorithm of an advance signal processing method on the impact response signal, the dataset of simulated signals is processed for fault identification.

CHAPTER 5 SIGNAL DECOMPOSITION AND FEATURE EXTRACTION

This chapter introduces a method to process the impact response signal. The obtained impact signal contains the noise generated from environmental interference such as that caused by a change of impact location or force. The signal noise masks the characteristic pattern of the signals, which contains information about the railway track structure and health. In this chapter, a laboratory experiment is introduced to collect the vibration and acoustic impact response signals. A control experiment was applied to prove the feasibility of the laboratory experiment. Then the signal obtained from the laboratory experiment was processed by different signal decomposition methods (EMD, EEMD, VMD, and LMD). The most suitable decomposition method was selected based on analysis of the energy-based results obtained by these four decomposition methods. Furthermore, a proposed algorithm is introduced for optimising the most suitable method. The parameter for initialising this method, which previously had to be selected manually, can now be selected automatically after the optimisation. And based on the proposed method, the most obvious characteristic patterns can be extracted and environmental interference, such as that caused by a change of impact location or force, can be eliminated. Finally, a pruning algorithm was used to automatically identify and remove the data obtained within the experiments that fell outside error and thus could not be used.

The contents of Section 5.2 have been reviewed and published in the journal *IET Signal Processing* (‘Decomposition methods for impact-based fault detection algorithms in railway inspection applications’). A discussion of the comparison of different signal decomposition methods was completed. From this publication, it can be understood which decomposition method is the most effective for processing the impact response signal.

Additionally, the contents of Section 5.3 have been reviewed and published in the journal *Applied Science* ('An improved VMD method for use with acoustic impact response signals to detect corrosion at the underside of railway tracks'). Based on the discussion in this paper, the effectiveness of the improvement of the impact response signal by the selected decomposition method was proved. And from this paper, it was indicated that the different types of corrosion on the underside of the rail can be identified.

5.1 Laboratory experiment for obtaining impact response signals

In Chapter 4, the simulated dataset was obtained. The similarity between the simulated and experimental signals has been validated in Section 4.3. The next step is to explore how to process the impact response signal. Firstly, like the simulated dataset, another dataset containing a sufficient experimental signal is obtained for the analysis. In order to identify the different rail faults, three kinds of fault types are considered. And in order to identify the different rail corrosion levels, three faults with different depths are designed. The geometry of the faults matches the shapes of the simulated faults shown in Figure 4.16, which supports validation. The detailed shapes and sizes of the rail corrosion are given in Table 5.1 and a picture of different corrosion types from the experiment is shown in Figure 5.1.

Figure 5.1 Shapes of the rail corrosion

Table 5.1 Shapes and sizes of rail corrosion

Fault type	Fault size	
Square corrosion	50 mm × 50 mm × 0.75 mm	
Rectangular corrosion	Through the whole bottom of the surface and with a depth of 0.75 mm	
		0.25 mm depth
Circular corrosion	25 mm radius	0.75 mm depth
		3 mm depth

These four types of rails will be used for fault identification. Before the identification, in order to understand the feasibility of the experiment, a control experiment comparing the results from two healthy rails was completed. In the impact response signal experiment, the vibration signals and acoustic signals from two healthy rails were obtained separately, as shown in Figure 5.2 and Figure 5.3.

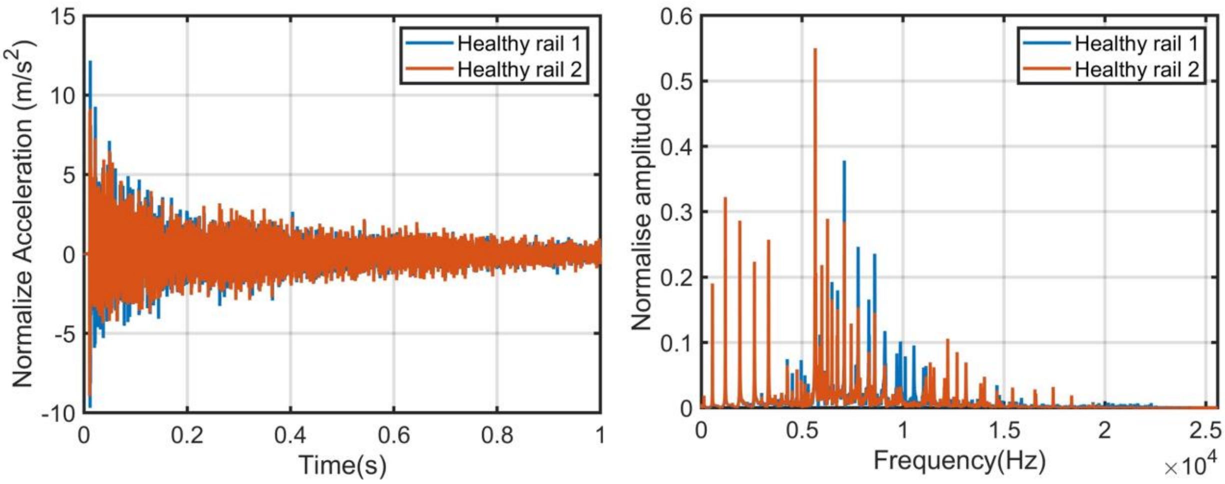


Figure 5.2 Vibration results for two healthy rails

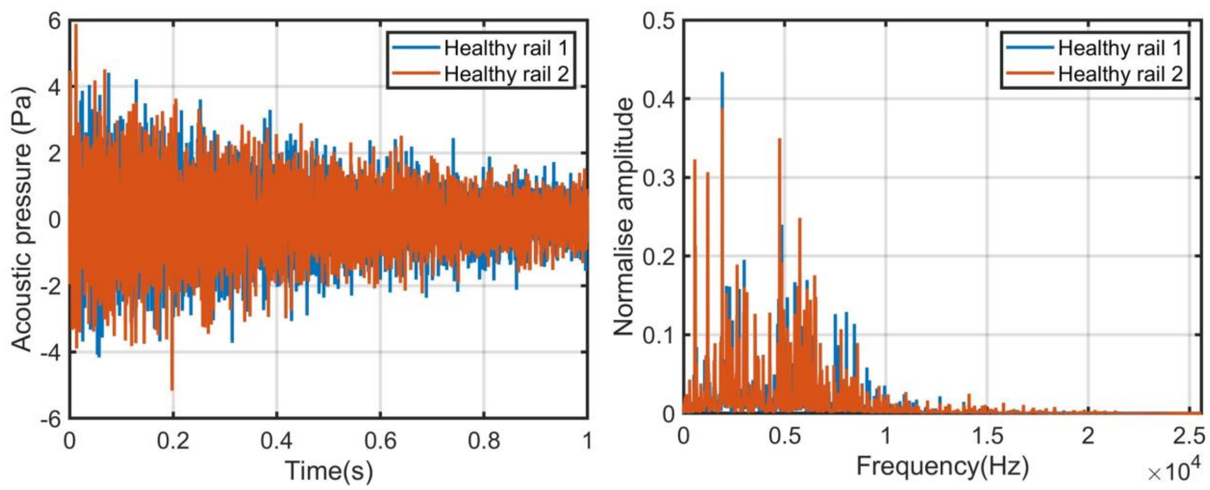


Figure 5.3 Acoustic results for two healthy rails

Figure 5.2 and Figure 5.3 indicate that the signals from the two rails were similar. In particular, the frequencies at which the peak points occurred were the same. The correlation coefficients introduced in Chapter 4 showed that the three coefficients obtained from the frequency results were all over 0.7, validating the similarity of the two signals. This proved that the subsequent experiment would not be impacted by the environment or other unpredictable factors. Hence, the signal obtained from the laboratory experiment could be used for further analysis. In the next section, decomposition methods are introduced for processing this impact response signal. The effectiveness of different decomposition methods is evaluated.

5.2 Decomposition methods for processing the impact response signal

The signals obtained from laboratory experiments will be used for the analysis of feature identification. Then the features can be used for classifying the different conditions of the rails. Firstly, the signal obtained from the laboratory experiments should be processed through a suitable signal processing method. In this section, different signal processing methods are used to analyse the effectiveness of extracting signal features.

Signal decomposition techniques are widely used in processing many different types of signals. They are particularly common in applications such as rotational machine monitoring (roller bearings) [111],[112] and medical examination (pulse, electroencephalogram, electrocardiogram, etc.) [96],[97],[98]. Unlike traditional Fast Fourier Transform (FFT) and wavelet transform methods, mode decomposition can identify and categorise minor faults more accurately [99]. As introduced in Chapter 2, common mode decomposition techniques include Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), Local Mean Decomposition (LMD) and Variational Mode Decomposition (VMD).

These four decomposition methods have individual characteristics and may be appropriate for detecting different signals. However, there is limited research focusing on direct comparison, particularly in impact response signal processing or in a railway environment. Additionally, little research has been conducted on direct comparisons, particularly in the context of railway impact response signal processing. Therefore, there are valuable research gaps that need to be explored. Firstly, there is a lack of research on railway impact site inspections, particularly in relation to impact response signal analysis. Secondly, while there have been comparative studies on various decomposition methods, none have focused on evaluating their effectiveness in processing impact response signals. Hence, there is a need for further research to explore the application of impact response signal analysis in railway impact site inspections, while evaluating the suitability of different decomposition methods.

In this section, comparison experiments were used to explore the effect of defects in the rail structure on the observed vibration signals. And based on the methodology in Chapter 3, the signal characteristics obtained from the experiments using the different decomposition methods are presented and energy-based methods used to evaluate the different approaches. Standard deviation was used to assess the similarity of parameter values from different Intrinsic Mode Functions (IMFs), while maximum and minimum differences from the mean value were used to consider the effectiveness of each technique in identifying specific damage types. The evaluation showed that VMD is a strong candidate for the rail application due to the similarity and clarity of results. A flow chart describing the methods used in this chapter is presented in Figure 5.4.

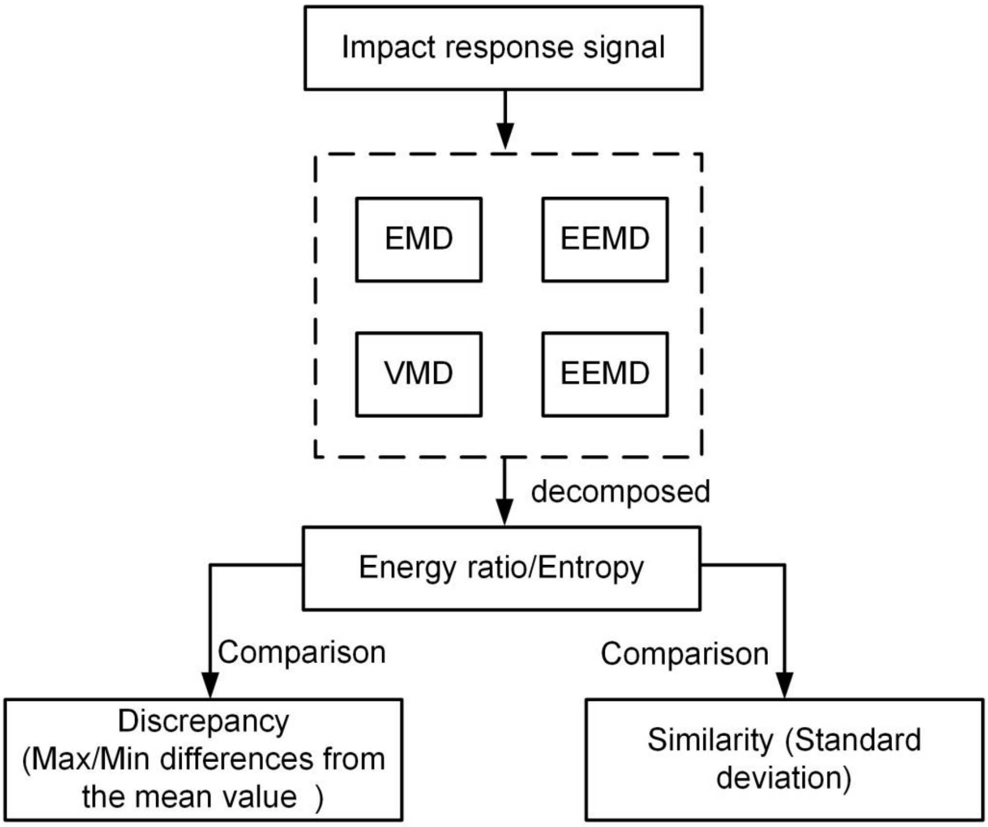


Figure 5.4 Flow chart of the proposed method

5.2.1 Application of methods

Based on the laboratory experiments, the vibration signal and acoustic signal can be obtained. The vibration signal contains less noise from the experiments. Hence, the vibration signal was used to investigate the effectiveness of the decomposition methods for the impact response signal. The normalised vibration signals (normalised using the MATLAB normalisation function) in the time and frequency domains are shown in Figure 5.5. Considering the impact response signal in the frequency domain, the majority of the energy was in the range of 0 kHz to 20 kHz; and the frequency peaks were present from 0 to 3.5 kHz, and from 5 to 9 kHz.

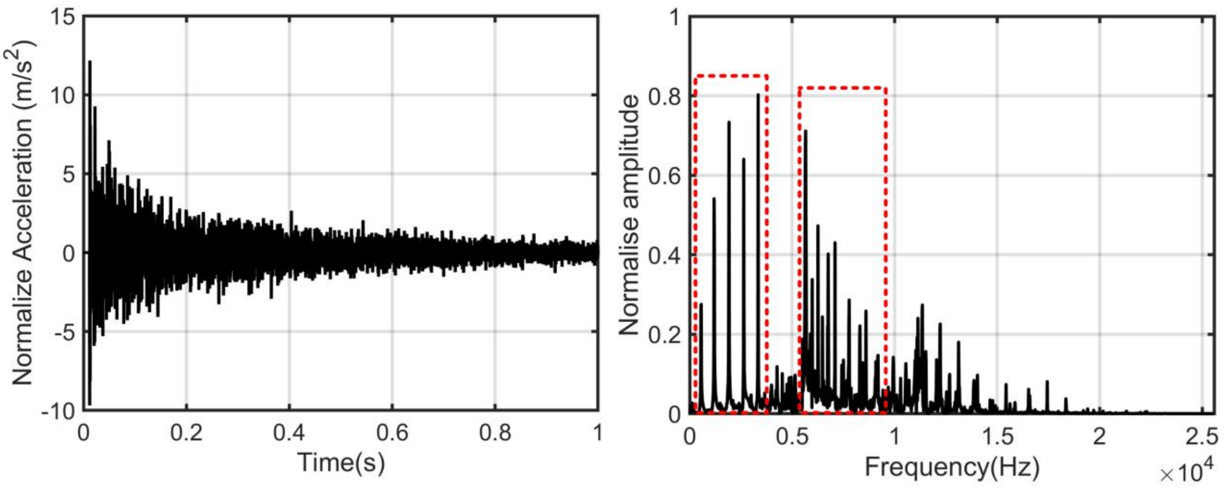


Figure 5.5 Frequencies of interest for healthy and faulty rail sections

During the simulated foot corrosion measurement, the impact response signal was processed by the four decomposition methods. Figure 5.6 illustrates the first six IMFs that were selected based on the differences between the IMFs and the original signal. In order to support the comparison, only significant IMFs were considered further across all the decomposition methods. It was observed that IMFs with a notable signal component often displayed visible peaks at specific frequencies. However, this was not always the case.

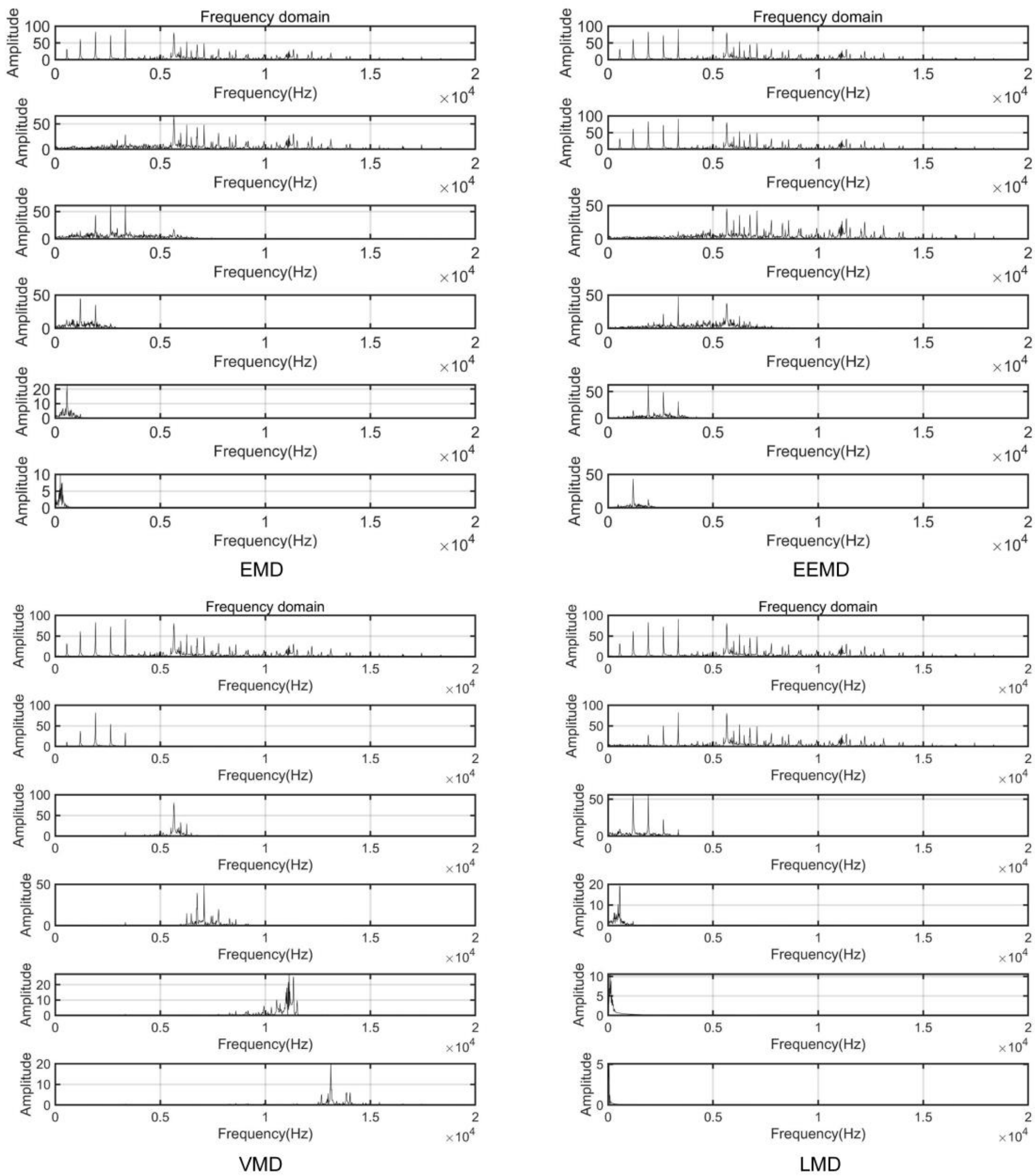


Figure 5.6 Decomposition results for different methods

The four decomposition methods were applied to the impact response signal and a discussion of their performance is provided below. The results were processed using MATLAB 2022a on an Intel® Core^{TM} i5-9500T CPU @ 2.20 GHz. Among the four decomposition methods, the EEMD approach was the most computationally intensive, with EMD, VMD, and LMD taking 9.5%, 43.1%, and 58.7% of the EEMD duration, respectively.

Based on the comparison experiment results illustrated in Figure 5.6, it was observed that the first six IMFs produced by the different decomposition methods varied. EMD, EEMD, and LMD preserved most of the information in one IMF (IMF1 for EMD, EEMD, and LMD) while VMD distributed features across several IMFs and a broader frequency range. Within the 0 to 5 kHz frequency range associated with fault features, EEMD and LMD contained information in a single IMF (IMF1 in EEMD, IMF2 in LMD), EMD had the majority of information in two IMFs (IMF3 and IMF4), and VMD dispersed energy across four IMFs. Additionally, mode mixing was more pronounced in EMD, EEMD, and LMD (IMF5 and IMF6 in EEMD and EEMD; IMF 4, IMF5, and IMF6 in LMD) than in VMD. These findings highlight the variation in performance of the different decomposition methods. To further analyse the effectiveness of these methods, specific post-processing techniques are employed in the next section to quantify the performance variation.

5.2.2 Energy-based analysis for post-processing

In this section, post-processing techniques are used to quantify the performance variation. Energy-based analysis is one of the main techniques used in the post-processing of IMF components. The energy ratio was employed to evaluate the effectiveness of the four decomposition techniques. Additionally, energy entropy was employed to validate the accuracy of the result analysed using the energy ratio approach.

Based on Equation 3.6, the IMF energy distribution is not dependent on excitation intensity. If the physical structure changes, the energy distribution of the IMF components would also change. The energy ratio and entropy are energy-based techniques, the output of which can be evaluated in order to identify changes in the IMF components corresponding to damage in the physical structure.

The energy ratio was first proposed by Wan and Peng [200] and has been widely applied for the inspection of faulty machine components [201]. If $x(t)$ is the impact response signal, the selected IMF components calculated by decomposition are given as $IMF_i(t)$. The energy of each $IMF_i(t)$ is given as in Equation 5.1.

$$E_i = \sum |IMF_i(t)|^2 \; (i = 1,2 \ldots M) \tag{5.1}$$

where *M* is the number of IMFs. The total energy *E* of *x(t)* can be identified through the summation of all $IMF_i(t)$ [202], as shown in Equation 5.2.

$$E = \sum_{i=1}^{M} E_i \tag{5.2}$$

Then, the energy ratio can be obtained as shown in Equation 5.3.

$$p_i = E_i / E \tag{5.3}$$

The result for each IMF was extracted from the impact response signal using Equations 5.3. The results indicated a variation in the energy ratio of some of the IMFs associated with the rail condition (healthy or faulty types). As an example, the averaged energy ratios obtained for application of the VMD method to the 300 measurements of each of the four rail conditions are presented in Figure 5.7.

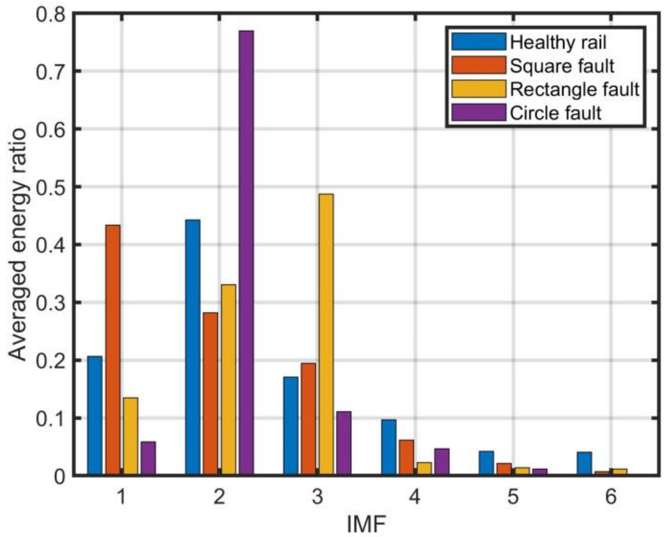


Figure 5.7 Energy ratio of impact response signals using the VMD method

The Standard Deviation (SD) of the mean values of the energy ratio results for each IMF is obtained for each rail condition. This indicates how uniform the distribution of energy over the IMFs is. The range of energy ratios across the different rail conditions is also identified for each IMF. It illustrates the IMF's suitability for differentiating rail conditions.

The mean value of the SD (MSD) of each rail condition can be calculated as per Equation 5.4, with the results presented in Figure 5.8.

$$\sigma_{ij} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(p_{ij} - \mu\right)^2}$$

$$MSD_i = \sum_{j=1}^{N}\sigma_{ij} \tag{5.4}$$

where σ_{ij} is the SD of the energy ratio corresponding to a type of damage in a single IMF; and *N* is the number of types of damage.

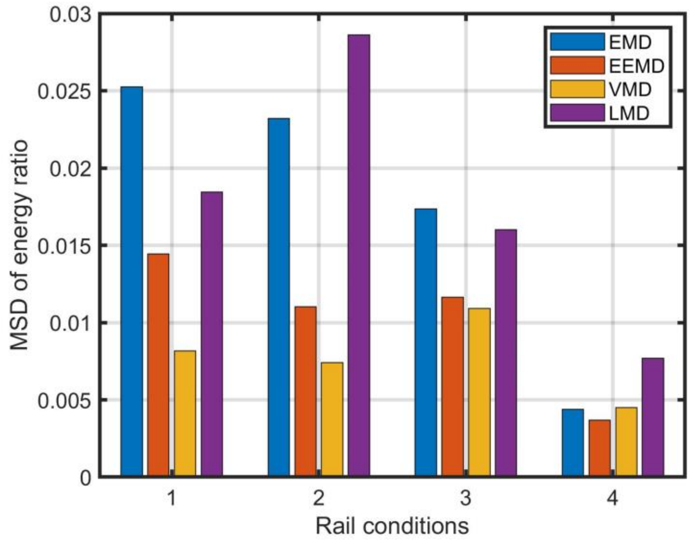


Figure 5.8 MSD of each damage type (energy ratio)

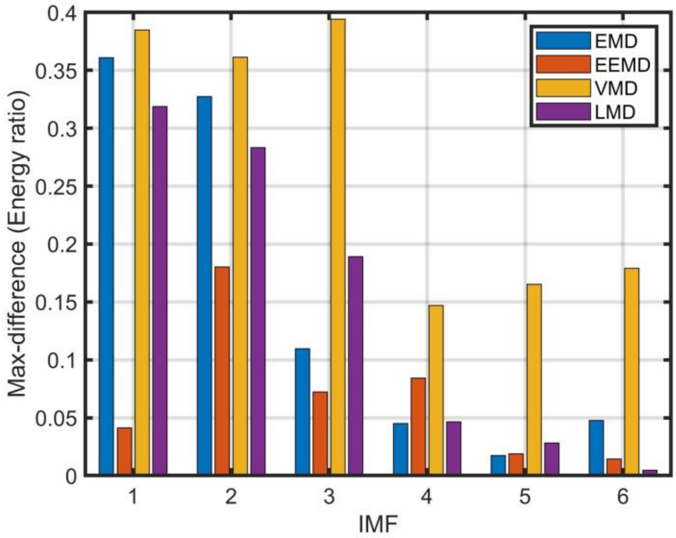
The MSD was calculated to reflect the uniformity of the energy distribution over the IMFs. If the MSD was lower, it indicated that the energy distribution was more uniform. In Figure 5.8, the lowest value was obtained from the use of VMD and the highest value was from LMD or EMD. This showed that the VMD method had the greatest level of uniformity among the methods considered.

The maximum and minimum differences from the mean values were used to define the index of the discrepancy between different damage types in each IMF. These two parameters, given in Equation 5.5, reflect the difference in the results obtained for different rail conditions.

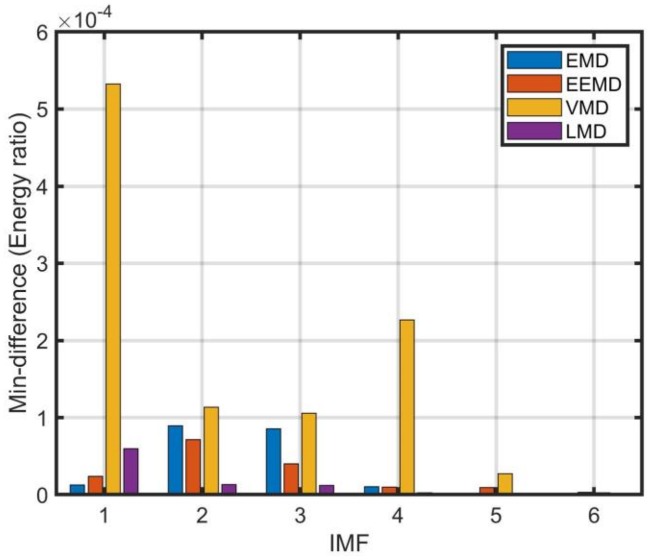
$$d_{max} = \left|\overline{D_i} - \overline{C_j}\right|, d_{min} = min\left|\overline{D_i} - \overline{C_j}\right| \quad 5.5$$

where $\overline{D_i}$ is the mean value of $IMF_i(t)$ and $\overline{C_j}$ (*j* is the rail condition) is the value of the energy ratio obtained for different conditions in the same IMF. The comparison of d_{max} and

d_{min} between each IMF decomposed from the impact response signal is illustrated in Figure 5.9.



(a)



(b)

Figure 5.9 Differences from the mean value of the energy ratio, (a) maximum, (b) minimum

Considering the maximum deviations from the mean values for each IMF (as shown in Figure 5.9a), the highest values were all obtained using the VMD method. Conversely, the lowest values in almost all IMFs were obtained using the EEMD method (IMF1, IMF2, and IMF3). The minimum deviations showed a similar pattern, with the largest differences corresponding to the VMD method.

Based on the energy ratio approach, the results shown for MSD and the difference from the mean approaches, VMD has been shown to have the greatest effectiveness in the decomposition of impact response signals over a range of rail conditions.

Further to the energy ratio evaluation, the use of energy entropy as an alternative post-processing method was applied to compare the different decomposition methods.

The energy entropy is an extension of entropy in the energy domain which is given in Equation 5.6 [203].

$$R_i = \int_{-\infty}^{+\infty} |IMF_i(t)^2| \, dt$$

$$P_i = -\sum_{i=1}^{M} R_i \log R_i \qquad 5.6$$

The entropy result, as per the energy ratio analysis, was analysed using MSD as well as the maximum and minimum difference from the mean value approach described above. The MSD results are presented in Figure 5.10.

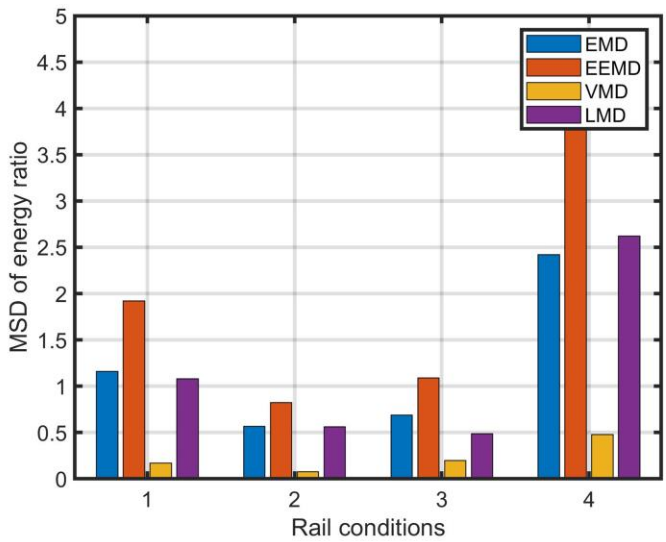
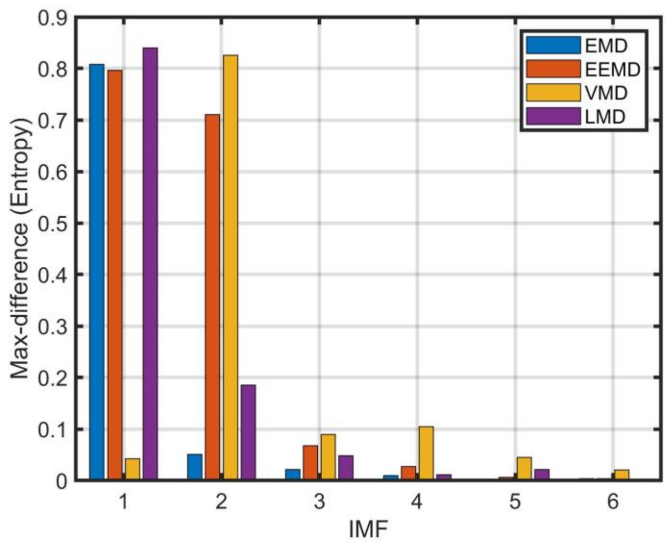
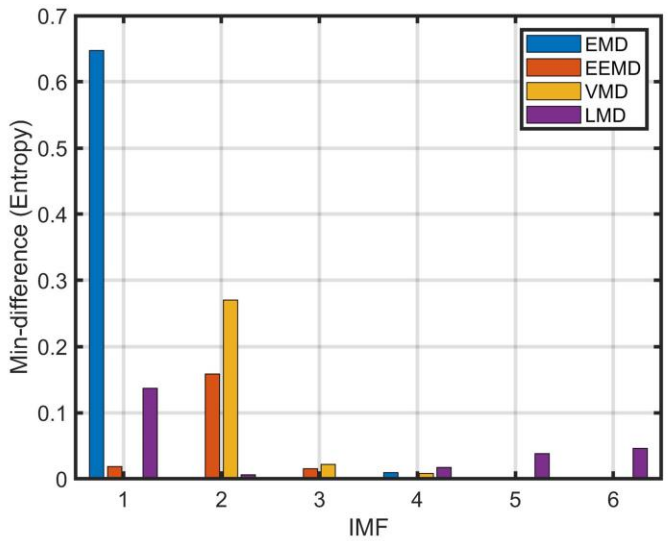


Figure 5.10 MSD for each damage type (energy entropy)

The results of the maximum and minimum difference from mean value processes are presented in Figure 5.10.



(a)



(b)

Figure 5.11 Difference from mean value of energy entropy, (a) maximum, (b) minimum

Considering Figure 5.10, most of the lowest values were generated using the VMD method, and the highest values in all IMFs were obtained using EEMD. This result is similar to the one obtained using energy ratio-based methods.

Additionally, considering Figure 5.11, most of the highest values for the max-difference from the mean value arose in the VMD results (IMF2, IMF3, IMF4, IMF5, and IMF6). Likewise, most of the highest values for the min-difference from mean value results were obtained using the VMD method (IMF2, IMF3, and IMF4). This result is also the same as the conclusion from the energy ratio.

Considering the analysis, both the decomposition performance (evaluated using MSD and the min/max differences from mean value) and computational complexity were considered. EMD was identified as having the advantage of minimal computation time. However, EMD performed comparably poorly in terms of effectiveness, whereas VMD demonstrated particularly strong performance and provided a uniform distribution of signals from the same category. VMD also provided the greatest variation in signals associated with rail condition. In this study (and most like it in the literature), the effectiveness of the technique was considered to be of greater significance than the computational complexity. It was, therefore, considered that VMD was the preferred method for the decomposition of impact response signals.

In conclusion, based on the results from both the energy ratio and energy entropy evaluations, it was found that the VMD method is the most effective for decomposition of impact response signals over a range of rail conditions. However, the effectiveness of VMD in classifying different rail fault types still needed to be explored. Hence, in the next section, based on VMD, an improved algorithm is designed to enhance the effectiveness of VMD on the impact response signal.

5.3 An improved VMD method for using impact response signals

Based on the conclusion in Section 5.2, the VMD method is the most effective for the decomposition of impact response signals over a range of rail conditions. Hence, VMD will be used for processing the impact response signals. Previous research indicates that the number of decomposition modes (K-value) is a significant parameter in the performance of VMD [118]. Current research in the selection of the optimal K-value for VMD generally focuses on two aspects. Firstly, the relationship between the original signal and sub-signals is known as the IMF. Yang et al. [204] applied kurtosis to develop an indicator of the effect of the relationship on the performance of the VMD algorithm. The improved VMD was then used to detect the early failure of bearings through chatter signals.

The second popular approach for K-value selection is to consider the Initial Centre Frequency (ICF) parameter. This approach is suitable for periodic signals for which it is easy to calculate a centre frequency [115],[205]. Jiang et al. proposed an ICF-guided VMD algorithm for extracting minor damage information from bearings; however, the impact response signals are different from the rotational signals considered in the current research [206]. Impact response signals are impulse signals with gradually decreasing energy, which is different from rotational signals which contain a periodic peak in the time domain. Hence, kurtosis-based methods are generally not considered suitable because of their dependence on the distribution of peaks within the signal. Additionally, as impact response signals are non-periodic signals, it is generally difficult to calculate their centre frequency. Hence, current techniques for determining K-value are not suitable for selecting the optimal K-value for impact response applications. Therefore, it is necessary to investigate an algorithm to select the most suitable decomposition modes for VMD when it is applied to impact response signals.

In this section, based on the impact response signal, a method for optimising VMD is introduced to extract the most obvious characteristic patterns and eliminate environmental interference such as that caused by a change of impact location or force. The optimised algorithm combines VMD, Singular Value Decomposition (SVD), and Shannon entropy techniques. The proposed improved VMD algorithm (IVMD-SVD) method, through its SVD component, automatically selects the initial parameters for the VMD element. SVD is then further applied to the VMD result to obtain Shannon entropy, which is used in the detection and classification of underside corrosion. A flow chart describing the structure of the proposed algorithm is provided in Figure 5.12.

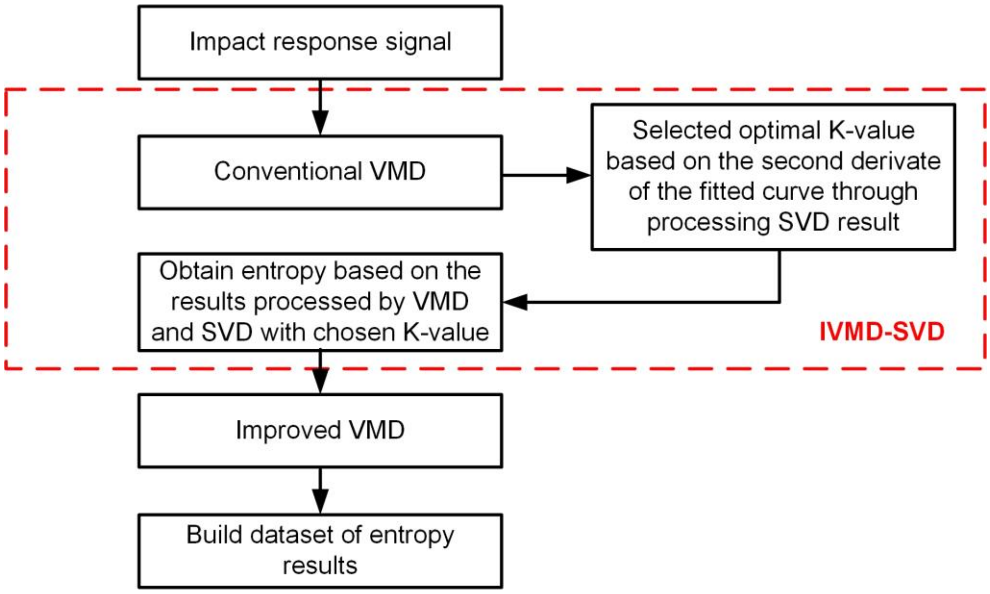


Figure 5.12 Flow chart for processing the impact response signal through the proposed method

5.3.1 Introduction of the IVMD-SVD with detailed signal

The principle of VMD is illustrated in Chapter 3. A brief introduction to SVD and Shannon entropy are given below.

SVD is an algorithm based on an orthogonal matrix transformation. The matrix $A(m \times n)$, when processed by SVD, can be decomposed into three matrices with different dimensions, as shown in Equation 5.7 [207].

$$A = USV^T \tag{5.7}$$

where $U = [u_1, u_2 \dots u_1] \in R^{m \times m}$ and $V = [v_1, v_2 \dots v_1] \in R^{n \times n}$. *S* is a diagonal matrix with *n* singular values with descending order shown as $S = [diag(\lambda_1, \lambda_2 \dots \lambda_n), 0] \in R^{m \times n}$. Singular values from *S* contain the information from the original signal.

The proposed IVMD-SVD method uses an algorithm to automatically select the number of modes (K-value) to be used in the decomposition of the impact response signal. The result, decomposed by the selected modes, is processed to obtain the Shannon entropy which is then used in further analysis.

The appropriate K-value is selected through consideration of the relationship between IMFs (sub-signals) and the original signal. First, SVD is used to obtain eigenvalues from both the IMFs and the original signal. Then, the ratio between the eigenvalues of each IMF and the original signal is calculated. This ratio is then used as an index in the evaluation of the relationship between sub-signals and the original signal. This evaluation process selects the optimal K-value, and hence the number of modes to be used in the decomposition of the response signal. The process for selecting the optimal K-value is described through an example in the following section.

The mechanism by which the optimal K-value is selected is summarised in Figure 5.13.

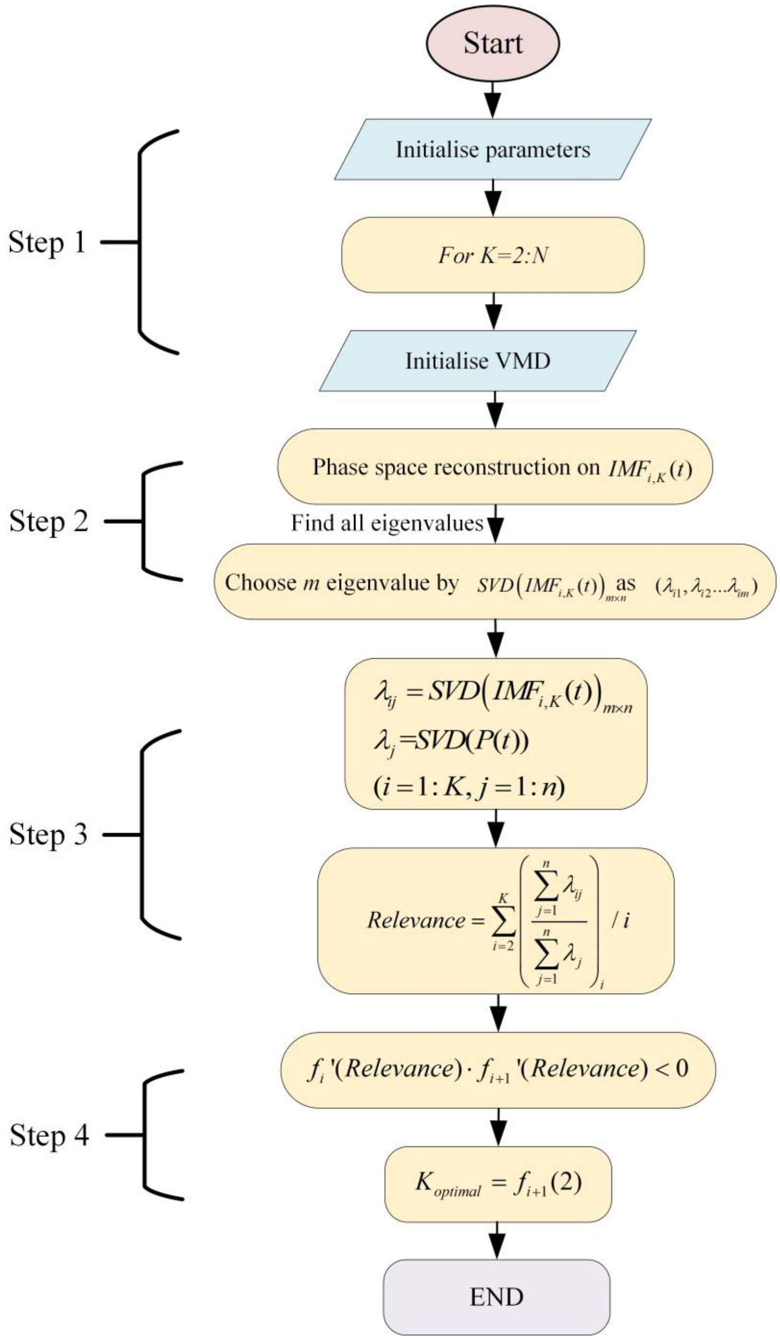


Figure 5.13 Process of selecting optimal K-value

The steps indicated in Figure 5.13 are as follows:

Step 1: Sub-signals $IMF_i(t)$ are obtained by performing VMD using a range of K-values from 2 to *N*, i.e. VMD (*i* = 1,2 … *N*).

Step 2: The IMFs calculated from each VMD are combined to form Hankel matrix representations [208] of each sub-signal using a phase-space reconstruction process as shown in Equation 5.8.

$$X = \begin{pmatrix} IMF_{i1} & IMF_{i2} & & IMF_{in} \\ IMF_{i2} & IMF_{i3} & \cdots & IMF_{i(n-1)} \\ \vdots & & \ddots & \vdots \\ IMF_{i(m-n+1)} & IMF_{i(m-n+2)} & \cdots & IMF_{im} \end{pmatrix} \quad 5.8$$

where m is the sample number of the signal and n is the dimension of the matrix.

Step 3: Eigenvalues are calculated for each Hankel matrix IMF representation using SVD, as shown in Equation 5.9.

$$SVD(IMF_i(t)) \Rightarrow \begin{pmatrix} \lambda_{i1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{ik} \end{pmatrix} \quad 5.9$$

At this stage, the original signal is also converted to a Hankel matrix representation and SVD is applied, as shown in Equation 5.10.

$$SVD((F(t)) \Rightarrow \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_k \end{pmatrix} \quad 5.10$$

The $\{Ratio\}_i$ defined in Equation 5.11 represents the relationship between each sub-signal and the original signal.

$$\{Ratio\}_i = \left(\frac{\sum_{j=1}^{k} \lambda_{ij}}{\sum_{j=1}^{k} \lambda_j} \right) = \begin{pmatrix} Ratio_{12} & Ratio_{13} & \cdots & Ratio_{1N} \\ Ratio_{22} & Ratio_{23} & \ddots & Ratio_{2N} \\ & Ratio_{33} & & \vdots \\ & & & Ratio_{NN} \end{pmatrix} \quad 5.11$$

As the number of IMFs and thus $\{Ratio\}$ varies with K-value, the mean value of $\{Ratio\}$ ($Relevance$) is calculated to allow direct comparison. $Relevance$ is a single measure of the relationship between the original signal and all of its sub-signals. It is generated for each K-value. This is shown in Equation 5.12.

$$Relevance = \sum_{i=2}^{N} \{ratio\}_i / i \qquad 5.12$$

Step 4: The *Relevance* obtained from different K-values can be plotted to form a curve. The second derivative is applied to this curve in the evaluation of the optimal K-value. The optimal value is selected at the point of inflection above a specified threshold. In the next section, the process is demonstrated with an example.

Having used the IVMD approach to identify the optimal K-value, a process using the selected K-value is then applied to achieve fault detection. VMD is performed using the selected K-value to generate IMFs; SVD is then applied to generate eigenvalues for each IMF. This is followed by filtering to generate Shannon entropy, which is then used for fault detection. Shannon entropy is a standard index for evaluating the energy level. The process of using the eigenvalues is as follows.

Each eigenvalue is divided by the sum of the eigenvalues for a particular IMF to obtain a set of ratio results. The first *l* eigenvalues are selected based on evaluating the value of the ratio against a threshold a to eliminate redundant information. The process is shown in Equation 5.13.

$$\sum_{j=1}^{l} \lambda_{ij} / \sum_{j=1}^{k} \lambda_{ij} > a \qquad 5.13$$

The total value of the first *l* eigenvalues is then used to calculate the Shannon entropy which is used in fault detection. By defining $p_{ij} = \lambda_{ij} / \sum_{j=1}^{l} \lambda_{ij}$, the Shannon entropy of each IMF is calculated by Equation 5.14.

$$E_i = \sum_{j=1}^{l} p_{ij} \, log \, p_{ij} \qquad 5.14$$

The Shannon entropies from all IMFs can be combined as shown in Equation 5.15 for use in fault diagnosis.

$$E = [E_1, E_2 \dots E_i] \qquad 5.15$$

In this section, a method based on a combination of VMD, SVD, and Shannon entropy techniques is introduced. The SVD component of the algorithm automatically selects the initial parameters for the VMD element. SVD is then further applied to the VMD result to obtain Shannon entropy, which is the feature used for the detection and classification of underside corrosion. The IVMD-SVD method uses an algorithm to automatically select the number of modes (K-value) to be used in the decomposition of the impact response signal.

5.3.2 Result analysis

In order to demonstrate the decomposition methodology described above, the impact response acoustic signal generated by impacting the healthy rail sample in the location is shown in Figure 5.14 and the acoustic impact response signals in the time and frequency domains are shown in Figure 5.15.

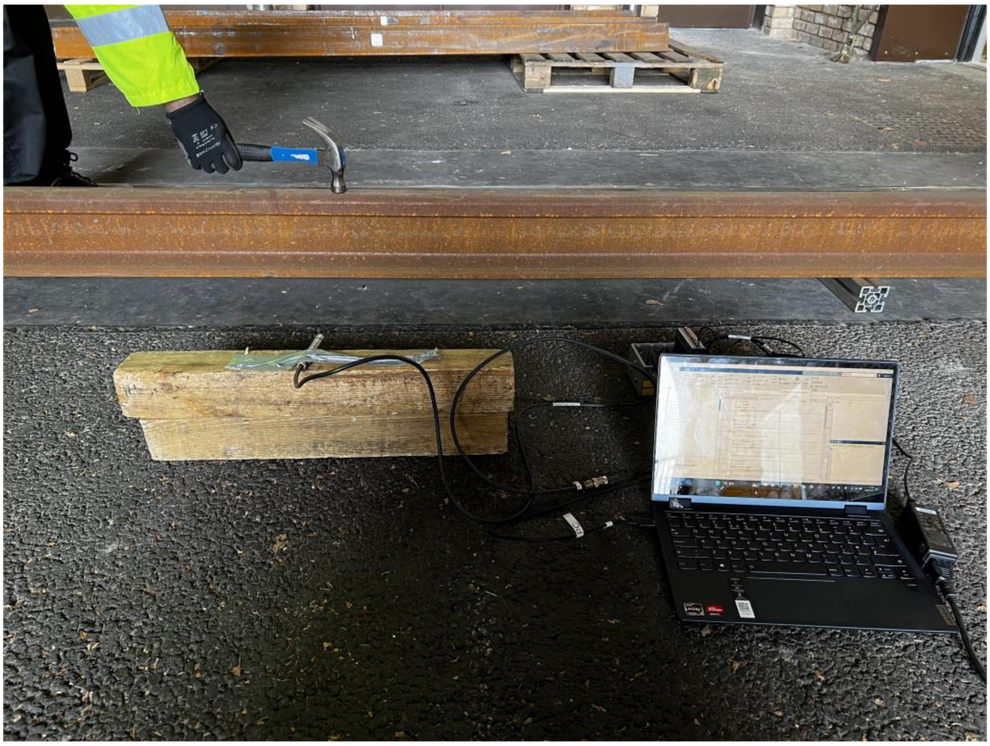


Figure 5.14 Experiment of collecting acoustic signals

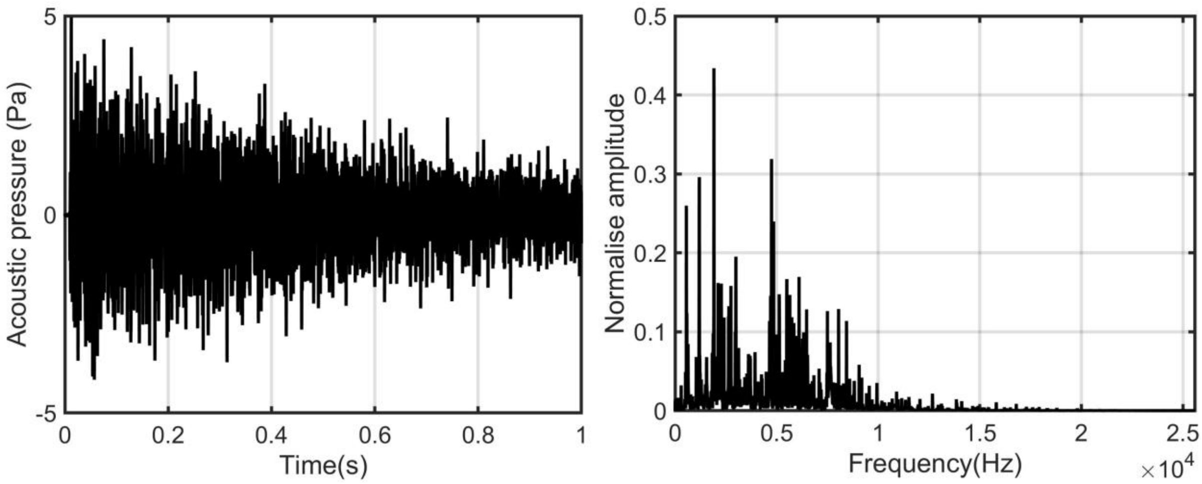


Figure 5.15 Acoustic impact response signal

Then the experimental signals were decomposed using the IVMD-SVD algorithm incorporating VMD, SVD, and Shannon entropy techniques. The K-value was selected automatically by examining the relationship between IMFs (sub-signals) and the original signal. The process began with the use of SVD to obtain eigenvalues from both the IMFs and the original signal. Then, the ratio between the eigenvalues of each IMF and the original signal was calculated and named $Relevance$. $Relevance$ serves as an index to evaluate the relationship between the sub-signals and the original signal. This evaluation process picks the optimal K-value, determining the number of modes to be used in the decomposition of the response signal. An example is provided to demonstrate the process of automatically selecting the K-value as follows.

In the current research on selecting the K-value, the value range is usually from 3 to 8 [209],[210],[211]. But in order to completely present the effectiveness of the VMD with different K-values, K-values from 3 to 12 were selected. By operating the VMD with these K-values, respectively, the $Relevance$ representing the relationship between the original signal and IMFs generated by VMD with a specific K-value is obtained. A curve connecting these $Relevance$ values was then generated as shown in Figure 5.16.

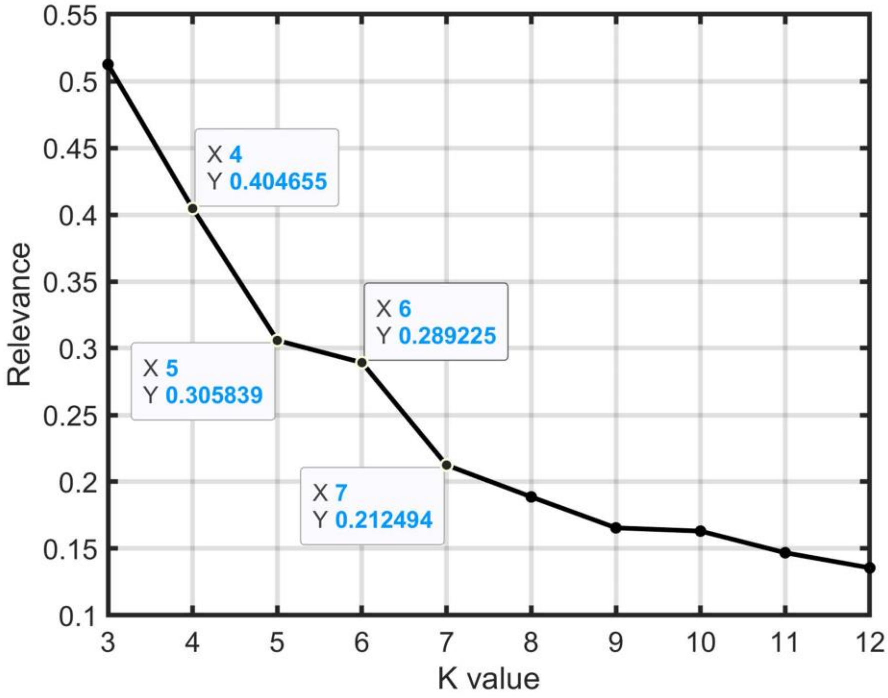


Figure 5.16 Optimal K-value selection from the acoustic impact response signal

The curve in Figure 5.16 shows a general downtrend. If the K-value was large, the sub-signals from the higher-numbered decomposed modes were generally less relevant to the original signal. Conversely, if the K-value was too small, useful information may have been excluded. The preferred K-value occurred at the first point of significant inflexion in the curve. This was identified by observing the sign and the difference in values of the second derivative of the curve. The value difference should have been at least 0.05, as identified by Wang et al. [115].

Using the example in Figure 5.16, the values of the second derivative at K-values of 5 and 6 were −0.082 and 0.060, respectively. The product of these two values was negative. And the absolute difference was greater than 0.05. Hence, the optimal K-value was chosen to be K = 6.

The Shannon entropy was then calculated using IVMD-SVD with the selected K-value. The process was applied to 30 measurements of each of the four rail conditions, as presented in Figure 5.17. The blue, red, yellow, and green bars represent healthy rail, square faults, rectangular faults, and circular faults, respectively.

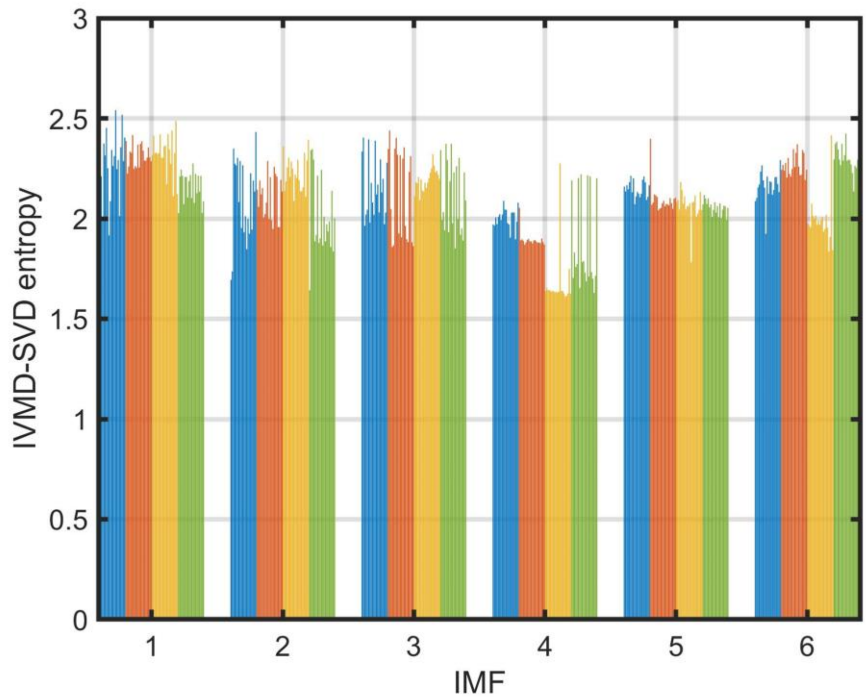


Figure 5.17 Entropy results from IVMD-SVD

In order to evaluate the effectiveness of the IVMD-SVD, it is compared to the energy ratio, which is a classic energy-based method [212]. The approach is based on the sum of the energy from the original signal considered in the frequency domain. The energy ratio is calculated as the ratio between each IMF and the total energy, as shown in Equation 5.16.

$$E_i = \sum |IMF_i(t)|^2 \ (i = 1,2 \dots m)$$

$$E = \sum_{i=1}^{m} E_i \quad 5.16$$

$$p_i = E_i / E$$

The comparison of Shannon entropy results obtained from IVMD-SVD and the energy ratio method is shown in Figure 5.18.

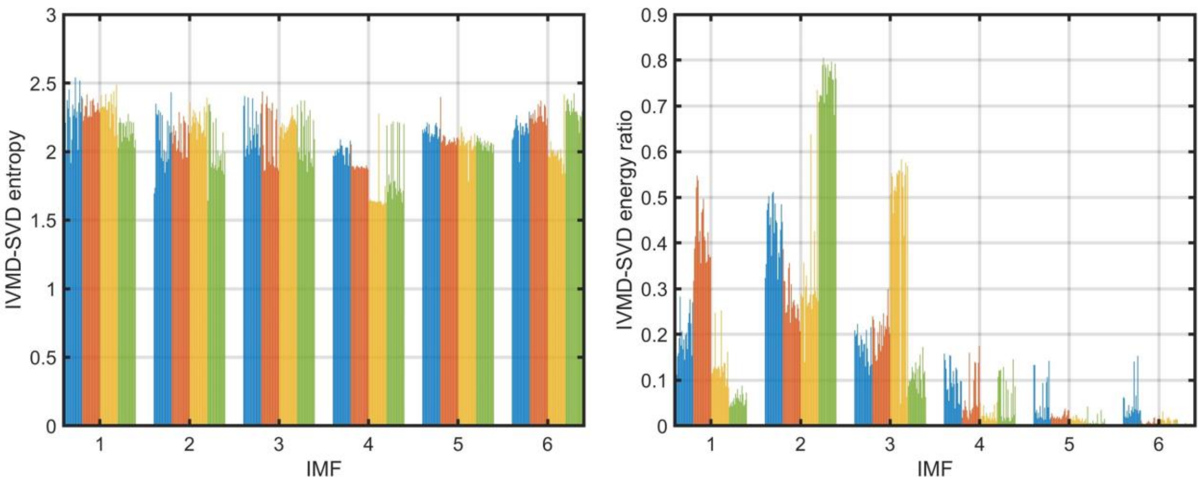


Figure 5.18 Comparison of results obtained using IVMD-SVD and energy ratio methods

The statistical parameter distinction degree (*D*) is used to evaluate the uniformity of the distribution of energy-based results obtained for the same rail conditions. Firstly, considering each rail condition, each entropy in each IMF (E_{ij}) is divided by the mean value of all entropies to obtain a set of ratios (R_{ij}), using Equation 5.17.

$$R_{ij} = E_{ij} / \left(\sum_{j=1}^{z} E_{ij} / z \right) \quad 5.17$$

where z is the number of signals obtained for a particular rail condition.

Secondly, the absolute difference (D_i) between each two adjacent ratios (R_{ij} and $R_{i(j+1)}$) is calculated. And then the distinction degree (*D*) is obtained by summing these differences (D_i) as in Equation 5.18.

$$D_i = \sum_{j=1}^{z} \left| R_{ij} - R_{i(j+1)} \right| \quad 5.18$$

Finally, the set of distinction degrees including all IMFs is represented as Equation 5.19.

$$D = [D_1, D_2 \ldots D_i] \quad 5.19$$

A comparison of the distinction degree obtained using the two energy-based methods is shown in Figure 5.19.

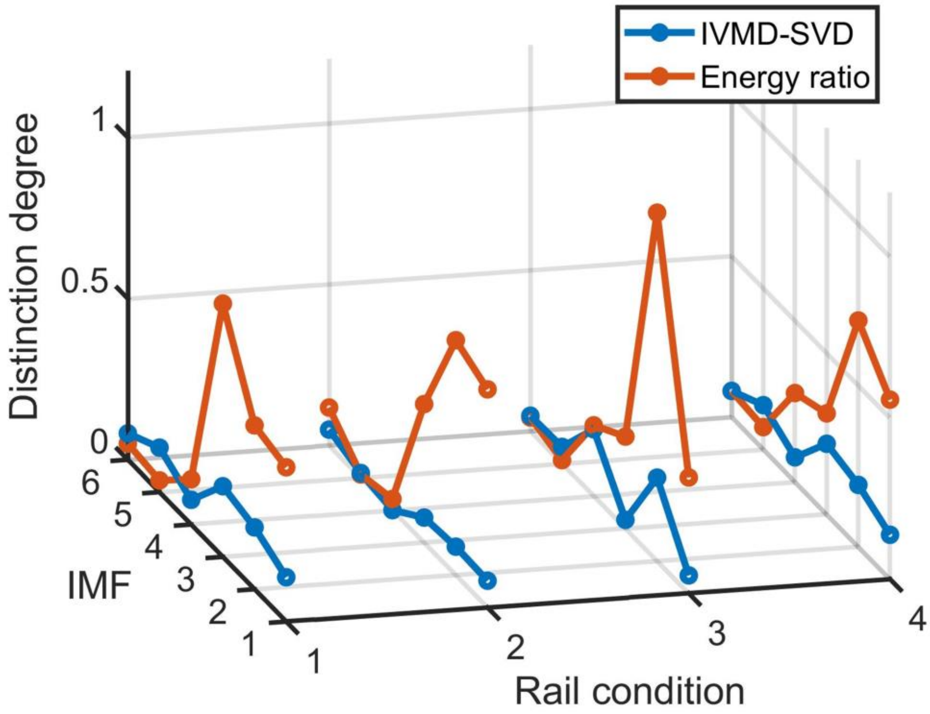


Figure 5.19 Comparison of distinction degree between two methods for acoustic signal

Figure 5.19 shows that, for all healthy conditions, the average distinction degree obtained from the entropy result is lower than that obtained using the energy ratio. This provides an indication that the entropy results processed using IVMD-SVD are more uniform than those obtained using the energy ratio method. The proposed method is, therefore, considered more robust against environmental interference such as that caused by a change of impact location or force.

In order to further validate the effectiveness of the IVMD-SVD algorithm, the evaluation process is repeated using vibration-based impact response signals. This is another type of impact response signal closely related to acoustic signals.

The optimal K-value selection for the vibration-based impact response signals is shown in Figure 5.20. The product of the second derivatives at K-values of 5 (0.066) and 6 (−0.018) is negative. The absolute difference between these two-second derivatives is 0.057 which is larger than the threshold of 0.05. This indicates that the optimal K-value is 6, the same conclusion as drawn for the acoustic version of the impact response signal.

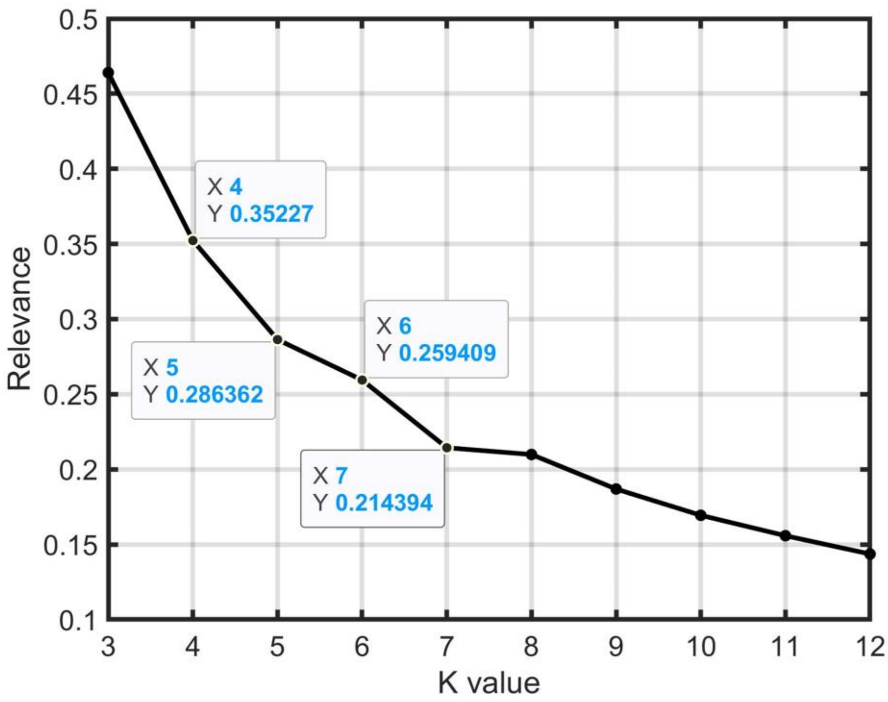


Figure 5.20 Optimal K-value selection from impact response vibration signal

The distinction degree results for IVMD-SVD entropy and energy ratio methods are presented in Figure 5.21.

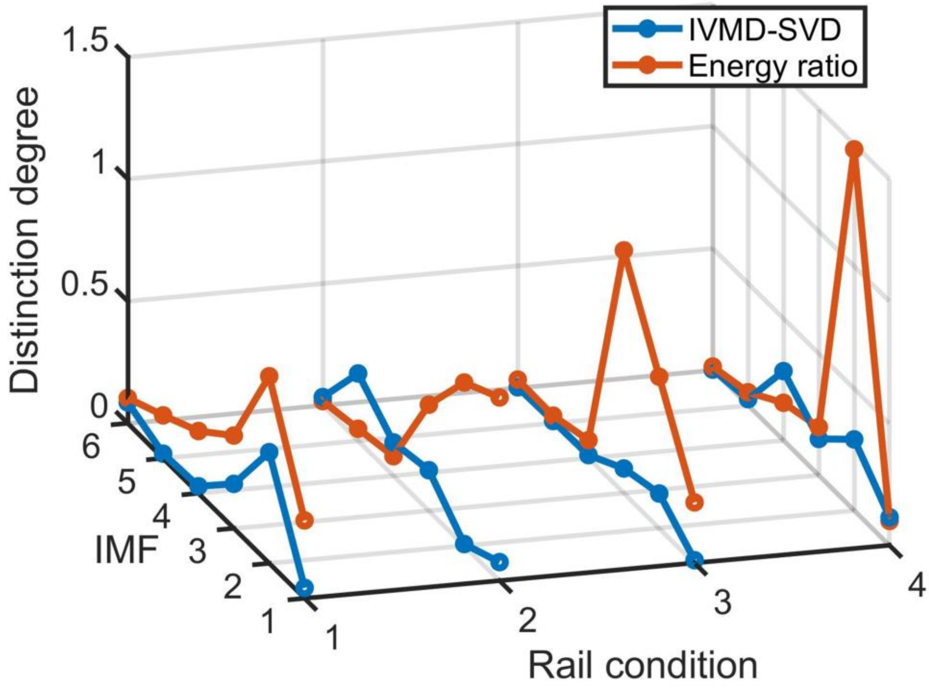


Figure 5.21 Comparison of distinction degree between two methods for vibration signal

The result for the vibration-based signal presented in Figure 5.21 is similar to that obtained for the acoustic signal. The distinction degree values calculated from IVMD-SVD are lower than those calculated from the energy ratio. This further demonstrates that IVMD-SVD produces more uniformly decomposed signals and that there is consistency when applied to both acoustic- and vibration-based signals. Hence, it is considered that the approach is suitable for reducing environmental effects in a range of recorded signals.

To illustrate the effectiveness of appropriately selecting the K-value for fault classification, an index representing the capacity for identification (C_k) of different rail conditions is calculated. The index can be described as a subtraction of the maximum and minimum differences from the mean value of entropy, as given in Equation 5.20.

$$c_{max\,i} = max\left|\overline{C_i} - E_{ij}\right|$$

$$c_{min\,i} = min\left|\overline{C_i} - E_{ij}\right| \tag{5.20}$$

$$C_k = \sum_{i=1}^{K} \left|c_{max\,i} - c_{min\,i}\right|$$

where $\overline{C_i}$ is the mean value of $IMF_i(t)$. The identification capacity index results for each K-value, generated for the acoustic impact response signals, are shown in Figure 5.22.

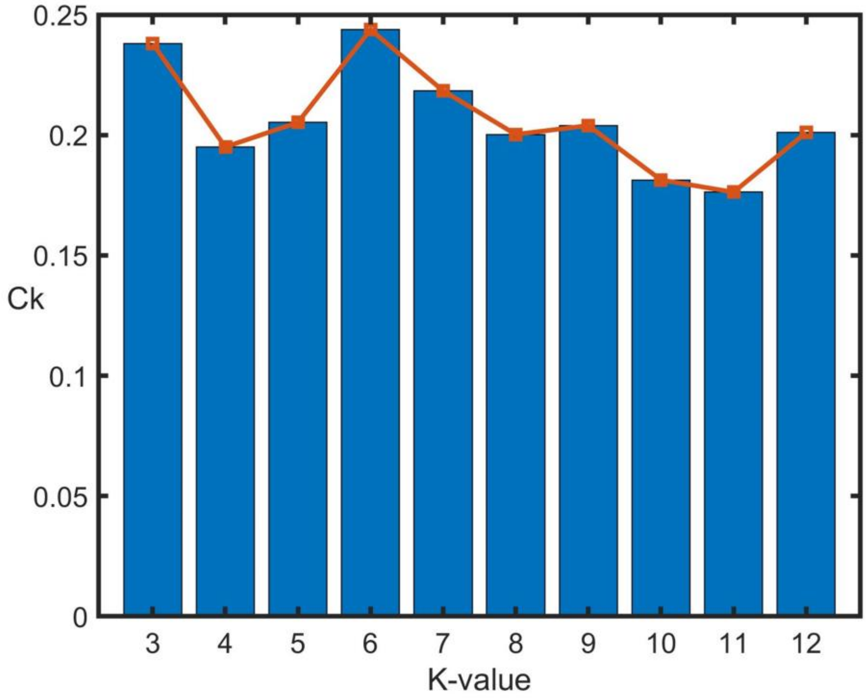


Figure 5.22 Identification capacity index values (acoustic signal)

The values presented in Figure 5.22 represent the differences in signals obtained for different rail conditions, considered by the K-value used in the processing. The figure shows that the highest value of C_k corresponded to K = 6. This further demonstrates that the selected K-value is the most appropriate for differentiating the different rail conditions.

And then, the analysis obtained identification capacity index values for the vibration signal, as for the evaluation of the distinction degree. The corresponding result is presented in Figure 5.23

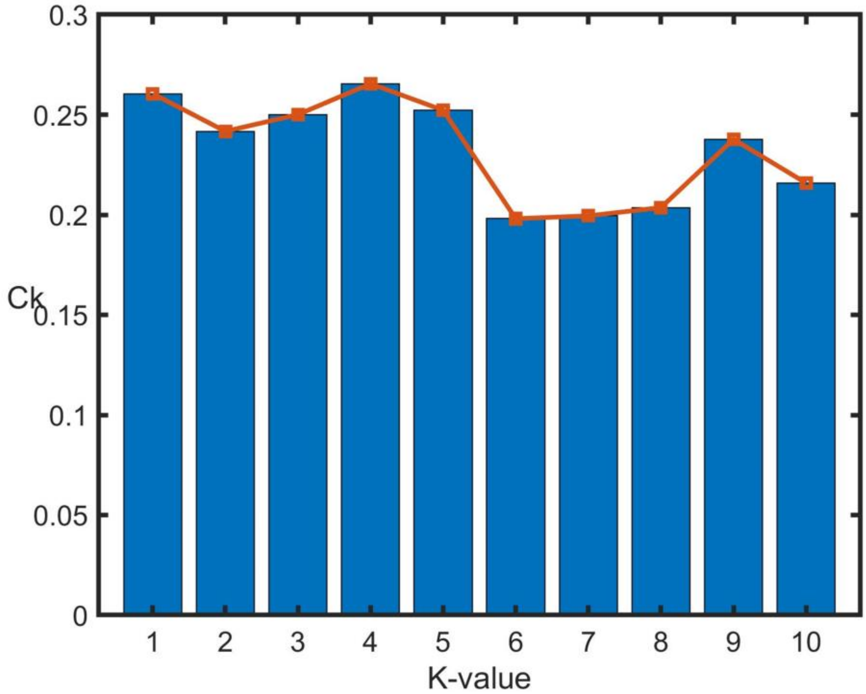


Figure 5.23 Identification capacity index values (vibration signal)

Considering the results of the identification capacity index values for vibration signals, the highest value of C_k also corresponded to K = 6. Additionally, the variance of the value of C_k was consistent with the conclusion for the acoustic signals.

In conclusion, in this section a novel algorithm (IVMD-SVD) has been introduced to automatically select initial parameters for the VMD algorithm in this section. The IVMD-SVD algorithm uses Shannon entropy to evaluate the significance of rail base corrosion. Additionally, the statistical parameter distinction degree was used to compare the uniformity of energy-based results for different rail conditions. The results demonstrated that the proposed entropy-based method is more resistant to environmental interference, such as changes in impact location or force. Another index, which was based on the maximum and minimum difference from the mean value, was used to evaluate the ability to identify various track conditions. Using the proposed method to select the K-value allowed for the effective identification of track conditions and faults from acoustic impact response signals.

The proposed method is able to maximise its efficiency in decomposing impact response signals and minimising noise interference. However, during the experiments, some data were collected with errors such as those arising from unexpected environmental noise or experimental inaccuracies during data collection, which could negatively impact identification results. Therefore, it is essential to eliminate these erroneous data before fault identification and classification. In the next section, an algorithm will be introduced to automatically remove these erroneous data generated during the experiments.

5.4 Data pruning

In this section, an algorithm is introduced to eliminate the erroneous data generated through the experiments. Some of the data acquired during the experiments and signal extraction are probably anomalous and unable to be used for further analysis. Additionally, it is difficult to find these erroneous data manually. Hence, an algorithm based on the extraction of principal components through SVD is designed for data pruning.

5.4.1 Introduction of the algorithm

The acoustic data collected from the experiment might contain experimental errors due to the precision of the impact site, location of the sensor, etc. And the acoustic data collected from the simulation might contain errors due to the model computation. Additionally, the simulated signal generated from the model is not exactly the same as the experimental signal. When there is an error in the experimental signal, the difference between the simulated signal and the experimental signal might be too obvious. Hence, when additional environmental noise is introduced, the signals become non-representative. In order to describe this difference, SVD is used to extract the principal components from the simulated and experimental signals. And by evaluating the differences of the principal components extracted from the SVD, the relationship between the simulated signal and the experimental signal can be evaluated. The details of this process are introduced as follows:

Firstly, the collected signal is given as $f(t)$. The signal is processed by IVMD-SVD. The obtained entropy can be used for the analysis. The set of entropies ($E = [E_1, E_2, \dots, E_n]$) is reconstructed and formed as a Hankel matrix $H(t)$, which is given as Equation 5.21.

$$H(t) = \begin{pmatrix} E_1 & E_2 & & E_n \\ E_2 & E_3 & \cdots & E_{(n-1)} \\ \vdots & & \ddots & \vdots \\ E_{(m-n+1)} & E_{(m-n+2)} & \cdots & E_m \end{pmatrix} \qquad 5.21$$

where n is the number of IMFs and m is the dimension of the matrix.

Then the Hankel matrix obtained from the simulated signal $H_s(t)$ and experimental signal $H_e(t)$ is decomposed by SVD, which is shown in Equation 5.22.

$$SVD(H_s(t)) = \begin{pmatrix} \lambda_{s1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{sm} \end{pmatrix}$$

$$SVD(H_e(t)) = \begin{pmatrix} \lambda_{e1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{em} \end{pmatrix} \qquad 5.22$$

Two sets of eigenvalues ($\{\lambda_{s1}, \lambda_{s2}, \dots \lambda_{sm}\}$, $\{\lambda_{e1}, \lambda_{e2}, \dots \lambda_{em}\}$) are obtained through Equation 5.22. The first eigenvalues obtained from all simulated signals $\lambda_s = \{\lambda_{s1}, \lambda_{s2}, \dots, \lambda_{sk}\}$ and all experimental signals $\lambda_e = \{\lambda_{e1}, \lambda_{e2}, \dots, \lambda_{el}\}$ are then used to evaluate the difference between simulated and experimental signals. k and j are the number of simulated and experimental signals, respectively.

Dynamic time warping (DTW) is an algorithm that is designed to measure the Euclidean metric between two audience data in the beginning [213]. It is also can evaluate the difference between the two signals [214]. The process of the DTW is given as follows:

1) Based on two time series signals $f_1(t) = (x_1, x_2 \dots x_m)$ and $f_2(t) = (y_1, y_2 \dots y_n)$, where *m* and *n* represent the sequences of the signal, a matrix (*M*) with an *m*-by-*n* path can be created as in Equation 5.23.

$$M = \begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & \cdots & x_m \end{array} \\ \begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} & \begin{pmatrix} & & \cdots & \\ & & & \\ \vdots & & \ddots & \vdots \\ & & \cdots & \end{pmatrix} \end{array} \qquad 5.23$$

2) $d(x_i, y_j)$ is defined as the distance of two time series, where the ranges of i and j are 0 to m and n, respectively.

3) A cost matrix C is obtained by creating a matrix $C(n+1, m+1)$ such that $C(i,j)$ contains the cumulative distance between x_1 to x_i and y_1 to y_j.
4) The method for calculating the cumulative distance starts from $C(1,1)$. And the cumulative distance between each pair of corresponding elements in the two time series is computed up to $C(1,1)$. The detailed cumulative distance is computed as follows:

   For each element in the first time series, find the corresponding element in the second time series that minimises the distance measure $d(x_i, y_j)$. This is called the 'warping path'. $C(i,j)$ is updated to be the sum of the distance measure $d(x_i, y_j)$ and the minimum of the three neighbouring values $C(i-1,j)$, $C(i,j-1)$, and $C(i-1,j-1)$.

5) Find the optimal warping path: Once the cost matrix is computed, the optimal warping path can be found by backtracking from $C(n,m)$ to $C(0,0)$. The optimal path is the one that minimises the cumulative distance between two time series.

In this thesis, based on the theories of the above applications, there are two steps for completing the pruning algorithm. Firstly, each experimental signal will be compared with another experimental signal. Specifically, each eigenvalue obtained from the experimental signal (λ_e) is compared with each eigenvalue obtained from other experimental signals (λ_e) through the DTW method. The results from DTW consist of a set. The size of the set is ($l \times l$). And the threshold (β) against the value of DTW results is used to analyse the relevance between each two experimental signals. If β is greater than 2, the experimental signal corresponding to this result will be removed. The new size of the dataset of the experimental signal is u.

Then, the rest of the experimental signals are compared with the simulated signals. One time series is defined as a set of the first eigenvalues obtained from simulated signals (λ_s).

And another series is defined as a set of the first eigenvalues obtained from the rest of the experimental signals (λ_{er}). Each eigenvalue obtained from a simulated signal (λ_s) is then compared with each eigenvalue obtained from an experimental signal (λ_{er}) through the DTW method. The results from DTW consist of a new set. The size of the set is ($g \times u$). The threshold (β) against the value of DTW results is used to analyse the relevance between the experimental and simulated signals. If β is greater than 2, the simulated signal corresponding to this result will be removed.

The process has the capability to automatically decrease the errors incurred during the experiment, as well as eliminate errors arising from the computation of the simulated model.

5.4.2 Data processing and analysis

A comparison of the entropy result before and after the pruning algorithm is completed. A scatter figure is applied to demonstrate the range of all entropy results from the dataset. As an example, one of the results considered the signals obtained from the healthy rail, which are shown in Figure 5.24 and Figure 5.25.

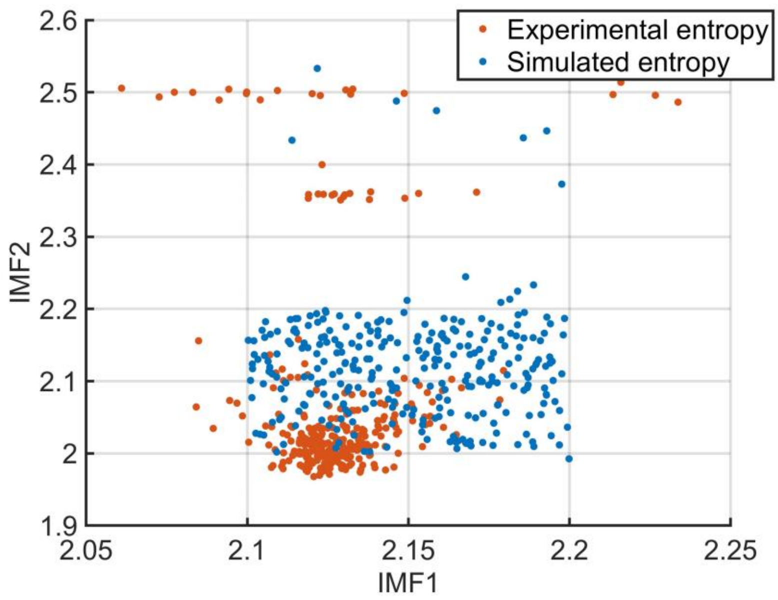


Figure 5.24 Distribution of entropy from experimental and simulated results before the pruning algorithm

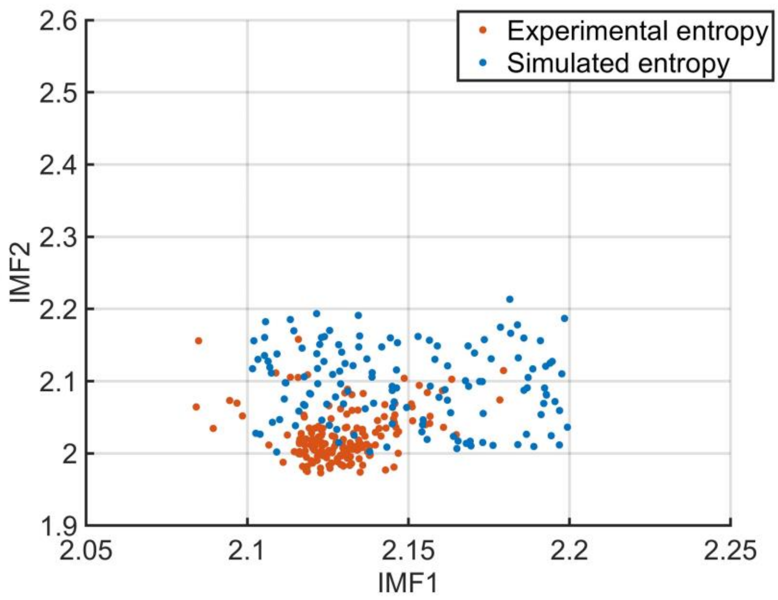


Figure 5.25 Distribution of entropy from the experimental and simulated results after the pruning algorithm

In Figure 5.24 and Figure 5.25, the blue and red spots represent the experimental and simulated results, respectively. Figure 5.24 shows the original results containing all entropies from the experimental and simulated signals. The results from the *x*-axis and *y*-axis show the entropy results of IMF 1 and IMF 2, respectively. Entropy values over 2.3 are too high and are significantly different to the majority of results. It can be observed that these results might contain an error due to an experimental mistake or calculation errors from the simulation model. The proposed pruning algorithm is designed to remove these results. As shown in Figure 5.25, on using the algorithm the entropy results are concentrated in the range of 1.9 to 2.2 on the IMF2 axis and 2.05 to 2.2 on the IMF 3 axis.

In conclusion, this section introduces a pruning algorithm which uses the techniques of SVD and DTW. This algorithm can automatically remove the erroneous results which can be generated due to experimental mistakes or simulation errors. From this operation, the processed dataset can be used for classification with better performance.

5.5 Conclusion

In this chapter, the impact response signal is processed in order to remove the noise and extract the information corresponding to fault features. Firstly, besides the simulated signal generated from a simulation model which was introduced in Chapter 4, a laboratory set-up is introduced for collecting the experimental signal. After obtaining the impact response signal through the experiment, four signal decomposition methods were selected for processing these impact response signals. According to the analysis, VMD performed best in decomposing the impact response signal compared to EMD, EEMD, and LMD. And then, VMD as the selected decomposition method was enhanced through a proposed method (IVMD-SVD). Using the proposed method, the appropriate initial parameter (K-value) was selected for automatically processing the impact response signal through VMD. The Shannon entropy results obtained for the same rail condition were more uniform than those obtained using the traditional method. By evaluating the entropy results, the effectiveness of the proposed method was proved. The entropy results from the proposed method were then able to be used for classification. However, due to the existence of experimental mistakes, potential data errors should be removed before classification. Hence, a pruning algorithm using the techniques of SVD and DTW was proposed for automatically removing erroneous data. The processed dataset will be used for classification in the next chapter.

In the next chapter, datasets with different data (simulated signal or experimental signal; vibration signal or acoustic signal) or different corrosion categories (corrosion types or corrosion level) will be classified through a designed machine learning-based classifier. According to the performance of the classifier, the corrosion at the underside of the rail can be completely identified.

CHAPTER 6 FAULT CLASSIFICATION OF CORROSION ON THE UNDERSIDE OF THE RAIL

This chapter introduces using a machine learning-based method to identify the different faults of corrosion on the underside of the rail. The basic theory of the 1-D CNN model is introduced in Chapter 3, and the 1-D CNN model will be used for the classification in this chapter. The simulated dataset from Chapter 4 and the experimental dataset obtained from Chapter 5 will be used for the fault classification. In this Chapter, based on the discussion in the literature review, 1-D CNN has been used for classifying the structural faults on the railway system with significant effectiveness. Therefore, it will be used for the classification in this thesis. Firstly, the different types of corrosion and the different levels of corrosion are separately classified through independent 1-D CNN models. Specifically, in the first model, the corrosion types, square, rectangular, and circular, are classified against a fourth option of a ‘healthy’ rail. In the second model, the level of circular corrosion is classified as either 0.25, 0.75, or 3 mm depth, or the rail may be identified as healthy. And then, the effectiveness of classifying the different types and levels of corrosion against a healthy rail simultaneously is evaluated. An integrated 1-D CNN model with a specific structure is designed for the classification. And in order to evaluate the effectiveness of the designed integrated 1-D CNN model, a conventional 1-D CNN model that simultaneously considers corrosion type and level is used for comparison. According to the analysis, the integrated model proves that corrosion on the underside of the rail can be identified by classifying the processed impact response signals. All of these classifiers have been evaluated using both simulated impact response signals, generated using the model described in Chapter 4, and signals obtained experimentally. Additionally, the simulated signal and experimental dataset were combined to show compatibility of the classification. The experimental dataset was used for testing the integrated 1-D CNN model which was trained through the simulated dataset. Analysis of the classification result indicates that the simulated signal could be used for identifying underside corrosion from the real world.

6.1 Model preparation

The fault classification will be completed through the 1-D CNN model. The basic structure of the 1-D CNN model is introduced in Chapter 3. However, before operating the fault classification, some information about the 1-D CNN model should be introduced. Firstly, preparation of the dataset is introduced. Secondly, in order to validate the performance of the 1-D CNN model, the cross-validation method is introduced. Thirdly, the hyper-parameters of the 1-D CNN model which are mentioned in Chapter 3 should be determined. In this section, a hyper-parameter tuning technique considering the automatic selection of hyper-parameters is introduced. And four metrics commonly used for evaluating the performance of machine learning-based methods are introduced.

6.1.1 Data preparation

The datasets were obtained from the acoustic results obtained from the experiment and those from the simulation model. Firstly, the results were labelled according to the different corrosion types – square corrosion, rectangular corrosion, circular corrosion, and healthy rail – to identify the shape of corrosion on the underside of the rail. And then, the results classified as circular corrosion were labelled according to four different corrosion levels (0.25, 0.75, or 3 mm depth, and healthy rail). Considering the experimental datasets, the results from the experiment were directly used for the classification.

When analysing the simulated signal, noise was added to simulate field noise like rotational machine noise and wind noise. To evaluate the effectiveness of the 1-D CNN model, simulated signals were added with different levels of noise power.

6.1.2 Cross-validation and K-fold method

Cross-validation is a method used in machine learning and statistical modelling to evaluate the performance of a model. It involves partitioning the data into multiple subsets, or folds, and using one or more of these folds as a validation set while the rest are used as the training set [215].

The general idea is to train the model on a subset of the data and test it on another subset, and to repeat this process multiple times with different subsets, to get a more robust estimate of the model's performance. Cross-validation is particularly useful when the dataset is small or when the model is prone to overfitting, as it helps to reduce bias and variance. There are several types of cross-validation methods, including K-fold cross-validation, leave-one-out cross-validation, and stratified cross-validation [215].

K-fold, as a common cross-validation method, is widely used to estimate prediction accuracy [216]. It entails dividing a dataset into K subsets or folds of roughly equal size. One fold is designated as the validation set, while the remaining $K - 1$ folds are utilised for training the model. This process is repeated K times, with each fold serving as the validation set once. Consequently, every data point is utilised for both training and validation purposes, and the model's performance is assessed based on the average of the0 K validation scores. K-fold cross-validation proves particularly advantageous when dealing with small datasets or models that are prone to overfitting. It provides a more accurate estimate of the model's performance than simply splitting the dataset into training and testing sets, as it uses all the available data for both training and validation. Additionally, it helps to reduce the bias and variance of the performance. Compared with the typically practised cross-validation method, the results from the K-fold method are closer to real performance [217]. Hence, K-fold cross-validation is used for evaluating the performance of the 1-D CNN model in this chapter.

6.1.3 Hyper-parameter tuning technique

Hyper-parameters in machine learning refer to parameters that are not learned during the training process but are manually set by the user before training. They play a crucial role in determining the performance and behaviour of the machine learning model. The model's accuracy and generalisation ability can be significantly affected by hyper-parameters. In this section, an adaptive application [187] is used for randomly searching the optimal hyper-parameters. As introduced in Chapter 3, the hyper-parameters considered include the number of filters, the number of CNN layers, weight decay, learning rate, and the number of filters in the dense layers [187]. The kernel_initialiser is selected between the He normal initialiser [176] and the Glorot uniform initialiser [218]. And the activation function is selected between Sigmoid, Tanh, and ReLU. Additionally, the number of epochs for training the model is 200, and 10 trials are used for random searching of hyper-parameters. Furthermore, selecting the optimal hyper-parameter of weight decay can effectively prevent a model from overfitting. Generally, these operations can gain the best effectiveness of 1-D CNN model for classification. The selected ranges of these hyper-parameters are shown in Table 6.1.

Table 6.1 Selection range of CNN hyper-parameters

CNN hyper-parameter	Selection range
Number of filters in the convolutional layers	32–512
CNN layers	1–3
Weight decay	1×10^{-7}–1×10^{-1}
Learning rate	1×10^{-6}–1×10^{-1}
Dense layers	1–3
Number of filters in the dense layers	10–512

Before classification, the validation metrics for evaluating the performance of the 1-D CNN model should be introduced. Accuracy, F1-score, precision, and recall are four important metrics commonly used for evaluating the performance of machine learning-based methods [219]. Accuracy is defined as the closeness of a measured value to the true value [220]. Precision and recall are metrics that evaluate the ability to retrieve data from the original dataset [221]. The F1-score is a measurement combining the precision and recall of the model to further reflect the accuracy of evaluating results [222].

6.2 Fault classification of different corrosion types by experimental signals

In the previous chapters, the impact response signal was processed through the proposed method (IVMD-SVD). The next stage is to explore how to classify the different rail conditions according to the impact response signals based on the machine learning-based method.

Firstly, in this section, the different types and levels of corrosion from the experimental dataset are separately classified through independent 1-D CNN models.

6.2.1 Classification of corrosion types through experimental signals

The dataset obtained from the experiment with different corrosion types and a healthy rail is classified in this section. Considering the types of corrosion on the underside of the rail mentioned in the previous chapter, a dataset containing the signals generated from the healthy rail, square corrosion, rectangular corrosion, and circular corrosion (0.75 mm) is labelled for the classification.

The dataset was trained through the 1-D CNN model for selecting hyper-parameters by the hyper-parameter tuning technique. The datasets with different noise power were trained by an individual 1-D CNN model. The hyper-parameter tuning technique was used for each individual 1-D CNN model every time. The selected ranges of hyper-parameters were based on Table 6.1. The hyper-parameters of the 1-D CNN model selected for classifying the dataset of corrosion types and a healthy rail from experiments are shown in Table 6.2.

Table 6.2 CNN model hyper-parameters selected for classifying corrosion types

Selected hyper-parameter	Value
CNN layers	1
Number of filters in the first convolutional layer	480
Weight initialisation	2.4040×10^{-3}
Learning rate	1.7141×10^{-3}
Number of filters in the dense layers	256
Kernel_initialiser	He normal
Activation function	Tanh

Then, by training the 1-D CNN model with these selected hyper-parameters, the performance of the 1-D CNN model for classifying the different corrosion types could be obtained. As mentioned in Section 6.1.2, cross-validation (K-fold) was applied to evaluate the performance. Firstly, the ‘validation performance’ demonstrates the effectiveness of the 1-D CNN model. The validation performance with 10-fold cross-validation is shown in Figure 6.1.

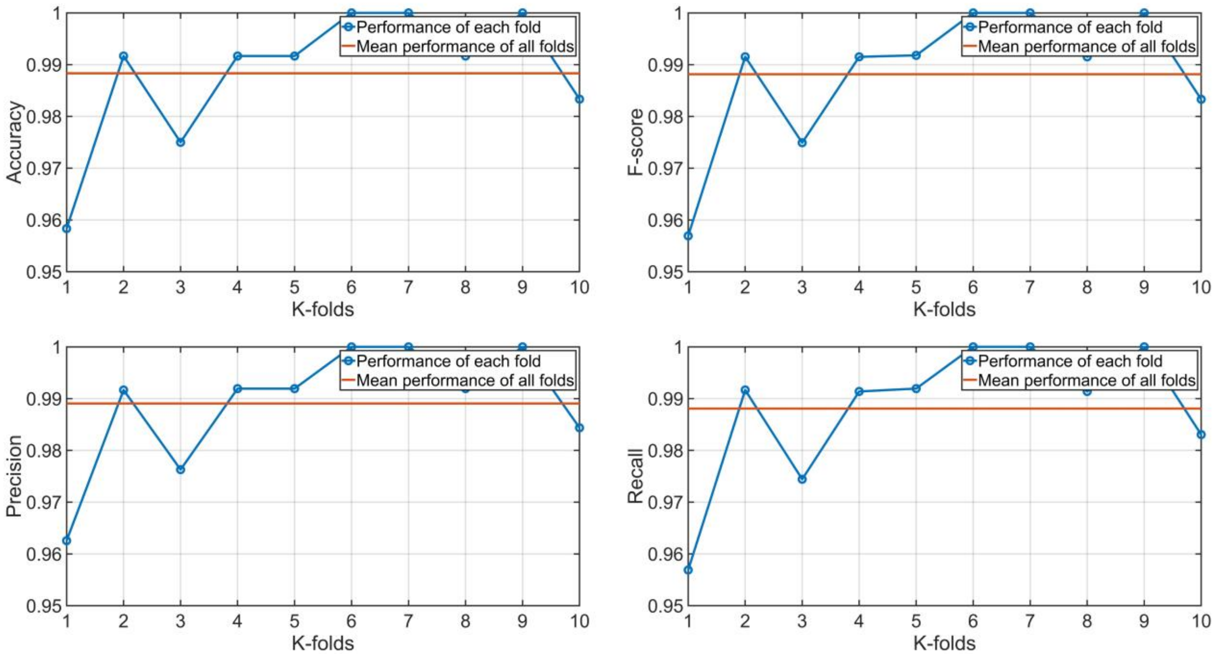


Figure 6.1 Validation classification of fault type performance

The validation performance in Figure 6.1 show that the mean performance value for each of the four metrics was over 0.98. In terms of each K-fold, the results for the four validation performance metrics were over 0.96, and the variance of the value from the validation performance metrics was lower than 0.02. The validation performance value for the four metrics was similar and stable, proving that the 1-D CNN model is efficient for the classification of the corrosion types.

And the ‘classification accuracy’ demonstrates the classification performance of the 1-D CNN model. A confusion matrix was used to present the classification performance, as shown in Figure 6.2.

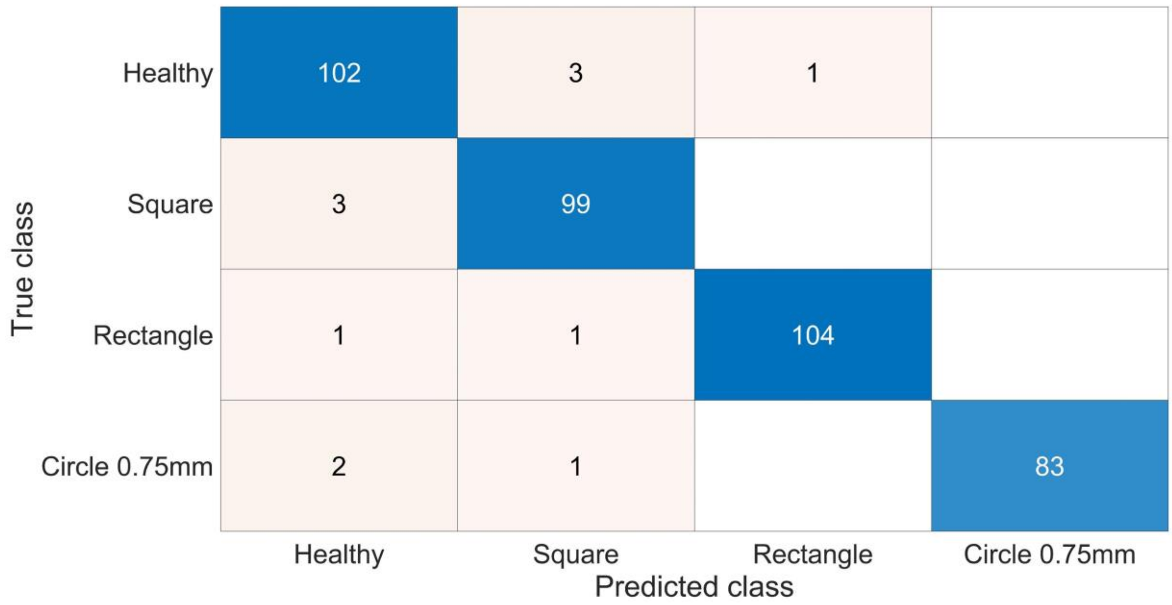


Figure 6.2 Confusion matrix for the testing classification of corrosion types from the experiment

The confusion matrix results in Figure 6.2 demonstrate the significant effectiveness of the model in classifying corrosion types from the experiment. Three different fault types and a healthy rail could be classified. The classification accuracy was 97.5%, proving that the 1-D CNN model is able to classify rail corrosion types through experimental impact response signals.

6.2.2 Classification of corrosion levels through experimental signals

The classification of underside corrosion focuses not only on the different types of corrosion but also on the different levels of corrosion. After classifying the types of corrosion of the underside of the rail, the corrosion level was then classified against a healthy rail through the 1-D CNN model. The process of the hyper-parameter tuning technique is the same as in Section 6.2.1. According to this technique, the hyper-parameters of the 1-D CNN model for classifying the corrosion levels from experimental signals were selected as shown in Table 6.3.

Table 6.3 CNN model hyper-parameters selected for classifying corrosion levels

Selected hyper-parameter	Value
CNN layers	1
Number of filters in the first convolutional layer	352
Weight initialisation	1.4187×10^{-5}
Learning rate	8.8770×10^{-3}
Number of filters in the dense layers	256
Kernel_initialiser	He normal
Activation function	ReLU

After training the 1-D CNN model with these selected hyper-parameters, the model's validation performance with 10-fold cross-validation was shown in Figure 6.3.

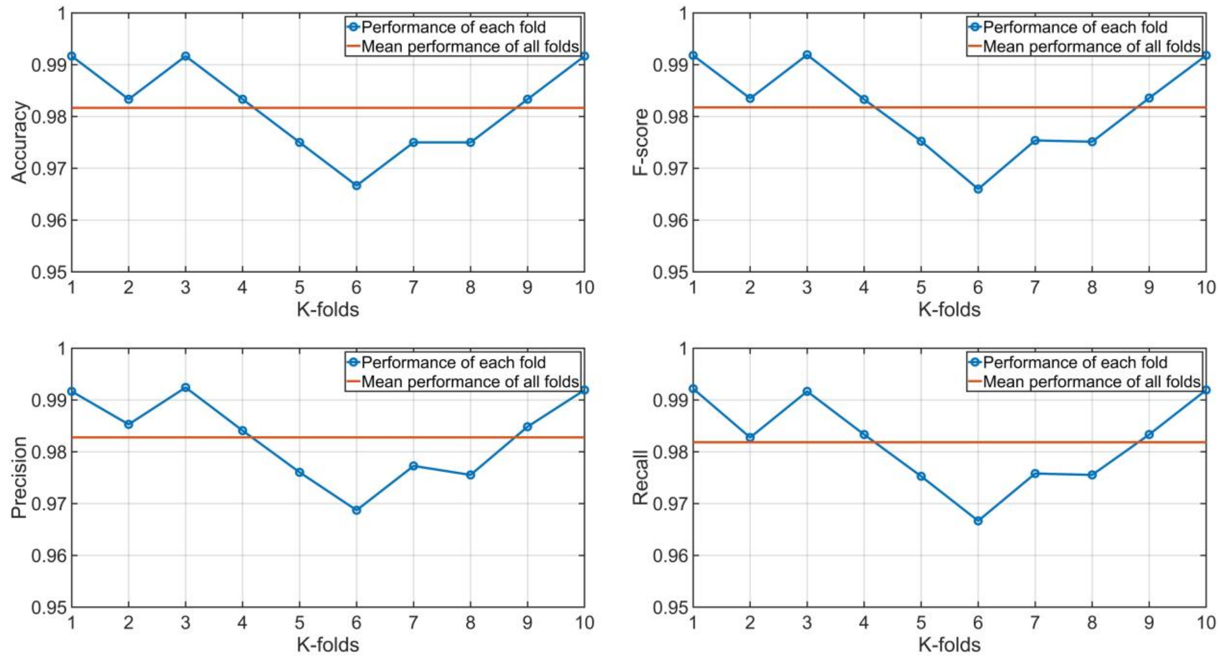


Figure 6.3 Validation performance of fault level classification

The validation performance in Figure 6.3 show that the mean value for each of the four metrics was over 0.98. In terms of each K-fold, the average results for the four validation performance metrics were over 0.96, and the variance of the value from the validation performance metrics was lower than 0.02. This indicates that the effectiveness of 1-D CNN model is good. In order to investigate the classification accuracy, a confusion matrix was used, as shown in Figure 6.4.

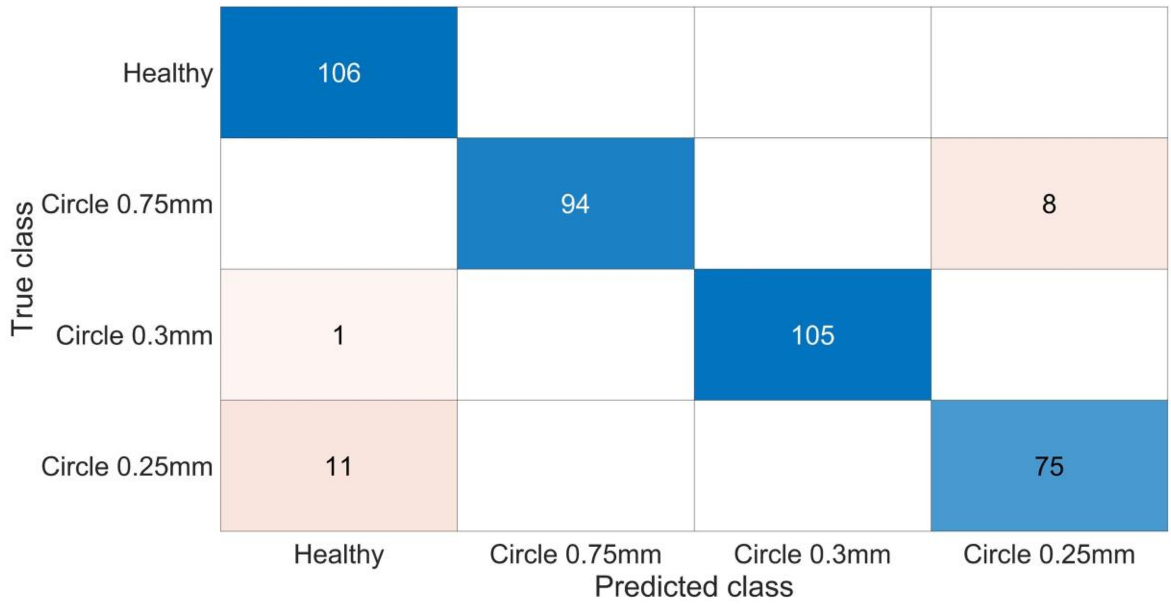


Figure 6.4 Confusion matrix for the testing classification of corrosion levels from the experiment

The results shown in Figure 6.4 demonstrate the significant effectiveness of the model in classifying corrosion levels from the experiment. In terms of the classification accuracy, that for 0.25 mm corrosion was worst because the corrosion is too small to be diagnosed. The classification results for 0.75 and 3 mm corrosion are acceptable. As mentioned in Chapter 2, based on the maintenance standard concerning corrosion on the underside of the rail [30], the rail should be replaced when the corrosion depth is over 0.75 mm. It is therefore not important to identify 0.25 mm corrosion compared with a healthy rail. The classification accuracy for classifying the corrosion levels was 95%, generally proving that the 1-D CNN model is able to identify significant levels of rail corrosion through experimental impact response signals.

In conclusion, this section demonstrates the effectiveness of the 1-D CNN model for the classification of different types and levels of corrosion on the underside of the rail against a healthy rail by using acoustic impact response signals. The performance of the 1-D CNN model was evaluated through four metrics, accuracy, F1-score, precision, and recall, and validated with 10 K-fold cross-validation. The confusion matrices indicate that the proposed method achieves high classification accuracy for corrosion types and corrosion levels, respectively. Overall, the processed impact response signal generated from the experiment can be used for identifying the type and level of corrosion of the underside of the rail.

In the next section, the dataset obtained from the simulation is classified through the 1-D CNN model. As for the classification in this section, the datasets of the corrosion type and level from the simulation will be classified against a healthy rail separately.

6.3 Fault classification of different corrosion types by simulated signals

In this section, simulated signals from the simulation model for different rail conditions (healthy or different conditions of corrosion) are used for the classification. Due to the simulated signal being generated without environmental noise, noise simulating field noise such as rotational machinal noise and wind noise was added into the simulated signal. To evaluate the capability of the 1-D CNN model, the simulated signals containing the different levels of noise power were evaluated.

Different levels of noise were added to the simulated signal based on the amplitude of the simulated signal, ranging from 0% to 100% of the signal amplitude with an increment of 10%.

6.3.1 Classification of corrosion types through simulated signals

In this section, the dataset obtained from the simulation model for different corrosion types and a healthy rail is classified, the same as for the experimental signals. The dataset containing the signals generated from a healthy rail, square corrosion, rectangular corrosion, and circular corrosion (0.75 mm) was labelled for the classification. The dataset of the simulated signal with different noise power was then trained by the 1-D CNN model. The model for classifying each simulated signal dataset was trained through the 1-D CNN model by selecting the optimal hyper-parameters (the hyper-parameter tuning technique). The selected ranges of hyper-parameters were based on Table 6.1.

As mentioned in Section 6.1.2, cross-validation (K-fold) was applied to evaluate the model's validation performance. After selecting the optimal hyper-parameters for the 1-D CNN model, the model's validation performance in classifying corrosion types and a healthy rail with different noise power was validated, as shown in Table 6.4.

Table 6.4 Validation performance of the 1-D CNN model in classifying corrosion types

Noise power	Accuracy (%)	F1-score (%)	Precision (%)	Recall (%)
0%	100.00	100.00	100.00	100.00
10%	100.00	100.00	100.00	100.00
20%	100.00	100.00	100.00	100.00
30%	100.00	100.00	100.00	100.00
40%	99.33	99.35	99.42	99.34
50%	99.89	99.89	99.90	99.89
60%	99.78	99.78	99.78	99.78
70%	99.78	99.78	99.78	99.78
80%	98.89	98.87	98.95	98.88
90%	99.33	99.33	99.33	99.34
100%	97.78	97.77	97.84	97.77

The results shown in Table 6.4 demonstrate the validation performance of the 1-D CNN model to classify types of corrosion on the simulated rail. The dataset with different noise power was analysed. The validation performance through the four metrics was shown by the average values. Considering the metric values of all folds, the standard deviation of the results was small (all standard deviations were lower than 0.05). Following the validation performance, the model was tested using previously unseen data, again with varying levels of noise added. The trend of the classification accuracy obtained using this new, unseen, data and with different noise powers is given in Figure 6.5.

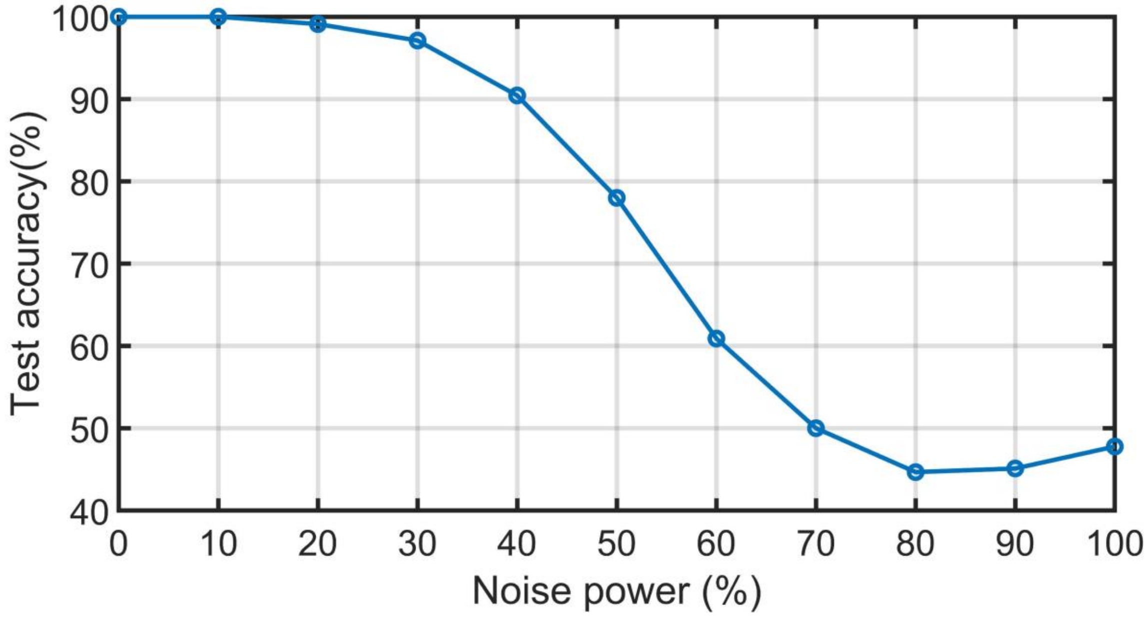


Figure 6.5 Trend of classification accuracy for corrosion types with different noise power

With an increase of noise power, the classification accuracy decreased. While the noise power was lower than 40% against the amplitude of the simulated signal, the values of the classification accuracy were all over 90%, proving that the 1-D CNN model can classify the corrosion type against a healthy rail from the simulation model through simulated signals containing a certain level of noise (less than 40%).

6.3.2 Classification of corrosion levels through simulated signals

After classifying the corrosion types against a healthy rail from the simulation model through the simulated signals, the corrosion level from the simulation model was then classified against a healthy rail through the 1-D CNN model. Different levels of circular corrosion and a healthy rail were applied for the analysis. Firstly, the optimal hyper-parameters for the 1-D CNN model were selected through the hyper-parameter tuning technique. The validated performance of classification of corrosion levels against a healthy rail with different noise power is shown in Table 6.5.

Table 6.5 Validation performance of the 1-D CNN model in classifying corrosion levels

Noise power	Accuracy (%)	F1-score (%)	Precision (%)	Recall (%)
0%	100.00	100.00	100.00	100.00
10%	100.00	100.00	100.00	100.00
20%	100.00	100.00	100.00	100.00
30%	100.00	100.00	100.00	100.00
40%	99.83	99.83	99.83	99.83
50%	99.00	99.00	99.14	99.00
60%	98.69	98.86	98.88	98.69
70%	98.74	98.73	98.73	98.74
80%	98.35	98.57	98.55	98.35
90%	99.22	99.22	99.22	99.22
100%	96.98	96.99	96.84	96.98

The results in Table 6.5 demonstrate the validation performance of the 1-D CNN model in classifying corrosion levels of the simulated model. Similar to the classification of the corrosion types, the validation performance of the 1-D CNN model in classifying corrosion levels through the four metrics was shown by the average values. And considering the metric values for all folds, the value of the standard deviation was also small (all standard deviations were lower than 0.04). Following the validation performance, the model was tested using previously unseen data, again with varying levels of noise added. The trend of the classification accuracy with different noise powers is given in Figure 6.6.

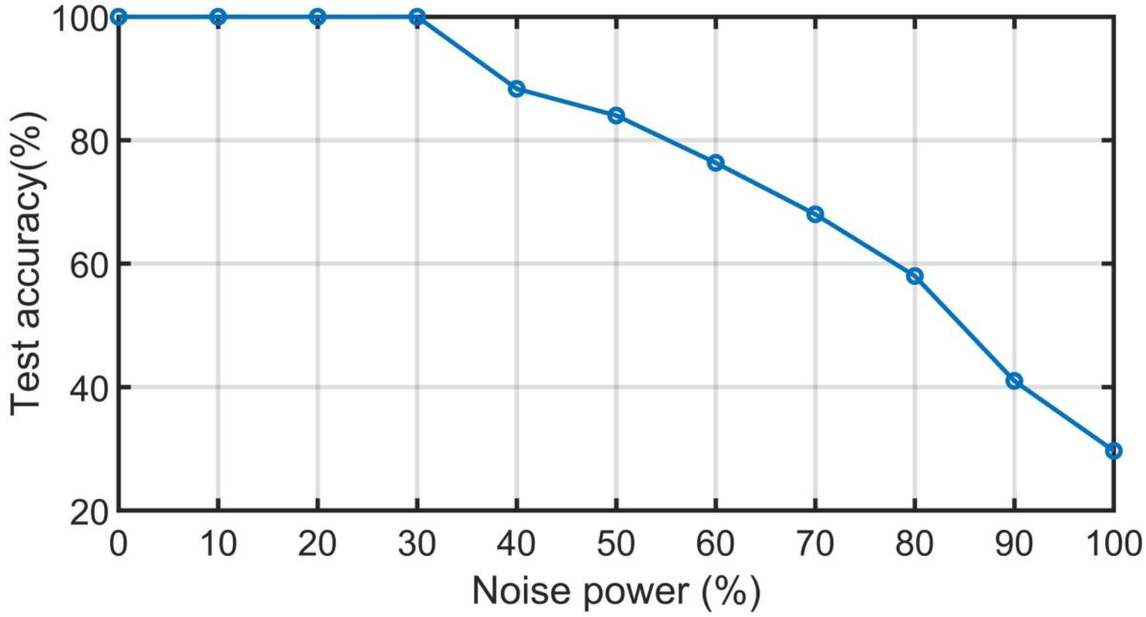


Figure 6.6 Trend of classification accuracy for corrosion levels with different noise power

The results in Figure 6.6 show that the accuracy was over 80% while the noise was lower than 50% of the amplitude of the simulated signal. As for the classification of corrosion type, with an increase of noise power, there was a decrease in the accuracy of classifying corrosion levels of the simulated model. The analysis indicates that the 1-D CNN is able to achieve significant performance in classifying the corrosion levels of the simulated model through the simulated signals containing noise of certain levels (less than 50%).

In conclusion, this section evaluates the effectiveness of the 1-D CNN model for identifying corrosion types and corrosion levels against a healthy rail separately from the simulation model. Simulated field noise was added to the simulated signals generated from the simulated model to create different datasets. The 1-D CNN model was trained to classify each simulated signal dataset, and hyper-parameter tuning was used to select the optimal hyper-parameters. The results show that the 1-D CNN model is able to classify the corrosion type or level against a healthy rail with high accuracy when the noise power is less than 40% of the amplitude of the simulated signal. However, with an increase in noise power, the accuracy of the classification decreases. Hence, the classification results prove that with a certain degree noise power, the 1-D CNN model is able to accurately classify corrosion types and levels separately through simulated signals.

In the next section, classification of the type and level of corrosion on the underside of the rail will be combined. The type and level will be classified simultaneously against a healthy rail. In order to obtain better accuracy, an integrated 1-D CNN model will be introduced for the classification.

6.4 Fault classification of rail base corrosion by an integrated 1-D CNN model

The contents this section have been reviewed and accepted in the journal MDPI Applied Science ('1-D CNN Based Detection and Localisation of Defective Droppers in Railway Catenary'). This paper introduces the structure of the integrated 1-D CNN model and reviews its advantages compared with the conventional 1-D CNN model. From this publication, it can be understood that the integrated 1-D CNN model can be applied for the classification of corrosion types and corrosion levels simultaneously.

6.4.1 Introduction of the integrated 1-D CNN model

In above sections, the types and levels of corrosion of the experimental rail and the simulation model are classified separately. However, if considering the problem of using 1-D CNN to classify two different fault categories simultaneously, it is found that a conventional machine learning technique is not efficient [223]. Additionally, the hyper-parameters selected for different 1-D CNN models (classifying the corrosion type/level) are changeable. When using a 1-D CNN model to classify the corrosion type and level simultaneously, it is hard to select the hyper-parameters to gain best performance for these two conditions. Hence, it is suggested to integrate separate 1-D CNNs for the corrosion type and the corrosion level. The multiple objectives for classification (corrosion types and levels) require different hyper-parameters to be selected in the CNN model in order to achieve optimal performance. An independent CNN model cannot be used to complete these two different classifications.

In this section, a method based on integrating multiple 1-D CNNs (i.e. an integrated 1-D CNN) is applied to classify the types and levels of corrosion of the rail base as two independent tasks. The structure of the integrated 1-D CNN model is shown in Figure 6.7. Specifically, Sub-model 1 considers the types of rail base corrosion defined by the experiment (which were introduced in Chapter 4) and Sub-model 2 considers the level of rail base corrosion. It should be noted that when the rail is identified as being healthy, it is not necessary to identify its corrosion level. Hence, Sub-model 2 is applied to train the part of the dataset which only contains the data labelled as corrosion categories (i.e., data not labelled as ‘healthy’). Additionally, for square and rectangular corrosion, only damage of 0.75 mm depth is considered. Hence, when data are identified as square or rectangular corrosion, they are not transferred to Sub-model 2 for classifying the corrosion level.

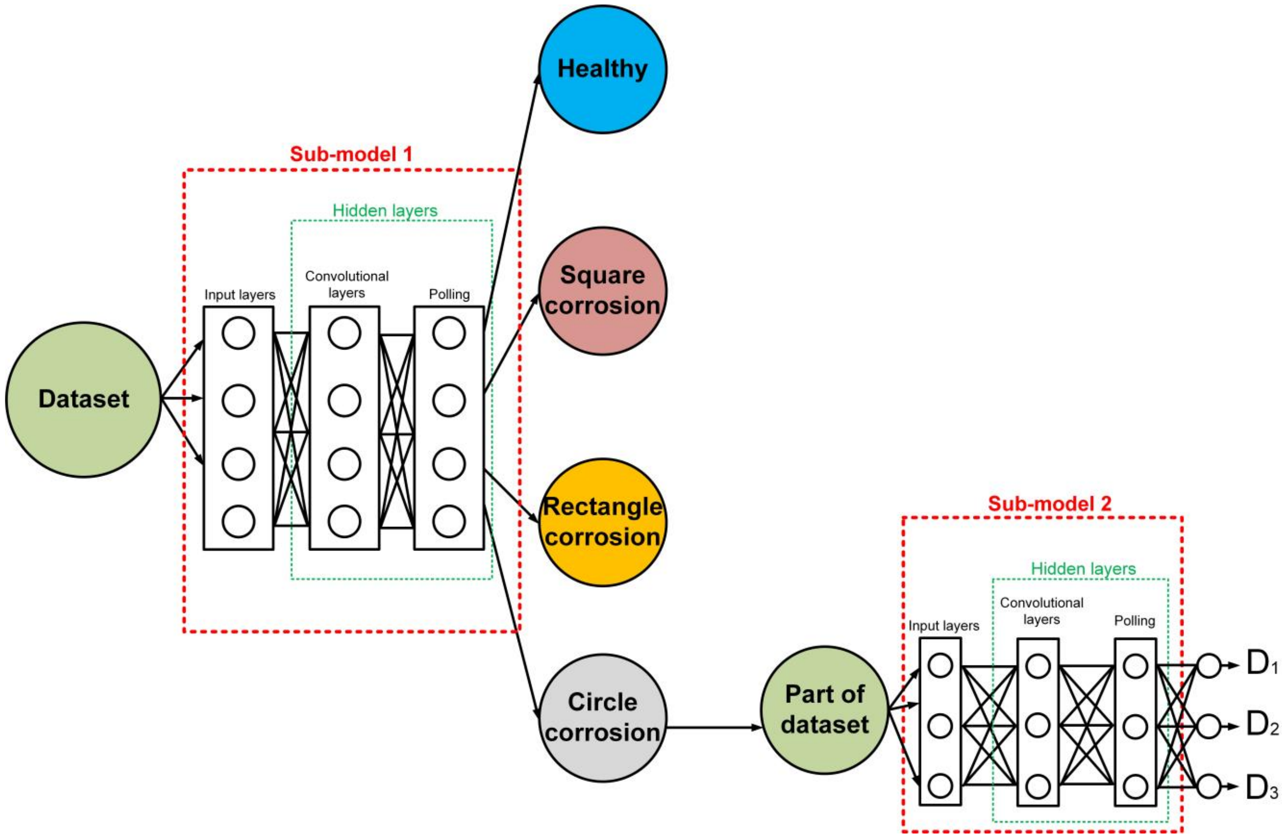


Figure 6.7 Integrated 1-D CNN model

To integrate two trained sub-models, a method for transferring the output result from Sub-model 1 to Sub-model 2 is proposed. The result from Sub-model 1 is given as a set of predicted probabilities corresponding to corrosion types. The probabilities of a 'healthy' label or different corrosion labels (corrosion types) are given as P1, P2, P3, and P4, respectively. P_{max} is the maximum value of P1, P2, P3, and P4. If P_{max} corresponds to the 'healthy' condition (P1), square corrosion (P2), or rectangular corrosion (P3), the process will stop and the predicted output of Sub-model 1 will be saved. If P_{max} corresponds to circular corrosion (P4), it will be fed into Sub-model 2 for further prediction. The predicted output of Sub-model 2 represents the level of rail corrosion. The testing performance of the integrated CNN model is the product of the predicted results of Sub-model 1 and Sub-model 2.

Accordingly, the proposed method can identify whether the rail is corroded and further identify the level of this corrosion type. The advantage of this integrated 1-D CNN model is that it is able to obtain and use the best hyper-parameters of both individual models (Sub-model 1 and Sub-model 2), while still producing a single classification. Additionally, it can simplify the dataset for each classification model. Specifically, the dataset for Sub-model 2 only now contains three labels (three different levels of circular corrosion). Compared with the independent level classification model, the data corresponding to the healthy label are removed and hence it is no longer possible to misclassify the level of a circular corrosion fault as healthy.

To demonstrate how the data are processed by the integrated 1-D CNN, a data flow diagram is shown in Figure 6.8.

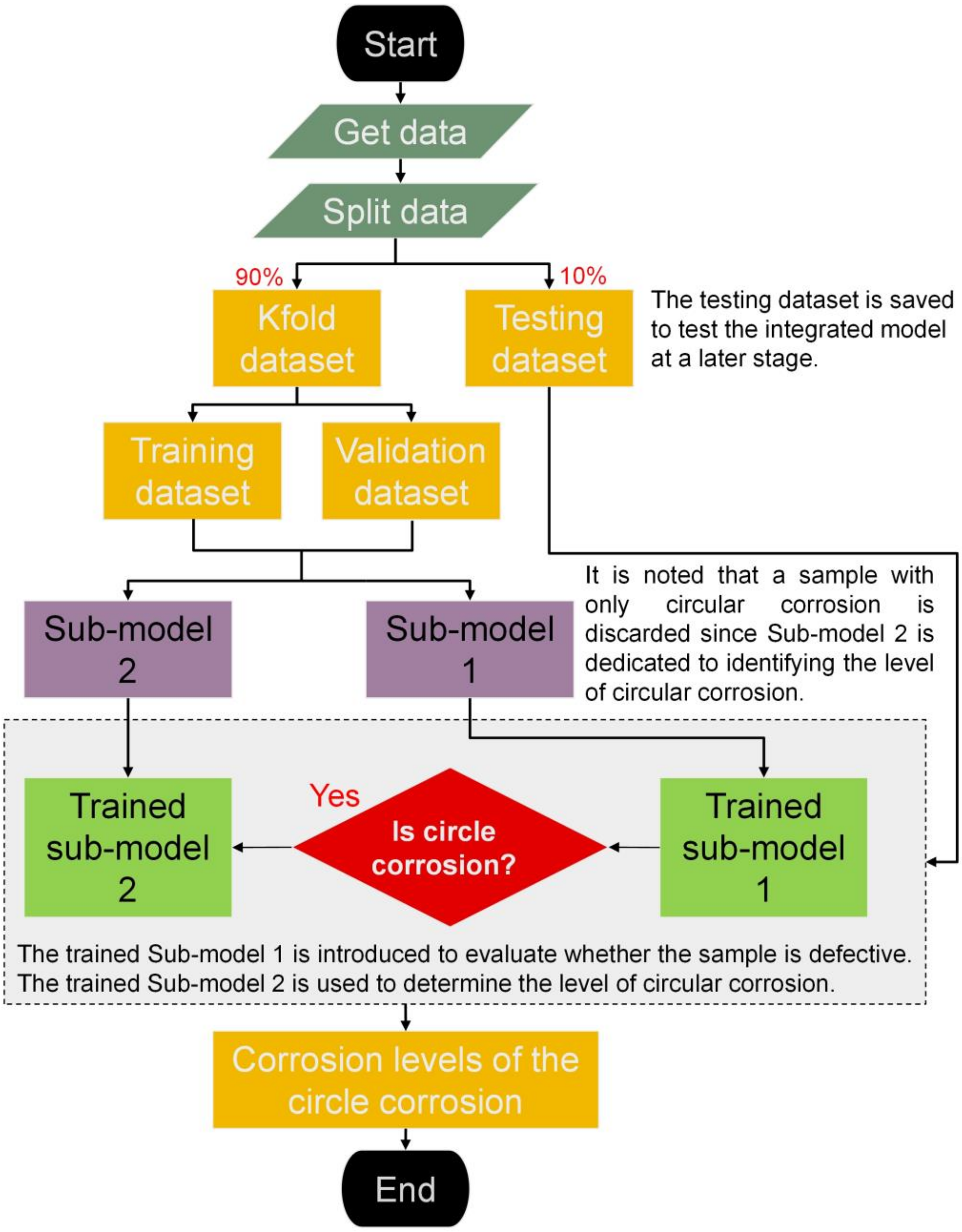


Figure 6.8 Data flow diagram

As can be seen from Figure 6.8, the dataset was split into two sets: a K-fold dataset and a testing dataset. The K-fold dataset was utilised to train and validate the two sub-models. After this, the performance of the integrated model consisting of two fine-tuned sub-models was appraised using the testing dataset.

The Z-score [224] was applied for normalising the dataset. It is notable that the datasets for validation and testing (validation and testing datasets) were normalised based on the training dataset. Hence, during the whole process of classification, the testing dataset for the integrated 1-D CNN model was not used for training and validation of the sub-models. This avoids data leakage and proves the authenticity of the validation results.

In the next section, to illustrate the performance of the integrated 1-D CNN model, the datasets obtained from the experiment and simulation are applied for classification.

6.4.2 Fault classification of rail base corrosion through experimental signals

In this section, the classification of the experimental results was completed. The corrosion types and levels were classified against a healthy rail through the integrated 1-D CNN model simultaneously. As introduced in Section 6.4.1, Sub-model 1 was applied for classifying the corrosion types and a healthy rail. Hence, the three different levels of circular corrosion (0.25, 0.75, and 3 mm) were labelled as being in the same category for the classification.

In order to demonstrate the advantages and effectiveness of the integrated 1-D CNN model, a conventional 1-D CNN model for classifying the corrosion types and corrosion levels against a healthy rail simultaneously was applied for comparison.

Firstly, two sub-models of the integrated 1-D CNN model and the conventional 1-D CNN model were trained to select optimal hyper-parameters according to the hyper-parameter tuning technique. The optimal hyper-parameters of three models are given in Table 6.6.

Table 6.6 Optimal hyper-parameters of the integrated and conventional 1-D CNN models

Optimal hyper-parameters of Sub-model 1	CNN layer	2
	Number of filters in the first convolutional layer	480
	Weight initialisation	1.0211×10^{-5}
	Learning rate	4.2800×10^{-4}
	Number of filters in the dense layers	96
	Kernel_initialiser	Glorot uniform
	Activation function	Tanh
Optimal hyper-parameters of Sub-model 2	CNN layers	2
	Number of filters in the first convolutional layer	224
	Weight initialisation	3.4762×10^{-5}
	Learning rate	1.8075×10^{-5}
	Number of filters in the dense layers	128
	Kernel_initialiser	Glorot uniform
	Activation function	ReLU
Optimal hyper-parameters of conventional 1-D CNN	CNN layers	1
	Number of filters in the first convolutional layer	320
	Weight initialisation	7.0016×10^{-3}
	Learning rate	1.6961×10^{-4}
	Number of filters in the dense layers	480
	Kernel_initialiser	He normal
	Activation function	ReLU

After training the conventional and integrated 1-D CNN models, the average validation performance assessed through the four metrics for these two models (including Sub-model 1, Sub-model 2) was determined as shown in Table 6.7 (the value of the four metrics was obtained from the average results of 10 K-folds).

Table 6.7 Validation performance of the integrated 1-D CNN model for experimental results

Validation Performance metric	Integrated 1-D CNN model		Conventional 1-D CNN model
	Sub-model 1	Sub-model 2	
Accuracy (%)	99.72	99.44	99.44
F1-score (%)	99.58	99.45	99.45
Precision (%)	99.59	99.48	98.50
Recall (%)	99.58	99.44	98.45

The results for the integrated and conventional 1-D CNN models show that these two models both achieved great classification performance of over 99% for all the validation metrics. To demonstrate the classification accuracy of these two models, a confusion matrix was used for the analysis. The confusion matrix for the conventional 1-D CNN model is shown in Figure 6.9.

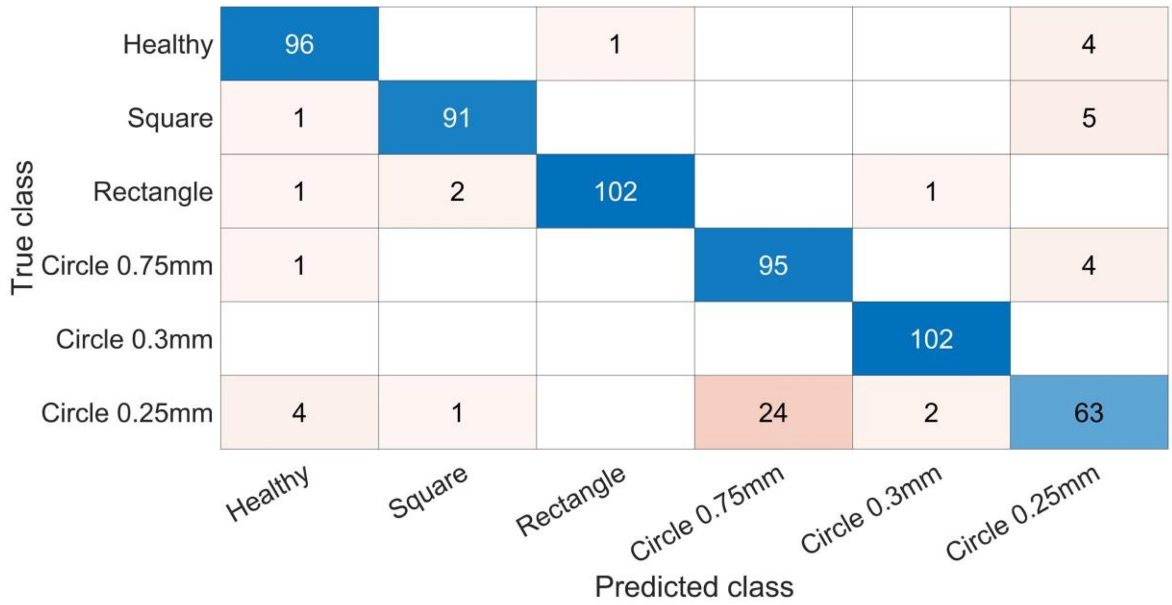
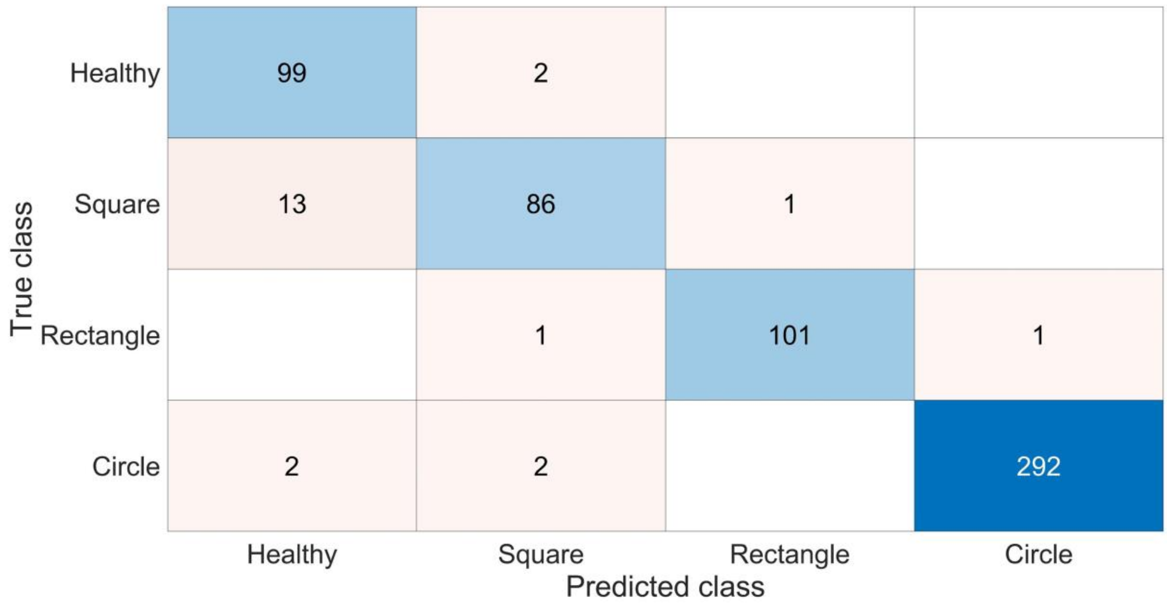
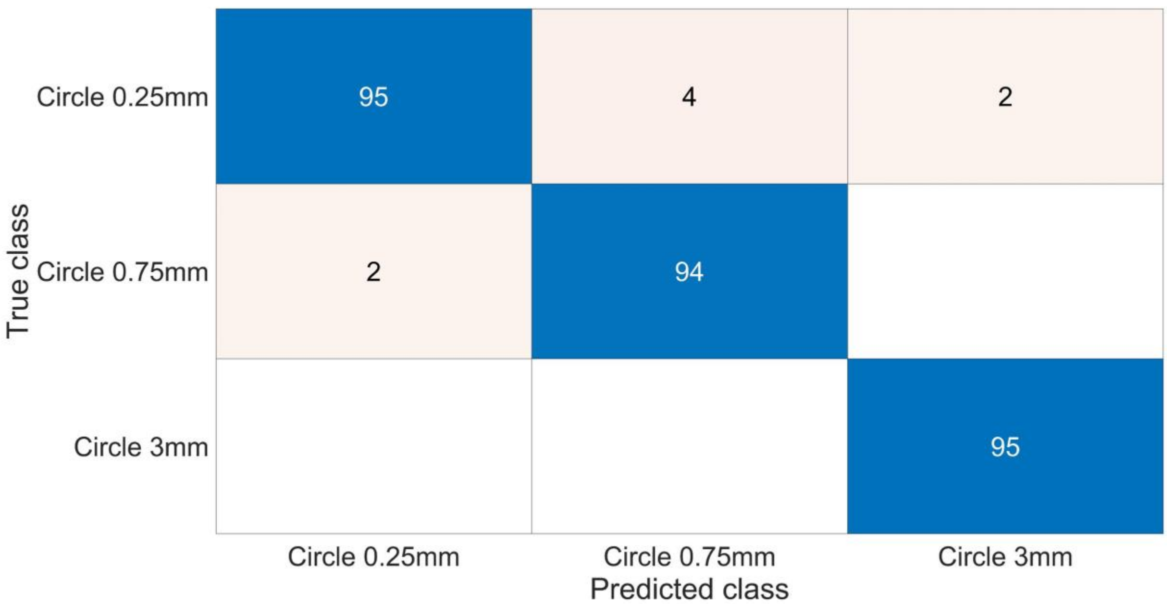


Figure 6.9 Confusion matrix for the conventional 1-D CNN model

The results shown in Figure 6.9 demonstrate the accuracy of the conventional 1-D CNN model in classifying the corrosion type and level against a healthy rail simultaneously. Compared with two independent classification results (classification of corrosion type and level), the classification results became worse, especially for classifying the three levels of circular corrosion. The accuracy of classifying 0.25 mm circular corrosion was only 67%. The testing classification of the integrated 1-D CNN model is shown as a confusion matrix in Figure 6.10.



a. Confusion matrix for Sub-model 1 of the integrated 1-D CNN model



b. Confusion matrix for Sub-model 2 of the integrated 1-D CNN model

Figure 6.10 Confusion matrix for the integrated 1-D CNN model

The results shown in Figure 6.10 demonstrate the testing performance of the integrated 1-D CNN model. Figure 6.10a demonstrates Sub-model 1's accuracy in classifying different corrosion types; Fig. 6.10b shows Sub-model 2's performance in classifying the corrosion levels. The accuracy in Fig. 6.10b was obtained from the predicted data (those 292 data in Fig. 6.10a) from Sub-model 1. The testing performance of the two sub-models was significant, especially the improved accuracy of classifying 0.25 mm circular corrosion compared with the result for the conventional 1-D CNN model. The reason is that the dataset trained for Sub-model 2 only contains the three categories of circular corrosion without the healthy categories. Compared with the conventional 1-D CNN model, these three categories won't be classified as healthy when classifying the corrosion levels. In particular, there is a strong possibility that 0.25 mm circular corrosion will be recognised as healthy, according to the conclusion drawn in Section 6.3.2 and result shown in Figure 6.9. Hence, the integrated 1-D CNN model can enhance the classification accuracy by avoiding these error classifications.

A comparison of the classification accuracy of the conventional 1-D CNN model and the integrated 1-D CNN model is given in Table 6.8.

Table 6.8 Comparison of classification accuracy of the two 1-D CNN models

1-D CNN model	Classification accuracy (%)
Conventional 1-D CNN model	91.11
Integrated 1-D CNN model	93.13

According to the classification accuracy in Table 6.8, it can be seen that the 1-D CNN model is able to classify the rail base corrosion from the experiment, displaying significant performance. The integrated 1-D CNN model performs better than the conventional 1-D CNN model, proving that the integrated 1-D CNN model provides enhanced classification accuracy when classifying two independent categories (corrosion type and level) simultaneously.

After classifying the corrosion faults from the experiment, the fault classification of rail corrosion from the simulated model was completed through the integrated 1-D CNN model.

6.4.3 Fault classification of rail base corrosion through simulated signals

In this section, classification using the simulation model is demonstrated. As the simulated signal was generated without environmental noise, noise simulating field noise such as rotational machinal noise and wind noise was added into the simulated signal, as for the analysis in Section 6.2. In order to evaluate the capability of the integrated 1-D CNN model and the conventional 1-D CNN model, simulated signals containing different levels of noise power were evaluated. The datasets with different noise power were trained by an individual integrated 1-D CNN model and an individual conventional 1-D CNN model.

The validated performance of the integrated 1-D CNN model with the selected optimal hyper-parameters (including Sub-model 1, Sub-model 2) for classifying the simulated dataset is given in Table 6.9.

Table 6.9 Validation performance of the integrated 1-D CNN model for simulation results

Noise power	Sub-model 1		Sub-model 2	
	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
0%	100.00	100.00	100.00	100.00
10%	100.00	100.00	100.00	100.00
20%	100.00	100.00	100.00	100.00
30%	100.00	100.00	99.41	99.41
40%	99.78	99.78	99.41	99.41
50%	99.19	99.14	98.82	98.81
60%	98.22	98.06	98.97	98.97
70%	97.37	97.19	97.21	97.21
80%	96.41	96.06	98.09	98.09
90%	96.48	96.07	96.91	96.65
100%	94.85	94.20	93.09	92.81
Noise power	Sub-model 1		Sub-model 2	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
0%	100.00	100.00	100.00	100.00
10%	100.00	100.00	100.00	100.00
20%	100.00	100.00	100.00	100.00
30%	100.00	100.00	99.44	99.40
40%	99.75	99.82	99.43	99.41
50%	99.43	98.92	98.85	98.82
60%	98.44	97.87	99.02	98.97
70%	97.39	97.01	97.22	97.21
80%	96.18	95.90	98.21	98.10
90%	96.52	95.69	97.14	96.82
100%	94.70	93.78	93.27	93.00

The validation performance of the conventional 1-D CNN model with the selected optimal hyper-parameters is given in Table 6.10.

Table 6.10 Validation performance of the conventional 1-D CNN model for simulation results

Noise power	Accuracy (%)	F1-score (%)	Precision (%)	Recall (%)
0%	100.00	100.00	100.00	100.00
10%	100.00	100.00	100.00	100.00
20%	100.00	100.00	100.00	100.00
30%	100.00	100.00	100.00	100.00
40%	99.33	99.35	99.42	99.34
50%	99.93	99.92	99.93	99.92
60%	96.59	96.49	96.80	96.51
70%	97.11	97.10	97.41	97.07
80%	98.81	98.82	98.90	98.82
90%	95.56	95.43	95.62	95.49
100%	93.33	93.33	93.70	93.27

The results in Table 6.10 indicate that the values of the four validation performance metrics get worse when the noise power increases, which is the same conclusion as drawn in Section 6.3 when classifying the corrosion types and corrosions levels separately. But the values of these four metrics are still over 93%. Hence, the validation performance of both models is still good.

The trend of the classification accuracy of the two 1-D CNN models with increased noise power is given in Table 6.11 and Figure 6.11.

Table 6.11 Classification accuracy of the two 1-D CNN models with increased noise power

Noise power	Classification accuracy (%)	
	Conventional 1-D CNN model	Integrated 1-D CNN model
0%	100.00	100.00
10%	100.00	100.00
20%	99.11	98.71
30%	94.00	97.00
40%	86.89	90.00
50%	78.00	80.50
60%	66.89	75.53
70%	59.10	64.38
80%	46.89	63.77
90%	45.11	55.19
100%	42.33	50.25

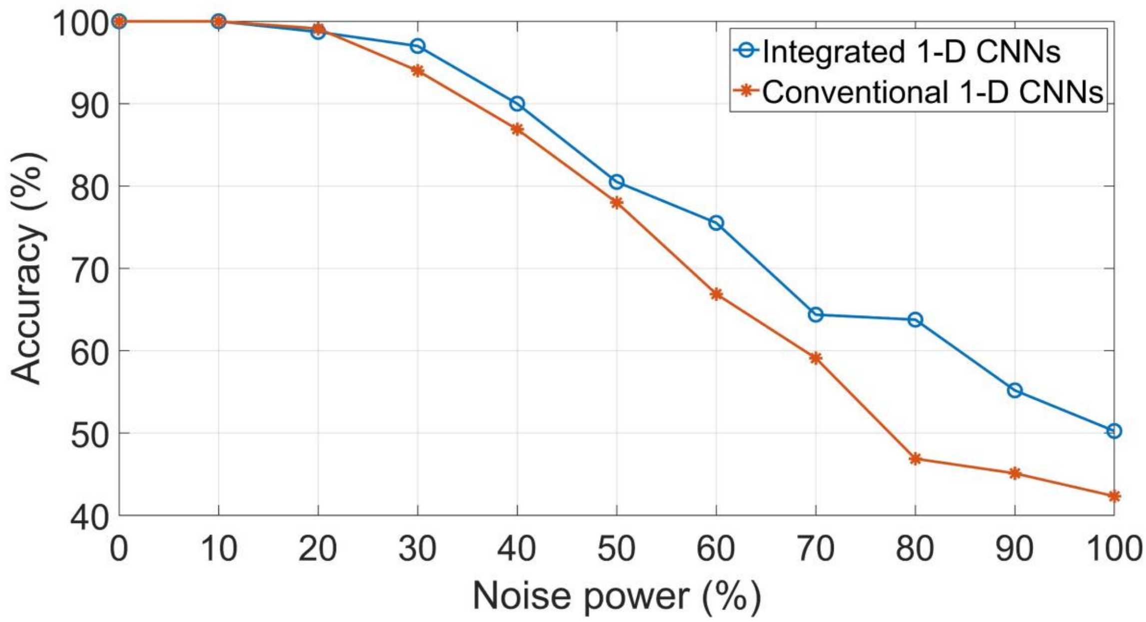


Figure 6.11 Trend of classification accuracy of the two 1-D CNN models with increased noise power

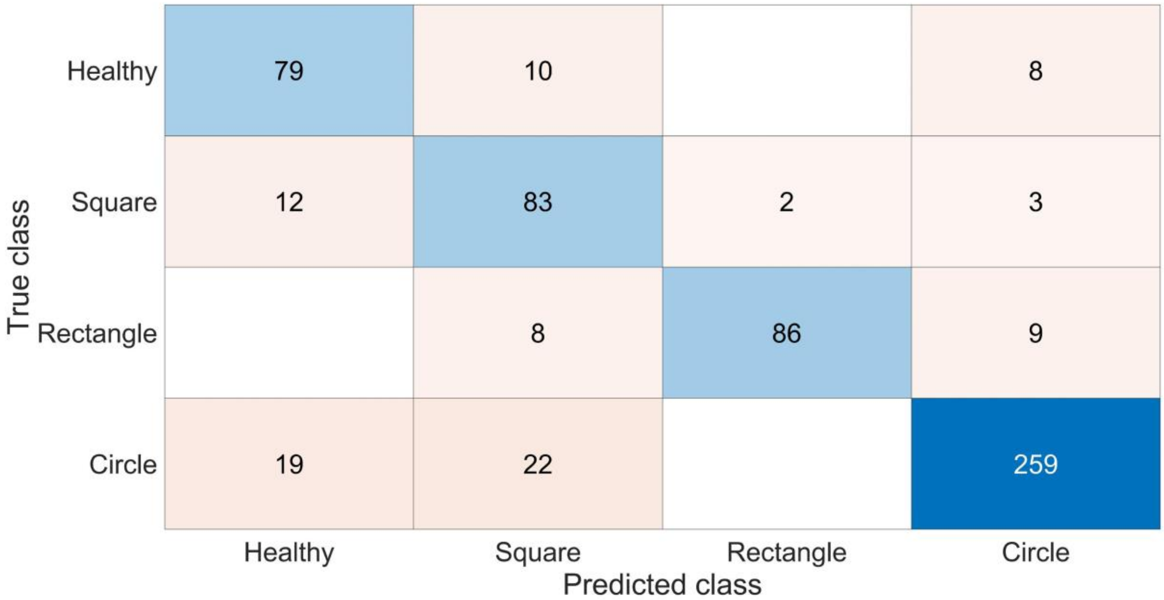
The results in Figure 6.11 indicate that the classification accuracy of the integrated 1-D CNN model is better than that of the conventional 1-D CNN model for almost all noise power levels. The conventional 1-D CNN model only had a better accuracy result (very small difference and less than 0.5%) with 20% power. Similar to the conclusion drawn in Section 6.3, good classification accuracy was achieved when the noise power was lower than 50%. The classification accuracy of the integrated CNN model was over 80% when the noise power was lower than 50%. This indicates that the 1-D CNN model is able to accurately classify corrosion types and levels simultaneously through simulated signals.

Additionally, the computational cost is a factor of concern. In this research, the computation was based on the Google Colaboratory platform with a P100 GPU. By training the conventional and integrated 1-D CNN models with their own optimal hyper-parameters, the computational costs from the two kinds of model (the integrated CNN model and the conventional 1-D CNN model) were compared. The computational time of the integrated 1-D CNN model was increased by around 2% to 3% in terms of the different datasets. But considering that accuracy is more important for corrosion classification, the integrated model showed better performance. And when the noise power is increased, the classification accuracy of the integrated 1-D CNN model is better than that of the conventional 1-D CNN model. Nevertheless, as mentioned in the conclusion in Section 6.4.2, the integrated 1-D CNN model can enhance the classification accuracy for smaller areas of corrosion. Hence, the effectiveness of the integrated model was proved again.

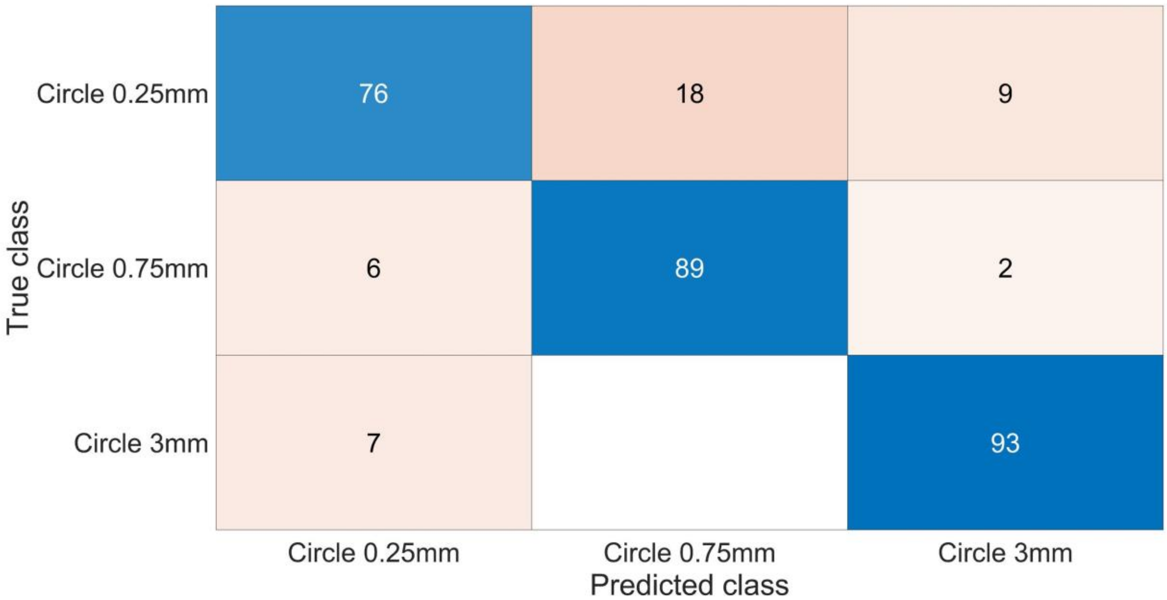
6.4.4 Fault classification of rail base corrosion through the combination of simulated and experimental signals

According to the above research, it has been proved that the integrated 1-D CNN model could be used for the classification of corrosion on the underside of the rail. In this section, the simulated and experimental signals are mixed as a whole dataset for the classification. This aims to explore the feasibility of using the 1-D CNN model trained by simulated data for identifying corrosion types and levels (or that a rail is healthy) in the real world (i.e., based on experimental data).

Firstly, the simulated data were applied to train the integrated 1-D CNN model. This process is the same as for training and validating the integrated 1-D CNN model for the simulated dataset in Section 6.3. The experimental dataset was then used for testing this integrated 1-D CNN model. The chosen simulated dataset was the simulated signal without noise (0% noise). After testing the integrated 1-D CNN model through the experiment dataset, the classification results were presented by a confusion matrix, as shown in Figure 6.12.



a. Confusion matrix for Sub-model 1 of the integrated 1-D CNN model



b. Confusion matrix for Sub-model 2 of the integrated 1-D CNN model

Figure 6.12 Confusion matrix for the integrated 1-D CNN model for classifying the experimental results

As shown in Figure 6.12, the performance was worse than for previous classifications: the classification accuracy of Sub-model 1 was 84.50% and that of Sub-model 2 was 86.00%. The total classification accuracy of the integrated 1-D CNN model was 72.67%. The reason for the decrease in classification accuracy is that, although the similarity of the simulated and experimental signals had been proved, there were still differences between these two datasets. This led to more errors in classification when using the integrated 1-D CNN model. Specifically, from Figure 6.12a it can be seen that some other data were identified as circular corrosion by mistake, and these data were predicted by Sub-model 2. Figure 6.12b shows that the best accuracy result was obtained for classification of 3 mm circular corrosion. The worst accuracy was still obtained for 0.25 mm corrosion, as in Sections 6.4.2 and 6.4.3. If the results for classification of 0.25 mm corrosion are removed (as this level of defect is not considered a fault within the standards [30]), the accuracy of Sub-model 2 increases to 92.39% and that of the integrated 1-D CNN model increases to 78.06%. The analysis indicates that the data generated from the simulation model can identify most of the corrosion conditions from the experiment, proving that it could be used for identifying corrosion on the underside of the rail in the real world.

6.5 Conclusion

In this section, corrosion on the underside of the rail was classified through a machine learning-based method. A 1-D CNN model was selected as the machine learning-based method for the classification. Firstly, the key parameters and processes of the model, such as cross-validation, K-fold validation, and hyper-parameter tuning technique, were introduced to ensure the best classification results. Secondly, considering the data from the experiment and the simulation model, the corrosion types and corrosion levels from each kind of dataset were classified against a healthy rail separately through two individual 1-D CNN models. The results showed that the 1-D CNN model is effective in classifying corrosion types and levels from both experimental and simulation data. Thirdly, the types and levels of corrosion from the experiment and the simulation model were classified against a healthy rail simultaneously. An integrated model was designed to obtain better classification results. The classification results showed that the designed integrated simulated model performs better than the conventional 1-D CNN model, particularly for minor corrosion (0.25 mm depth). The acoustic impact response signal was proven to identify the corrosion on the underside of the rail, and different corrosion types and corrosion levels were classified. The integrated 1-D CNN model trained by the simulation results was then used to simultaneous classifying the corrosion type and corrosion level from the experimental data. The classification results show that the trained 1-D CNN model could be used effectively for classifying corrosion on the underside of the experiment rail and also prove that the simulated signal generated from the simulation model can be used for identifying corrosion on the underside of the rail from the real world.

In conclusion, according to the classification through the integrated 1-D CNNs model, it can not only identify whether there is a corrosion but also identify different its corrosion types and corrosion levels. The integrated 1-D CNNs model designed in this Chapter can classify the corrosion types and corrosion levels simultaneously. The advantage of the integrated 1-D CNNs model is that it is able to obtain and use the best hyper-parameters of both individual models (Sub-model 1 and Sub-model 2), while still producing a single classification. Compared with the effectiveness of conventional 1-D CNNs model, the integrated 1-D CNNs model obtained better classification results. Especially while classifying the faults with stronger noise, the effectiveness of the integrated 1-D CNNs model is more significant.

Additionally, it should note that the dataset of using training the 1-D CNNs model is based on the simulated faults. It contains 3 different fault types based on approximated corrosion shapes (square, rectangular, and circular corrosion). The corrosion levels of the circular corrosion are separated as 3 different levels and with a healthy condition. While it is accepted that these are only approximate faults based on simplified shapes, the system is shown to work with both different fault shapes and levels suggesting good differentiation that would likely translate to variations in real faults. In the future work section, in order to identify the real corrosion on the railway, it only needs to design different geometry of the faults on the rail bottom (more realistic corrosion, more corrosion levels) in the simulation model base on the corrosion conditions on the practical scenario. And according to training and validating the simulated signals generated from the updated simulated model, testing the experimental signals obtained from the field test, it can identify the corrosion types and corrosion levels from the realistic world.

CHAPTER 7 CONCLUSION AND FUTURE WORKS

This thesis contributes to research into the inspection of corrosion on the underside of railway rails using acoustic techniques. Specifically, this research supports the author's hypothesis, and proves that it is possible to automatically detect and quantify corrosion of the underside of the rail by acoustically inspecting an impact response signal. The research questions presented in this thesis have been answered through the research subsequently discussed. In this section, conclusions are drawn and the contributions to research are summarised. Consideration is also given to potential future activities that may further develop this line of research.

7.1 Conclusion

This thesis aimed to develop a novel approach for the detection and classification of the condition of railway corrosion, specifically the underside corrosion of rails, through the application of signal processing, machine learning, and simulation. The research gap was identified through a literature review, which introduced different approaches for railway inspection and highlighted the potential of vibration and acoustic signals for detecting underside corrosion.

The initial methodology and key techniques were then presented. These include a process for the coupling of multi-physical fields for modelling and simulation, the adaptation and application of appropriate signal decomposition algorithms, and the implementation of an integrated multi-stage machine learning model for corrosion classification.

Chapter 4 details the creation and validation of a Finite Element Method (FEM) simulation model to obtain an acoustic impact response signal for different rail geometries representing different rail fault types. The similarity of the simulated signals to the experimental signal was evaluated based on the similarity of the frequency peaks, variance of energy, and correlation coefficients. After these validations, the simulation model was considered an appropriate tool to be used to generate an enlarged dataset of simulated signals, as required for further analysis.

Chapter 5 focuses on signal decomposition and feature extraction. Firstly, a laboratory experiment was introduced for collecting experimental signals. Secondly, the experimental signal obtained from the laboratory experiment was processed by different signal decomposition methods identified from the literature as likely to be suitable (EMD, EEMD, VMD, and LMD). The most suitable decomposition method was selected based on analysis of the energy-based results obtained using these four decomposition methods. The analysis indicated that VMD is the most effective for processing impact response signals. Thirdly, the VMD technique was enhanced using the newly proposed IVMD-SVD method, with effectiveness demonstrated through evaluation of the entropy results. Finally, an algorithm using SVD and DTW was proposed for removing potential data errors. The impact response signal was processed through the most suitable classification method; the processed signals constituted the dataset for the classification in the following chapter.

Chapter 6 initially presents results of the application of independent 1-D CNN models for the classification of corrosion fault types and levels of severity. The techniques were shown to be effective when applied both to experimental data and those generated from the numerical models presented. The different corrosion types and corrosion levels were then classified simultaneously using an integrated multi-level 1-D CNN model that was designed to achieve better classification performance than either independent classifier or a single combined classification model. The results from the integrated classifier demonstrated that acoustic impact response signals are an appropriate source with which to identify corrosion on the underside of the rail and to identify and quantify different corrosion types and levels. Finally, the integrated multi-level 1-D CNN model that had been trained using signals from the simulation model was used to simultaneously classify corrosion types and corrosion levels from acoustic impact response signals obtained from experimental data. The results indicated that the trained 1-D CNN model could be used effectively for the classification of underside corrosion from the experiment rail; they also proved that the simulated signal generated from the simulation model can be used for identifying corrosion on the underside of the rail from the real world.

Overall, the thesis demonstrates the potential of building a simulation model, optimising the signal processing technique, and applying a machine learning method to develop a new, non-invasive, approach for the inspection of railway corrosion, specifically underside corrosion of rails through an acoustic technique. The approach can potentially be used by railway maintainers for real time monitoring and evaluation of the health condition of the rail. Maintainers can safely deploy microphones on the wayside and collect the acoustic signals generated at positions of interest, where corrosion is likely to occur. The concept of using impact response signals to generate suitable acoustic signals has been demonstrated, and a potential for generation of such signals by vehicles passing Insulated Rail Joints (IRJs) or crossing noses has been identified. The maintainer can thus use the proposed approach to evaluate the health condition of the rail to identify whether underside corrosion exists at a target location.

7.2 Contribution to knowledge

The key contribution of this thesis is the use of acoustic techniques for inspecting for corrosion on the underside of rails. The literature review identified that impact response signals could potentially be used for such inspection, but that a number of developments in data management, signal processing, and classification would be required. Hence this thesis includes three sub-investigations each with their own contributions:

In Chapter 4, a method for creating simulated impact response signals for different rail geometries using an FEM simulation model is presented. The accuracy of the simulation model was validated by comparing the frequency response function and the similarity of simulated signals to laboratory test results. As obtaining sufficient experimental data for certain classification and analysis techniques is known to be challenging, it was demonstrated that this model could contribute through the generation of acoustic impact response signals representative of a hammer (or passing rolling stock) impact on the railway. The simulated signals generated were used to create a simulated dataset for further analysis. The use of simulated datasets was subsequently shown to reduce the need for extensive and costly laboratory experiments and hence allow for more efficient and faster development of fault identification methods.

Chapter 5 presents research demonstrating a contribution in the form of a method for processing impact response signals in order to extract information related to railway track structure and health, while eliminating environmental interference. An algorithm is proposed to automatically select the best initial parameters to be applied with the VMD technique to eliminate external influences, and to improve consistency, when decomposing impact response signals. The improved VMD algorithm contributes by automatically maximising the features extracted from the impact response signals and hence supporting further classification activities. Additionally, the methods and metrics used in the comparison of four decomposition methods, as presented in this chapter, may be used as a reference for other researchers focusing on processing impact response signals.

In Chapter 6, the research demonstrates three key contributions. Firstly, a machine learning-based method for the classification of corrosion on the underside of the rail. Secondly, a demonstration that an integrated multi-level 1-D CNN model can be more effective than both independent models and a single model for the simultaneous classification of corrosion types and levels. Thirdly, Chapter 6 demonstrates that a simulation model such as the one designed in Chapter 4 can be used to train classifiers for identifying and quantifying underside corrosion in practically obtained experimental data such as those which may be obtained from an actual railway.

7.3 Limitations of the work

The inspection of corrosion on the underside of rails through acoustic techniques has been researched and the fundamental concepts required are described in this thesis. However, further development is still required relating to limitations identified due to various factors encountered during the research, which can be summarised as follows:

1) Concerning corrosion faults, the corrosion geometries were designed as simple shapes (square, rectangle, and circle), which differ from real-world corrosion. Additionally, in accordance with corrosion standards, only the American standard was used to determine the depth of corrosion (corrosion geometry) for further testing and analysis. However,

considering the corrosion faults, these simplified shapes effectively demonstrate the capability of the proposed identification method in classification of corrosion at the underside of the rail. In the laboratory stage, these simple shapes are easier to be made than the realistic faults in the world. Additionally, the use of the American corrosion standard for depth determination is justified because maintenance standards across countries do not significantly differ. Hence, the American standard can be used as a reasonable reference for designing the fault size.

2) In terms of the experiments, several limitations, including rail length, noise, and testing conditions, need to be mentioned. To fit the requirements of simulation model, the rail length in the laboratory experiments was shorter than that in the real world (2 meters). But according to comparing the variance of the FRF from different length of the rail, the 2 meters rail is shown to be sufficient for FRF testing, and thus suitable for use in the laboratory experiments. Regarding noise in the experiments, only environmental noise from the laboratory experiment was considered. The other noise potentially generated in field tests is not considered. The choice to consider only environmental noise from laboratory experiments aligns with the conditions imposed by the simulation model, ensuring compatibility. Furthermore, the tested rail was placed on the floor, which differs from ballasted tracks with concrete bases in the real world. Again this is because the laboratory experiment is designed based on the simulation model. It requires more complexed computational capability when adding more components (concrete bases), and may bring sources of error into the calculation.

3) Regarding the simulation model, as mentioned in the limitation of experiment, more complex geometries, not only rail concrete bases, but also rail bolts, and longer rails, are not considered in this research. The reason is that it requires more complexed

computational capability when adding more components. And although the primary goal of the simulation model is to generate signals that closely resemble those challenging to obtain in field tests, it is more important to focus on achieving an accurate simulated model matched with laboratory experiments first. Therefore, a simpler model, with reduced potential for computational errors, was considered more appropriate for the initial validation stage. Subsequently, as Chapter 4 validated the capability of the simulation model, the research plans to progress to the next stage, which involves the incorporation of more components into the model. This expansion will allow for a more comprehensive assessment of practical conditions. Therefore, the research takes a stepwise approach, starting with a simplified model (without complexed components) to ensure accuracy in signal generation, and gradually enhancing complexity to align more closely with field conditions.

4) In terms of the proposed signal processing algorithm (VMD), the research was mainly on considering the variance of K-values during algorithm enhancement. And the result proves its effectiveness on processing the acoustic impact response signal. However, when initialising the VMD algorithm, other initial parameters are involved in triggering the VMD. Further investigation is needed to better understand and optimise these initial parameters for the effectiveness of the VMD. But the research of the proposed signal processing algorithm still contains their significance and values. The research successfully demonstrated the capacity of the algorithm to select the optimal K-values for enhancing the effectiveness of processing of acoustic signals. And the method of selecting the optimal K-values provides a valuable idea for further optimisation of the algorithm (for other initialising parameters) in the future.

5) In this thesis, the dataset was constructed based on acoustic signals simulated by the FEM model. However, the dataset size is limited. It includes only three simple corrosion labels to assess corrosion types and three different corrosion depths (0.25mm, 0.75mm, and 3mm) to evaluate corrosion levels. Consequently, the current dataset is suitable only for evaluating basic corrosion conditions on the underside of the rail. Classifying corrosion with complex geometries or depths not corresponding to the dataset's labels presents challenges for achieving effective classification. However, it is important to note that although the dataset is relatively small, it still allows for a meaningful analysis of simple corrosion scenarios. This small dataset has demonstrated the capability of serving the classification model to identify corrosions with different sizes and different levels. And the process of designing these simplified fault geometries provides valuable insights and experience that can serve for developing more complex geometries in the future.
6) In terms of the machine learning method, a designed 1-D CNNs model (integrated 1-D CNNs model) were employed for classification in this thesis. The limitation of the classification is that the comparison of 1-D CNNs model with other machine learning methods were not completed. But the reason that why1-D CNNs was selected for classifying in this thesis was addressed in the literature review (i.e., Considering the signals generated from the railway, 1-D CNN has been used for the classification of the structural fault on the railway system with the significant effectiveness). Additionally, the effectiveness of the designed 1-D CNNs model (integrated 1-D CNNs model) is compared with the conventional 1-D CNNs model, which is also proved the advantages of the integrated 1-D CNNs model.

7.4 Future work

The previous section described some of the limitations of this work. Suggestions for future consideration in these areas are presented here:

1) Regarding corrosion faults. It would be possible to create or simulate more realistic faults that would be more representative of real corrosion shapes, such as pitting corrosion, crevice corrosion, or galvanic corrosion. These would encompass different depths of corrosion for a more comprehensive analysis. The results obtained from analysing these faults would contribute to a larger dataset for classification, enabling a more thorough evaluation of corrosion health conditions on the railway.
2) Considering experiments. A more realistic testing platform could be established. This would involve considering various rail components, such as rail bolts, rail pads, and other structural elements. The data collected from this experimental platform would more closely resemble the signals that would be obtained during field test or deployment in the real railway.
3) Considering the simulation model. It is possible that this could be expanded to include various rail components that correspond to the components to be added to the new experimental platform, including rail bolts, rail pads, and other structural elements. By incorporating these details, the simulation could better replicate the sounds generated during train operations, leading to a more accurate representation of the railway environment. It would also be possible to enhance the simulations by adding more realistic noise, including mechanical noise and environmental noise (anthropogenic or biotic). Additionally, the simulation model could be extended beyond impact response signals generated by an impact hammer to simulate signals that occur when vehicles

pass through discontinuous sites on the railway, such as crossing noses or IRJs. This incorporation of real-world elements into the simulation might enable the automatic monitoring of the health condition of various elements of the railway infrastructure, potentially detecting not only corrosion but also other faults or issues, such as rail cracks, misalignments, or excessive vibration.

4) Considering the signal processing algorithm., In order to improve the signal processing techniques used for impact detection and analysis, further analysis could be conducted on the VMD algorithm. The objective would be to identify other parameters (beyond the K-value) that could be used to optimise the initialisation process of the VMD algorithm, leading to improved signal decomposition and extraction of impact-related information. Additionally, the improved VMD algorithm (IVMD-SVD) could potentially be applied to other types of acoustic signals.
5) When considering the machine learning method. The research presented has only evaluated the 1-D CNNs model for classification. In future work, other classification methods could be employed to assess their effectiveness and to compare them with the 1-D CNNs model. Additionally, the current study only classified corrosion levels based on a dataset of circular corrosion. In future work, in order to realise the classification of the corrosion in practical scenarios, it would be appropriate to modify the geometries of the corrosion in the simulation to fit the more practical faults found on the underside of the rail (more realistic corrosion, more corrosion levels). Specifically, additional corrosion levels should be considered for each corrosion type to ensure a comprehensive classification; and the classification model should be expanded to incorporate realistic corrosion types that correspond to the corrosion types designed for the new experimental platform. Subsequently, it would be necessary to train and validate the new simulated

dataset, which would include various more practical corrosion geometries and more corrosion levels. This dataset would be used to develop a model capable of identifying different corrosion types and levels in real-world settings. After testing this updated classification model with results obtained from field tests, it could be used to accurately identify the corrosion types and levels present in practical scenarios.

The future work listed above can be grouped into a number of themed activities or projects:

1) Advanced simulation of acoustic signals for realistic fault inspection (Future Work 1,2,3): Develop an advanced simulation model that focuses on collecting acoustic signals from more realistic faults in railway infrastructure. The simulation model should be designed for various practical fault types, including pitting corrosion, crevice corrosion, and galvanic corrosion, with different depths and shapes. It should also simulate signals from other structural issues like rail cracks, wear, plastic flow, etc. Creating such a highly detailed and realistic acoustic simulation model could provide a comprehensive dataset for fault inspection, enabling more accurate and efficient condition monitoring in the railway industry. This research would not only aim to solve the corrosion issues but also a broader range of potential fault sources, making it highly valuable for practical applications.
2) Advanced signal processing for comprehensive condition monitoring (Future Work 1,2,4): This would see the enhancement of signal processing techniques for comprehensive condition monitoring in railway infrastructure using enhanced signal processing algorithms that could detect not only corrosion but also other structural defects, such as rail cracks, wear, plastic flow, etc. It would investigate the use of advanced signal decomposition methods for improved fault inspection. The research

could also include a system based for data collection and analysis that integrates signals from various microphones and simulation models deployed in the railway network.

3) Machine learning for predictive maintenance and risk assessment (Future Work 4,5): This activity would develop a machine learning framework that would utilise data from both simulation models and railway networks for predictive maintenance and risk assessment. This research would employ advanced machine learning techniques to predict when and where maintenance is needed. Additionally, it could assess the overall risk to railway infrastructure components based on their corrosion levels and other structural factors. The research could be undertaken in collaboration with a railway operator to demonstrate the effectiveness of the advanced machine learning methods.

The above potential research projects could build up an advanced condition monitoring system for on-line monitoring of the condition at the underside of the rail. Such a system could support the rail industry in reducing maintenance costs, optimising normal railway operation, and enhancing passenger safety. By utilising environmental data sources, such as weather forecasts, historical rainfall records, and accident reports, areas with a heightened risk of corrosion within the railway network could be determined. The advanced condition monitoring system could then be deployed at these areas on the wayside of the track in proximity to wheel-rail impact sites (IRJs or crossing noses) that could generate appropriate impact response signals. The acoustic signals generated by wheel-rail impact from such impact sites would be collected and used for in the analysis undertaken by the advanced condition monitoring system. These results would then be compared with those produced by consideration of the simulated dataset to evaluate if any structural damages exist in these areas, and any to alert the maintenance centre appropriately. Using such a system, railway maintainers could realise on line monitoring to detect and quantify corrosion at the underside of the rail and hence avoid the risk of accidents due to the corrosion. Furthermore, such an approach could reduce the frequency of rail maintenance to enhance the effectiveness of railway operation, and hence bring significant profit for the company.

LIST OF REFERENCES

[1] Department for Transport, Concessionary travel statistics, England: Year ending March 2022, 2023. Available: https://www.gov.uk/government/statistics/concessionary-travel-statistics-year-ending-march-2022/concessionary-travel-statistics-england-year-ending-march-2022

[2] International Energy Agency, The Future of Rail, 2019.

[3] Abtus Limited. Available: http://www.abtus.com/cn/contact/

[4] S. L. Grassie, "Rolling contact fatigue on the British railway system: Treatment," Wear, vol. 258, no. 7-8, pp. 1310-1318, 2005.

[5] A. Kapoor, F. Franklin, S. Wong, and M. Ishida, "Surface roughness and plastic flow in rail wheel contact," Wear, vol. 253, no. 1-2, pp. 257-264, 2002.

[6] G. V. Duncheva and J. T. Maximov, "A new approach to enhancement of fatigue life of rail-end-bolt holes," Engineering Failure Analysis, vol. 29, pp. 167-179, 2013.

[7] W. Xu et al., "Corrosion of rail tracks and their protection," Corrosion Reviews, vol. 39, no. 1, pp. 1-13, 2021.

[8] F. C. R. Hernandez, K. Koch, and G. P. Barrera, Rail Base Corrosion Detection and Prevention (Contractor's Final Report for TCRP Project D-7, Task 14). Washington, DC: The National Academies Press, 2007.

[9] Britishsteel. (2023). Why Corrosion Protect Rail? Available: https://britishsteel.co.uk/what-we-do/rail/zinoco/the-need-for-corrosion-protection/

[10] D. Cannon, K.-O. Edel, S. Grassie, and K. Sawley, "Rail defects: An overview," Fatigue & Fracture of Engineering Materials & Structures, vol. 26, no. 10, pp. 865-886, 2003.

[11] R. J. N. Clark, "Rail flaw detection: Overview and needs for future developments," NDT & E International, vol. 37, no. 2, pp. 111-118, 2004.

[12] H. Zhang et al., "MRSDI-CNN: Multi-model rail surface defect inspection system based on convolutional neural networks," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11162-11177, 2021.

[13] W. Gong, M. F. Akbar, G. N. Jawad, M. F. P. Mohamed, and M. N. A. Wahab, "Nondestructive testing technologies for rail inspection: A review," Coatings, vol. 12, no. 11, art. no. 1790, 2022.

[14] Y. Tan, "Sensing localised corrosion by means of electrochemical noise detection and analysis," Sensors and Actuators B: Chemical, vol. 139, no. 2, pp. 688-698, 2009.

[15] J. Yang, E. Stewart, and M. Entezami, "Decomposition methods for impact-based fault detection algorithms in railway inspection applications," IET Signal Processing, vol. 16, no. 8, pp. 935-944, 2022.

[16] Transportation Research Board, National Academies of Sciences, Engineering, and Medicine, Guidelines for Rail Base Inspection and Rail Condemnation Limits for Corrosion-Induced Material Loss. Washington, DC: The National Academies Press, 2009.

[17] L. Lesley, H. Al Nageim, and LR55, "Static and dynamic load testing of a new low height rail system," in Proceedings of the Institution of Civil Engineers-Transport, vol. 117, no. 1, pp. 2-5, 1996.

[18] G. Tian, B. Gao, Y. Gao, P. Wang, H. Wang, and Y. Shi, "Review of railway rail defect non-destructive testing and monitoring," Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument, vol. 37, no. 8, pp. 1763-1780, 2016.

[19] X.-S. Jin, J. L. Guo, X.-B. Xiao, and Z.-F. Wen, "An investigation into effect of train curving on wear and contact stresses of wheel and rail," Tribology International, vol. 42, no. 3, pp. 475-490, 2008.

[20] W. Wang, J. Guo, Q. Liu, and M. Zhu, "Effect of laser quenching on wear and damage of heavy-haul wheel/rail materials," Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, vol. 228, no. 1, pp. 114-122, 2014.

[21] E. E. Magel, Rolling Contact Fatigue: A Comprehensive Review. U.S. Department of Transportation Federal Railroad Administration, 2011.

[22] S. B. Palmer, S. Dixon, R. S. Edwards, and X. Jian, "Transverse and longitudinal crack detection in the head of rail tracks using Rayleigh wave-like wideband guided ultrasonic waves," in Nondestructive Evaluation and Health Monitoring of Aerospace Materials, Composites, and Civil Infrastructure IV, 2005, vol. 5767, pp. 70-80.

[23] F. Sadeghi, B. Jalalahmadi, T. S. Slack, N. Raje, and N. K. Arakere, "A review of rolling contact fatigue," Journal of Tribology, vol. 131, no. 4, art. no. 041403, 2009.

[24] L. Smith, High performance rail for the long haul, 2018. Available: https://britishsteel.co.uk/what-weve-done/rail/heavy-haul-high-performance-rail/

[25] F. R. Hernández, G. Plascencia, and K. Koch, "Rail base corrosion problem for North American transit systems," Engineering Failure Analysis, vol. 16, no. 1, pp. 281-294, 2009.

[26] B. Panda, R. Balasubramaniam, and G. Dwivedi, "On the corrosion behaviour of novel high carbon rail steels in simulated cyclic wet–dry salt fog conditions," Corrosion Science, vol. 50, no. 6, pp. 1684-1692, 2008.

[27] Y. Sun, H. Li, C. Wang, and G. Shao, "Plastic strain accumulation behaviour of AH32 steel in a cyclic stress-corrosion environment," Journal of Constructional Steel Research, vol. 145, pp. 1-9, 2018.

[28] OKOndt GROUP™, Application of the EDDY CURRENT FLAW DETECTOR ETS2-77 for detection of surface defects rolling contact fatigue type in rails. Available: https://www.okondt.com/applications/rails-and-welds-testing/eddy-current-testing-of-rails

[29] L. Trykoz, S. Kamchatnaya, D. Borodin, A. Atynian, and R. Tkachenko, "Protection of railway infrastructure objects against electrical corrosion," Anti-Corrosion Methods and Materials, vol. 68, no. 5, pp. 380-384, 2021.

[30] Ministry of Defence, MOD UK Railways Permanent Way Design and Maintenance Standards, Issue 5, 2017. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974993/MOD_UK_Rly_PW_DM_-_issue_5_final_v3.pdf

[31] Gantrail. Available: https://www.gantrail.com/resources%20%20/

[32] M. Evans, A. Lucas, and I. Ingram "The inspection of level crossing rails using guided waves," Construction and Building Materials, vol. 179, pp. 614-618, 2018.

[33] Y. Chang, N. Li, J. Zhao, Y. Wang, and Z. Yang, "A novel ultrasonic guided wave-based method for railway contact wire defect detection," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-9, 2022.

[34] M. Mićić, L. Brajović, L. Lazarević, and Z. Popović, "Inspection of RCF rail defects–Review of NDT methods," Mechanical Systems and Signal Processing, vol. 182, art. no. 109568, 2023.

[35] Y. Jiang et al., "Non-contact ultrasonic detection of rail surface defects in different depths," in 2018 IEEE Far East NDT New Technology & Application Forum (FENDT), 2018, pp. 46-49.

[36] I. Ghafoor, W. T. Peter, N. Munir, and A. J. C. Trappey, "Non-contact detection of railhead defects and their classification by using convolutional neural network," Optik, vol. 253, art. no. 168607, 2022.

[37] J. Wu et al., "Ultrasonic guided wave-based switch rail monitoring using independent component analysis," Measurement Science and Technology, vol. 29, no. 11, art. no. 115102, 2018.

[38] S. Mariani et al., "Noncontact ultrasonic guided wave inspection of rails," Structural Health Monitoring, vol. 12, no. 5-6, pp. 539-548, 2013.

[39] E. O. Hall, "Variation of hardness of metals with grain size," Nature, vol. 173, pp. 948-949, 1954.

[40] S. Li, G. Wu, and H. Wu, "Study on acoustic emission characteristics of a semirigid base of dense skeleton type during complete uniaxial compression tests," Advances in Materials Science and Engineering, vol. 2016, art. no 9238698, 2016.

[41] V. Murav'ev and A. T. Baiteryakov, "Effect of traffic volume on acoustic structural noises in rails," Russian Journal of Nondestructive Testing, vol. 52, pp. 657-663, 2016.

[42] X. Zhang, T. Sun, Y. Wang, K. Wang, and Y. Shen, "A parameter optimized variational mode decomposition method for rail crack detection based on acoustic emission technique," Nondestructive Testing and Evaluation, vol. 36, no. 4, pp. 411-439, 2021.

[43] S. Liu, Q. Wang, and Y. Luo, "A review of applications of visual inspection technology based on image processing in the railway industry," Transportation Safety and Environment, vol. 1, no. 3, pp. 185-204, 2019.

[44] E. Deutschl, C. Gasser, A. Niel, and J. Werschonig, "Defect detection on rail surfaces by a vision based system," in IEEE Intelligent Vehicles Symposium, 2004, 2004, pp. 507-511.

[45] C. Mandriota, E. Stella, M. Nitti, N. Ancona, and A. Distante, "Rail corrugation detection by Gabor filtering," in Proceedings of the 2001 International Conference on Image Processing (Cat. No. 01CH37205), 2001, vol. 2, pp. 626-628.

[46] C. Mandriota, M. Nitti, N. Ancona, E. Stella, and A. Distante, "Filter-based feature selection for rail defect detection," Machine Vision and Applications, vol. 15, pp. 179-185, 2004.

[47] J. Ye, E. Stewart, and C. Roberts, "Use of a 3D model to improve the performance of laser-based railway track inspection," Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 233, no. 3, pp. 337-355, 2019.

[48] Z. Liu, W. Li, F. Xue, J. Xiafang, B. Bu, and Z. Yi, "Electromagnetic tomography rail defect inspection," IEEE Transactions on Magnetics, vol. 51, no. 10, pp. 1-7, 2015.

[49] J. Rajamäki, M. Vippola, A. Nurmikolu, and T. Viitala, "Limitations of eddy current inspection in railway rail evaluation," Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 232, no. 1, pp. 121-129, 2018.

[50] T. Shirotori, "A method for wayside bogie monitoring based on internal yard running vibration," Proceedings of the Institution of Mechanical Engineers Part F: Journal of Rail and Rapid Transit, vol. 235, no. 4, pp. 440-449, Apr 2021.

[51] B. Zhang, A. C. C. Tan, and J. H. Lin, "Gearbox fault diagnosis of high-speed railway train," Engineering Failure Analysis, vol. 66, pp. 407-420, Aug 2016.

[52] C. Yi, Y. Q. Li, X. M. Huo, and K. L. Tsui, "A promising new tool for fault diagnosis of railway wheelset bearings: SSO-based kurtogram," ISA Transactions, vol. 128, pp. 498-512, Sep 2022.

[53] M. Son and J. Seong, "Estimation of nondestructive compressive strength of concrete using impact force response signal," Journal of Testing and Evaluation, vol. 51, no. 5, doi: 10.1520/JTE20220376.

[54] W. X. Wang, J. Y. Li, W. G. Fan, and C. Y. Zhao, "Belt grinding mechanism-based method for roughness profile prediction of the rail surface," Journal of the Brazilian Society of Mechanical Sciences and Engineering, vol. 44, no. 3, art. no. 84, Mar 2022.

[55] B. Hou, Z. Shao, and Q. Zeng, "Correlation between rail vibration and sound radiation using a hammer impact," in IOP Conference Series: Materials Science and Engineering, 2020, vol. 841, no. 1, art. no. 012013.

[56] M. Oregui, Z. Li, and R. Dollevoet, "Identification of characteristic frequencies of damaged railway tracks using field hammer test measurements," Mechanical Systems and Signal Processing, vol. 54-55, pp. 224-242, 2015.

[57] D. P. Jena, S. N. Panigrahi, and R. Kumar, "Gear fault identification and localization using analytic wavelet transform of vibration signal," Measurement, vol. 46, no. 3, pp. 1115-1124, Apr 2013.

[58] A. G. Espinosa, J. A. Rosero, J. Cusido, L. Romeral, and J. A. Ortega, "Fault detection by means of Hilbert-Huang transform of the stator current in a PMSM with demagnetization," IEEE Transactions on Energy Conversion, vol. 25, no. 2, pp. 312-318, Jun 2010.

[59] R. Q. Yan and R. X. Gao, "Hilbert-Huang transform-based vibration signal analysis for machine health monitoring," IEEE Transactions on Instrumentation and Measurement, vol. 55, no. 6, pp. 2320-2329, Dec 2006.

[60] Y. Fujii, "Impact response measurement of an accelerometer," Mechanical Systems and Signal Processing, vol. 21, no. 5, pp. 2072-2079, Jul 2007.

[61] M. Oregui, M. Molodova, A. Nunez, R. Dollevoet, and Z. Li, "Experimental investigation into the condition of insulated rail joints by impact excitation," Experimental Mechanics, vol. 55, no. 9, pp. 1597-1612, Nov 2015.

[62] S. J. C. S. i. N. T. Kaewunruen and Evaluation, "Monitoring structural deterioration of railway turnout systems via dynamic wheel/rail interaction," vol. 1, pp. 19-24, 2014.

[63] S. Kaewunruen and A. Remennikov, "Field trials for dynamic characteristics of railway track and its components using impact excitation technique," NDT & E International, vol. 40, no. 7, pp. 510-519, 2007.

[64] S. Kaewunruen and A. Remennikov, "Application of vibration measurements and finite element model updating for structural health monitoring of ballasted railtrack sleepers with voids and pockets," in Mechanical Vibrations: Measurement, Effects and Control, R. C. Sapri, Ed. Nova Science Publishers, Inc., 2009.

[65] X. Han, M. Sarwar Islam, G. Newaz, L. Favro, and R. Thomas, "Finite element modeling of the heating of cracks during sonic infrared imaging," Journal of Applied Physics, vol. 99, no. 7, art. no. 074905, 2006.

[66] B. Raj, "Reliable solutions to engineering problems in testing through acoustic signal analysis," in Non-Destructive Testing '92, 1992, pp. 1-8.

[67] D. Zhang and D. Yu, "Multi-fault diagnosis of gearbox based on resonance-based signal sparse decomposition and comb filter," Measurement, vol. 103, pp. 361-369, 2017.

[68] R. Shafique et al., "A novel approach to railway track faults detection using acoustic analysis," Sensors, vol. 21, no. 18, art. no. 6221, 2021.

[69] W. Yao et al., "Rail defect inspection based on coupled structural–acoustic analysis," International Journal of Modern Physics B, vol. 33, no. 01n03, art. no. 1940031, 2019.

[70] F. P. García Márquez, P. J. Bernalte Sánchez, and I. Segovia Ramírez, "Acoustic inspection system with unmanned aerial vehicles for wind turbines structure health monitoring," Structural Health Monitoring, vol. 21, no. 2, pp. 485-500, 2022.

[71] A. Falamarzi, S. Moridpour, and M. Nazem, "A review on existing sensors and devices for inspecting railway infrastructure," Jurnal Kejuruteraan, vol. 31, no. 1, pp. 1-10, 2019.

[72] T. Nowakowski, F. Tomaszewski, P. Komorski, and G. Szymański, "Tram gearbox condition monitoring method based on trackside acoustic measurement," Measurement, vol. 207, art. no. 112358, 2023.

[73] D. Zhang, M. Xie, J. Yang, and T. Wen, "Multi-sensor graph transfer network for health assessment of high-speed rail suspension systems," IEEE Transactions on Intelligent Transportation Systems, 2023, doi: 10.1109/TITS.2023.3265401.

[74] R. Gunarathna and V. Sivahar, "Challenges in monitoring metallic pipeline corrosion using ultrasonic waves—A review article," Engineer: Journal of the Institution of Engineers, Sri Lanka, vol. 54, no. 1, pp. 67-75, 2021.

[75] A. Aljarah, N. Vahdati, and H. Butt, "Magnetic internal corrosion detection sensor for exposed oil storage tanks," Sensors, vol. 21, no. 7, art. no. 2457, 2021.

[76] L. Li, M. Chakik, and R. Prakash, "A review of corrosion in aircraft structures and graphene-based sensors for advanced corrosion monitoring," Sensors, vol. 21, no. 9, art. no. 2908, 2021.

[77] M. N. Akhtar, Guidelines for Rail Base Inspection and Rail Condemnation Limits for Corrosion-Induced Material Loss (Contractor's Final Report for TCRP Project D-7/Task 17). Washington, DC: The National Academies Press, 2009.

[78] J. Ye, E. Stewart, D. Zhang, Q. Chen, and C. Roberts, "Method for automatic railway track surface defect classification and evaluation using a laser-based 3D model," IET Image Processing, vol. 14, no. 12, pp. 2701-2710, 2020.

[79] Y. Zhan, X. Dai, E. Yang, and K. Wang, "Convolutional neural network for detecting railway fastener defects using a developed 3D laser system," International Journal of Rail Transportation, vol. 9, no. 5, pp. 424-444, 2021.

[80] Z. Zhang, M. Liang, and Z. Liu, "A novel decomposition model for visual rail surface inspection," Electronics, vol. 10, no. 11, art. no. 1271, 2021.

[81] A. Sophian, G. Tian, and M. Fan, "Pulsed eddy current non-destructive testing and evaluation: A review," Chinese Journal of Mechanical Engineering, vol. 30, no. 3, pp. 500-514, 2017.

[82] F. Aymerich and W. J. Staszewski, "Impact damage detection in composite laminates using nonlinear acoustics," Composites Part A: Applied Science and Manufacturing, vol. 41, no. 9, pp. 1084-1092, 2010.

[83] J. Guarino, R. Hamilton, and W. Fischer, "Acoustic detection of bolt detorquing in structures," in Proceedings of Meetings on Acoustics 157ASA, 2009, vol. 6, no. 1, art. no. 065002.

[84] J. Yang, E. Stewart, J. Ye, M. Entezami, and C. Roberts, "An improved VMD method for use with acoustic impact response signals to detect corrosion at the underside of railway tracks," Applied Sciences, vol. 13, no. 2, art. no. 942, 2023.

[85] D. Thompson, Railway noise and vibration: mechanisms, modelling and means of control. Elsevier, 2008.

[86] W. H. Organization, Environmental noise guidelines for the European region. World Health Organization. Regional Office for Europe, 2018.

[87] S. Venkataraman, R. Rumpler, S. Leth, M. Toward, and T. J. S. o. t. T. E. Bustad, "Improving strategic noise mapping of railway noise in Europe: Refining CNOSSOS-EU calculations using TWINS," vol. 839, p. 156216, 2022.

[88] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," The International Journal of Advanced Manufacturing Technology, vol. 89, pp. 3487-3494, 2017.

[89] A. Benedetto, F. Tosti, L. B. Ciampoli, and F. D’Amico, "An overview of ground-penetrating radar signal processing techniques for road inspections," Signal Processing, vol. 132, pp. 201-209, 2017.

[90] N. K. Mutlib, M. N. Ismael, and S. Baharom, "Damage detection in CFST column by simulation of ultrasonic waves using STFT-based spectrogram and welch power spectral density estimate," Structural Durability & Health Monitoring, vol. 15, no. 3, p. 227-246, 2021.

[91] A. J. D. Krupa, S. Dhanalakshmi, and R. Kumar, "Joint time-frequency analysis and non-linear estimation for fetal ECG extraction," Biomedical Signal Processing and Control, vol. 75, art. no. 103569, 2022.

[92] D. Mustafa, Z. Yicheng, G. Minjie, H. Jonas, and F. Jürgen, "Motor current based misalignment diagnosis on linear axes with short-time Fourier transform (STFT)," Procedia CIRP, vol. 106, pp. 239-243, 2022.

[93] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proceedings: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, pp. 903-995, 1998.

[94] A.-X. Pan and Z.-G. Yang, "Cause analysis and countermeasure on premature failure of a driven gear for the high-speed train," Engineering Failure Analysis, vol. 139, art. no. 106487, 2022.

[95] R. Nishat Toma, C.-H. Kim, and J.-M. Kim, "Bearing fault classification using ensemble empirical mode decomposition and convolutional neural network," Electronics, vol. 10, no. 11, art. no. 1248, 2021.

[96] M. Sareen, P. Prakash, and S. Anand, "Wavelet decomposition and feature extraction from pulse signals of the radial artery," in 2008 International Conference on Advanced Computer Theory and Engineering, 2008, pp. 551-555.

[97] T. Chen et al., "Emotion recognition using empirical mode decomposition and approximation entropy," Computers & Electrical Engineering, vol. 72, pp. 383-392, 2018.

[98] G. Han, B. Lin, and Z. Xu, "Electrocardiogram signal denoising based on empirical mode decomposition technique: An overview," Journal of Instrumentation, vol. 12, no. 3, art. no. P03010, 2017.

[99] S. Mohanty, K. K. Gupta, and K. S. Raju, "Comparative study between VMD and EMD in bearing fault diagnosis," in 2014 9th International Conference on Industrial and Information Systems (ICIIS), 2014, pp. 1-6.

[100] J. Zheng, J. Cheng, and Y. Yang, "Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis," Mechanical Systems and Signal Processing, vol. 40, no. 1, pp. 136-153, 2013.

[101] Z. Zhang, M. Entezami, E. Stewart, and C. Roberts, "Enhanced fault diagnosis of roller bearing elements using a combination of empirical mode decomposition and minimum entropy deconvolution," Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 231, no. 4, pp. 655-671, 2017.

[102] Y. Liu, G. Yang, M. Li, and H. Yin, "Variational mode decomposition denoising combined the detrended fluctuation analysis," Signal Processing, vol. 125, pp. 349-364, 2016.

[103] W. Guo and P. W. Tse, "A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals," Journal of Sound and Vibration, vol. 332, no. 2, pp. 423-441, 2013.

[104] T. Wang, M. Zhang, Q. Yu, and H. G. Zhang, "Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal," Journal of Applied Geophysics, vol. 83, pp. 29-34, 2012.

[105] J. Ma, J. Wu, Y. Fan, and X. Wang, "The rolling bearing fault feature extraction based on the LMD and envelope demodulation," Mathematical Problems in Engineering, vol. 2015, art. no. 429185, 2015.

[106] H. Jiang, C. Li, and H. Li, "An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis," Mechanical Systems and Signal Processing, vol. 36, no. 2, pp. 225-239, 2013.

[107] H. Wang, J. Chen, and G. Dong, "Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform," Mechanical Systems and Signal Processing, vol. 48, no. 1-2, pp. 103-119, 2014.

[108] Y. Amirat, V. Choqueuse, and M. Benbouzid, "EEMD-based wind turbine bearing failure detection using the generator stator current homopolar component," Mechanical Systems and Signal Processing, vol. 41, no. 1-2, pp. 667-678, 2013.

[109] X. Sun, M. Guo, M. Ma, and R. W. Mankin, "Identification and classification of damaged corn kernels with impact acoustics multi-domain patterns," Computers and Electronics in Agriculture, vol. 150, pp. 152-161, 2018.

[110] L. Yongbo, S. Shubin, L. Zhiliang, and L. Xihui, "Review of local mean decomposition and its application in fault diagnosis of rotating machinery," Journal of Systems Engineering and Electronics, vol. 30, no. 4, pp. 799-814, 2019.

[111] Y. Xu, K. Zhang, C. Ma, S. Li, and H. Zhang, "Optimized LMD method and its applications in rolling bearing fault diagnosis," Measurement Science and Technology, vol. 30, no. 12, art. no. 125017, 2019.

[112] Z. Wang et al., "Application of an improved ensemble local mean decomposition method for gearbox composite fault diagnosis," Complexity, vol. 2019, art. no. 1564243, 2019.

[113] Z. Wang, J. Wang, Y. Kou, J. Zhang, S. Ning, and Z. Zhao, "Weak fault diagnosis of wind turbine gearboxes based on MED-LMD," Entropy, vol. 19, no. 6, art. no. 277, 2017.

[114] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531-544, 2013.

[115] R. Wang, L. Xu, and F. Liu, "Bearing fault diagnosis based on improved VMD and DCNN," Journal of Vibroengineering, vol. 22, no. 5, pp. 1055-1068, 2020.

[116] C. Yi, Y. Lv, and Z. Dang, "A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition," Shock and Vibration, vol. 2016, art. no. 9372691, 2016.

[117] X. An and L. Pan, "Bearing fault diagnosis of a wind turbine based on variational mode decomposition and permutation entropy," Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, vol. 231, no. 2, pp. 200-206, 2017.

[118] Y. Miao, M. Zhao, Y. Yi, and J. Lin, "Application of sparsity-oriented VMD for gearbox fault diagnosis based on built-in encoder information," ISA Transactions, vol. 99, pp. 496-504, 2020.

[119] Y. Shrivastava and B. Singh, "A comparative study of EMD and EEMD approaches for identifying chatter frequency in CNC turning," European Journal of Mechanics - A/Solids, vol. 73, pp. 381-393, 2019.

[120] D. Chen, X. Zhang, H. Zhao, and H. Ding, "Development of a novel online chatter monitoring system for flexible milling process," Mechanical Systems and Signal Processing, vol. 159, art. no. 107799, 2021.

[121] K. Thameur, T. Marc, and G. Raynald, "Comparison between the efficiency of LMD and EMD algorithms for early detection of gear defects," Mechanics & Industry, vol. 14, no. 2, pp. 121-127, 2013.

[122] A. N. Wardana, "A comparative study of EMD, EWT and VMD for detecting the oscillation in control loop," in 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), 2016, pp. 58-63.

[123] S. Hadiyoso, E. Dewi, and I. Wijayanto, "Comparison of EMD, VMD and EEMD methods in respiration wave extraction based on PPG waves," in Journal of Physics: Conference Series, 2020, vol. 1577, no. 1, art. no. 012040.

[124] B. K. Pradhan, M. Jarzębski, A. Gramza-Michałowska, and K. J. N. Pal, "Automated detection of caffeinated coffee-induced short-term effects on ECG signals using EMD, DWT, and WPD," Nutrients, vol. 14, no. 4, art. no. 885, 2022.

[125] M. Zhang and G. Wei, "An integrated EMD adaptive threshold denoising method for reduction of noise in ECG," PLoS One, vol. 15, no. 7, art. no. e0235330, 2020.

[126] P. Singh and G. Pradhan, "Variational mode decomposition based ECG denoising using non-local means and wavelet domain filtering," Australasian Physical & Engineering Sciences in Medicine, vol. 41, pp. 891-904, 2018.

[127] G. Piao, J. Li, L. Udpa, J. Qian, and Y. Deng, "Finite-element study of motion-induced eddy current array method for high-speed rail defects detection," IEEE Transactions on Magnetics, vol. 57, no. 12, art. no. 6201010, 2021.

[128] T. Islam, M. Zaman, M. Hoque, and F. Rashid, "A numerical comparative study on heat generation pattern for different materials based on eigenfrequencies," in AIP Conference Proceedings, 2021, vol. 2324, no. 1, art. no. 030023.

[129] Z. Yang, P. Zhang, and L. Wang, "Wheel-rail impact at an insulated rail joint in an embedded rail system," Engineering Structures, vol. 246, art. no. 113026, 2021.

[130] C. Shen, Z. Li, and R. Dollovoet, "A novel method for railway crossing monitoring based on ambient vibration caused by train-track interaction," in Advances in Dynamics of Vehicles on Roads and Tracks: Proceedings of the 26th Symposium of the International Association of Vehicle System Dynamics, IAVSD 2019, August 12-16, 2019, Gothenburg, Sweden, 2020, pp. 133-141.

[131] I. O. Popp, "Analysis of the eigenfrequencies of the basalt machine parts by the FEM," in MATEC Web of Conferences, 2021, vol. 343, art. no. 08002.

[132] Z. He, G. Li, Z. Zhong, A. Cheng, G. Zhang, and E. Li, "An improved modal analysis for three-dimensional problems using face-based smoothed finite element method," Acta Mechanica Solida Sinica, vol. 26, no. 2, pp. 140-150, 2013.

[133] F. Immovilli, C. Bianchini, M. Cocconcelli, A. Bellini, and R. Rubini, "Bearing fault model for induction motor with externally induced vibration," IEEE Transactions on Industrial Electronics, vol. 60, no. 8, pp. 3408-3418, 2012.

[134] Q. Li, D.-P. Li, M. Hu, and X.-H. Xiao, "Optimization design of spiral bevel gear milling machine based on test FEM simulation and sensitivity analysis," in 2010 IEEE 17th International Conference on Industrial Engineering and Engineering Management, 2010, pp. 444-448.

[135] Y. B. Huang, Y. Ma, Q. C. Sun, W. Sun, W. B. Yue, and X. Sun, "Dynamic modeling and optimal design of a horizontal machining center based on experimental modal analysis," Applied Mechanics and Materials, vol. 868, pp. 271-276, 2017.

[136] M. Pletz, W. Daves, and H. Ossberger, "A wheel set/crossing model regarding impact, sliding and deformation—Explicit finite element approach," Wear, vol. 294, pp. 446-456, 2012.

[137] M. Pletz, W. Daves, W. Yao, and H. Ossberger, "Rolling contact fatigue of three crossing nose materials—Multiscale FE approach," Wear, vol. 314, no. 1-2, pp. 69-77, 2014.

[138] G. Schupp, C. Weidemann, and L. Mauer, "Modelling the contact between wheel and rail within multibody system simulation," Vehicle System Dynamics, vol. 41, no. 5, pp. 349-364, 2004.

[139] F. S. Iglesias and A. F. López, "Rayleigh damping parameters estimation using hammer impact tests," Mechanical Systems and Signal Processing, vol. 135, art. no. 106391, 2020.

[140] T. Shuku and K. Ishihara, "The analysis of the acoustic field in irregularly shaped rooms by the finite element method," Journal of Sound and Vibration, vol. 29, no. 1, pp. 67-76, 1973.

[141] Y. Kagawa, T. Yamabuchi, and A. Mori, "Finite element simulation of an axisymmetric acoustic transmission system with a sound absorbing wall," Journal of Sound and Vibration, vol. 53, no. 3, pp. 357-374, 1977.

[142] Y. Kagawa, T. Yamabuchi, K. Sugihara, and T. Shindou, "A finite element approach to a coupled structural-acoustic radiation system with application to loudspeaker characteristic calculation," Journal of Sound and Vibration, vol. 69, no. 2, pp. 229-243, 1980.

[143] S. Kopuz and N. Lalor, "Analysis of interior acoustic fields using the finite element method and the boundary element method," Applied Acoustics, vol. 45, no. 3, pp. 193-210, 1995.

[144] M. G. Sause and S. Richler, "Finite element modelling of cracks as acoustic emission sources," Journal of Nondestructive Evaluation, vol. 34, art. no. 4, 2015.

[145] A.-M. Zelenyak, M. A. Hamstad, and M. G. Sause, "Modeling of acoustic emission signal propagation in waveguides," Sensors, vol. 15, no. 5, pp. 11805-11822, 2015.

[146] L. L. Vergeynst, M. G. Sause, M. A. Hamstad, and K. Steppe, "Deciphering acoustic emission signals in drought stressed branches: The missing link between source and sensor," Frontiers in Plant Science, vol. 6, art. no. 494, 2015.

[147] F. Wang, S. C. M. Ho, and G. Song, "Modeling and analysis of an impact-acoustic method for bolt looseness identification," Mechanical Systems and Signal Processing, vol. 133, art. no. 106249, 2019.

[148] M. Mohammed, M. B. Khan, and E. B. M. Bashier, Machine Learning: Algorithms and Applications, CRC Press, 2016.

[149] K. Gulati, S. S. Kumar, R. S. K. Boddu, K. Sarvakar, D. K. Sharma, and M. Nomani, "Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic," Materials Today: Proceedings, vol. 51, pp. 38-41, 2022.

[150] S. A. Rajakarunakaran et al., "Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques," Advances in Engineering Software, vol. 173, art. no. 103267, 2022.

[151] P. William, A. Badholia, V. Verma, A. Sharma, and A. Verma, "Analysis of data aggregation and clustering protocol in wireless sensor networks using machine learning," in Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2021, Springer, 2022, pp. 925-939.

[152] M. Rella Riccardi, F. Mauriello, A. Scarano, and A. Montella, "Analysis of contributory factors of fatal pedestrian crashes by mixed logit model and association rules," International Journal of Injury Control and Safety Promotion, vol. 30, no. 2, pp. 195-209, 2022.

[153] A. Mekrache, A. Bradai, E. Moulay, and S. Dawaliby, "Deep reinforcement learning techniques for vehicular networks: Recent advances and future trends towards 6G," Vehicular Communications, vol. 33, art. no. 100398, 2022.

[154] M. Nasir, A. R. Javed, M. A. Tariq, M. Asim, and T. Baker, "Feature engineering and deep learning-based intrusion detection framework for securing edge IoT," The Journal of Supercomputing, vol. 78, pp. 8852–8866, 2022.

[155] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "IntruDTree: A machine learning based cyber security intrusion detection model," Symmetry, vol. 12, no. 5, art. no. 754, 2020.

[156] I. H. Sarker, Y. B. Abushark, and A. I. Khan, "Contextpca: Predicting context-aware smartphone apps usage based on machine learning techniques," Symmetry, vol. 12, no. 4, art. no. 499, 2020.

[157] W. I. D. Mining, "Data mining: Concepts and techniques," vol. 10, pp. 559-569, 2006.

[158] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," 2013. Available: arXiv:1302.4964 [cs.LG].

[159] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[160] S. L. Cessie and J. C. Van Houwelingen, "Ridge estimators in logistic regression," Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 41, no. 1, pp. 191-201, 1992.

[161] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN Computer Science, vol. 2, no. 3, art. no. 160, 2021.

[162] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, pp. 81-106, 1986.

[163] T. Elomaa, "In defense of C4. 5: Notes on learning one-level decision trees," in Machine Learning Proceedings 1994, 1994, pp. 62-69.

[164] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers-A survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 35, no. 4, pp. 476-487, 2005.

[165] W.-Y. Loh, "Classification and regression trees," WIREs Data Mining and Knowledge Discovery, vol. 1, no. 1, pp. 14-23, 2011.

[166] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, pp. 11-26, 2017.

[167] H. Kim, W.-K. Jung, Y.-C. Park, J.-W. Lee, and S.-H. Ahn, "Broken stitch detection method for sewing operation using CNN feature map and image-processing techniques," Expert Systems with Applications, vol. 188, art. no. 116014, 2022.

[168 Y. Li, "Research and application of deep learning in image recognition," in 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), 2022, pp. 994-999.

[169] Y. Ding et al., "Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification," Neurocomputing, vol. 501, pp. 246-257, 2022.

[170] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[171] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13, 2014, pp. 818-833.

[172] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013. Available: arXiv:1312.4400 [cs.NE].

[173] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904-1916, 2015.

[174] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1-9.

[175] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251-1258.

[176] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[177] N. Ganapathy, R. Swaminathan, and T. M. Deserno, "Deep learning on 1-D biosignals: A taxonomy-based survey," Yearbook of Medical Informatics, vol. 27, no. 1, pp. 98-109, 2018.

[178] P. P. San, S. H. Ling, and H. T. Nguyen, "Deep learning framework for detection of hypoglycemic episodes in children with type 1 diabetes," in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 3503-3506.

[179] B. Taji, A. D. Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 5, pp. 1124-1131, 2017.

[180] J. Yang, H. Duan, L. Li, E. Stewart, J. Huang, and R. J. A. S. Dixon, "1D CNN Based Detection and Localisation of Defective Droppers in Railway Catenary," vol. 13, no. 11, p. 6819, 2023.

[181] S. Teng, G. Chen, Z. Yan, L. Cheng, and D. J. S. H. M. Bassir, "Vibration-based structural damage detection using 1-D convolutional neural network and transfer learning," vol. 22, no. 4, pp. 2888-2909, 2023.

[182] C. Bapat and S. Sankar, "Single unit impact damper in free and forced vibration," Journal of Sound and Vibration, vol. 99, no. 1, pp. 85-94, 1985.

[183] A. J. Nimunkar and W. J. Tompkins, "R-peak detection and signal averaging for simulated stress ECG using EMD," in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 1261-1264.

[184] J. S. Smith, "The local mean decomposition and its application to EEG perception data," Journal of the Royal Society, Interface, vol. 2, no. 5, pp. 443-454, 2005.

[185] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout, "Steel defect classification with max-pooling convolutional neural networks," in The 2012 international joint conference on neural networks (IJCNN), 2012, pp. 1-6: IEEE.

[186] W. X. Cheng, P. N. Suganthan, and R. J. A. S. C. Katuwal, "Time series classification using diversified ensemble deep random vector functional link and resnet features," vol. 112, p. 107826, 2021.

[187] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, no. 2, pp. 281-305, 2012.

[188] Y. Song, Z. Wang, Z. Liu, and R. Wang, "A spatial coupling model to study dynamic performance of pantograph-catenary with vehicle-track excitation," Mechanical Systems and Signal Processing, vol. 151, art. no. 107336, 2021.

[189] G. Zhao, "Method for determining the rigidity of railway track," China Railway Science, vol. 26, pp. 1-6, 2005.

[190] R. Sampaio, N. Maia, and J. Silva, "Damage detection using the frequency-response-function curvature method," Journal of Sound and Vibration, vol. 226, no. 5, pp. 1029-1042, 1999.

[191] J. Krejčí and P. Beneš, "A comparison of two modal analysis methods: An impact hammer versus laser vibrometry," Acta Polytechnica, vol. 60, no. 5, pp. 420-427, 2020.

[192] V. Yantchev, P. Turner, and V. Plessky, "COMSOL modeling of SAW resonators," in 2016 IEEE International Ultrasonics Symposium (IUS), 2016, pp. 1-4.

[193] H. Benzeroual, A. Khamlichi, and A. Zakriti, "Inspection of transverse flaws for railways using phased array ultrasonic technique," International Review of Applied Sciences and Engineering, vol. 12, no. 2, pp. 119-126, 2021.

[194] D. J. Rixen and P. L. van der Valk, "An impulse based substructuring approach for impact analysis and load case simulations," Journal of Sound and Vibration, vol. 332, no. 26, pp. 7174-7190, 2013.

[195] K. L. Elmore and M. B. Richman, "Euclidean distance as a similarity metric for principal component analysis," Monthly Weather Review, vol. 129, no. 3, pp. 540-549, 2001.

[196] E. Kobayashi, T. Fushimi, K. Saito, and T. Ikeda, "Similarity search by generating pivots based on Manhattan distance," in PRICAI 2014: Trends in Artificial Intelligence: 13th Pacific Rim International Conference on Artificial Intelligence, Gold Coast, QLD, Australia, December 1-5, 2014, Proceedings 13, 2014, pp. 435-446.

[197] É. O. Rodrigues, "Combining Minkowski and Chebyshev: New distance proposal and survey of distance metrics using k-nearest neighbours classifier," Pattern Recognition Letters, vol. 110, pp. 66-71, 2018.

[198] R. Taheri, M. Ghahramani, R. Javidan, M. Shojafar, Z. Pooranian, and M. Conti, "Similarity-based Android malware detection using Hamming distance of static binary features," Future Generation Computer Systems, vol. 105, pp. 230-247, 2020.

[199] S. McLeod, "Correlation definitions, examples & interpretation," Simply Psychology, 2018.

[200] S. Wan and B. Peng, "The FERgram: A rolling bearing compound fault diagnosis based on maximal overlap discrete wavelet packet transform and fault energy ratio," Journal of Mechanical Science and Technology, vol. 33, pp. 157-172, 2019.

[201] T. Pattabiraman, K. Srinivasan, and K. Malarmohan, "Assessment of sideband energy ratio technique in detection of wind turbine gear defects," Case Studies in Mechanical Systems and Signal Processing, vol. 2, pp. 1-11, 2015.

[202] Y. Yu and C. Junsheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," Journal of Sound and Vibration, vol. 294, no. 1-2, pp. 269-277, 2006.

[203] Z. Zhang, H. Li, G. Meng, X. Tu, and C. Cheng, "Chatter detection in milling process based on the energy entropy of VMD and WPD," International Journal of Machine Tools and Manufacture, vol. 108, pp. 106-112, 2016.

[204] K. Yang, G. Wang, Y. Dong, Q. Zhang, and L. Sang, "Early chatter identification based on an optimized variational mode decomposition," Mechanical Systems and Signal Processing, vol. 115, pp. 238-254, 2019.

[205] H. Liu, D. Li, Y. Yuan, S. Zhang, H. Zhao, and W. Deng, "Fault diagnosis for a bearing rolling element using improved VMD and HT," Applied Sciences, vol. 9, no. 7, art. no. 1439, 2019.

[206] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," Journal of Sound and Vibration, vol. 435, pp. 36-55, 2018.

[207] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in A Practical Approach to Microarray Data Analysis, D. P. Berrar, W. Dubitzky, and M. Granzow, Eds. Norwell, MA: Kluwer, 2003, pp. 91-109.

[208] M. Zhao and X. Jia, "A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery," Mechanical Systems and Signal Processing, vol. 94, pp. 129-147, 2017.

[209] S. Wang, L. Wei, and L. Zeng, "Ultra-short-term photovoltaic power prediction based on VMD-LSTM-RVM model," in IOP Conference Series: Earth and Environmental Science, 2021, vol. 781, no. 4, art. no. 042020.

[210] X. Wu, Q. Zhang, and M. Tian, "Photovoltaic prediction method based on VMD-SE and optimized support vector machine," Power Science and Engineering, vol. 33, no. 9, p. 29-36, 2017.

[211] Y. Li, D. Huang, and Z. Qin, "A classification algorithm of fault modes-integrated LSSVM and PSO with parameters' optimization of VMD," Mathematical Problems in Engineering, vol. 2021, art. no. 6627367, 2021.

[212] Z. Q. Sun and H.J. Zhang, "Application of empirical mode decomposition based energy ratio to vortex flowmeter state diagnosis," Journal of Central South University of Technology, vol. 16, no. 1, pp. 154-159, 2009.

[213] X. Guan, C. Huang, G. Liu, X. Meng, and Q. Liu, "Mapping rice cropping systems in Vietnam using an NDVI-based time-series similarity measurement based on DTW distance," Remote Sensing, vol. 8, no. 1, art. no. 19, 2016.

[214] A. Makarova and V. Sulimova, "High-performance DTW-based signals comparison for the brain electroencephalograms analysis," in CEUR Workshop Proceedings, 2017, pp. 18-24.

[215] D. Berrar, "Cross-validation," in Encyclopedia of Bioinformatics and Computational Biology, Volume 1, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds. Elsevier, 2019, pp. 542-545.

[216] M. Stone, "Cross-validatory choice and assessment of statistical predictions," Journal of the Royal Statistical Society. Series B (Methodological), vol. 36, no. 2, pp. 111-133, 1974.

[217] S. R. Esmaeelzadeh, A. Adib, and S. Alahdin, "Long-term streamflow forecasts by Adaptive Neuro-Fuzzy Inference System using satellite images and K-fold cross-validation (Case study: Dez, Iran)," KSCE Journal of Civil Engineering, vol. 19, pp. 2298-2306, 2015.

[218] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249-256.

[219] A. Mishra and U. S. Reddy, "A novel approach for churn prediction using deep learning," in 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017, pp. 1-4.

[220] A.-M. Šimundić, "Measures of diagnostic accuracy: Basic definitions," EJIFCC, vol. 19, no. 4, pp. 203-211, 2009.

[221] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Semantic annotation and retrieval of music and sound effects," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 2, pp. 467-476, 2008.

[222] H. Huang, H. Xu, X. Wang, and W. Silamu, "Maximum F1-score discriminative training criterion for automatic mispronunciation detection," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 4, pp. 787-797, 2015.

[223] Y. Q. Chen, O. Fink, and G. Sansavini, "Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction," IEEE Transactions on Industrial Electronics, vol. 65, no. 1, pp. 561-569, 2017.

[224] C. Cheadle, M. P. Vawter, W. J. Freed, and K. G. Becker, "Analysis of microarray data using Z score transformation," Journal of Molecular Diagnostics, vol. 5, no. 2, pp. 73-81, 2003.