Intelligent real-time monitoring of critical rail infrastructure

By

Richard Culwick

A thesis submitted to

The University of Birmingham

for the degree of

Doctor of Philosophy

School of Metallurgy and Materials

College of Engineering and Physical Sciences

University of Birmingham

September 2022

Abstract

Rail networks are key national infrastructure assets, providing mass transport capability for both goods and passengers. The reliability of these networks is therefore critical to avoid widescale disruption. To ensure reliability, they must be effectively monitored and maintained. The monitoring of the structural health of the Rail network is challenging with rolling stock of different weights and speeds travelling on the same lines. Current structural health monitoring techniques are unable to effectively monitor in real time defects forming on the rail. Therefore, to ensure continued reliability of the rail network new monitoring methods must be found.

This work investigates the use and advancement of acoustic emission techniques in monitoring the real time structural health of critical rail infrastructure. This work will focus on the monitoring of R220 and R260 grade steels used for plain track and cast manganese steel used for rail crossings.

Acoustic emission is already used across a wide number of industries for the detection of crack growth. This work initially looks at the feasibility of a commercially avaible system, procured from Physical Acoustics for the monitoring of fatigue crack growth in rail steels. Three key acoustic parameters were focussed on, Energy, Duration and Counts. A good correlation was found between increasing crack growth and these parameters with increasing crack severity for both the R220 and R260 steels, and to a lesser extent with the cast manganese steel. This correlation shows the potential for applying these techniques to the monitoring of fatigue crack growth in the rail environment.

It is proposed that these commercial systems are limited in their accuracy and capacity for real-time monitoring as the acoustic data is packaged into hits removing much of the avaible data.

An alternative approach is therefore proposed using a customised acoustic emission monitoring system that captures and analysis the complete acoustic waveform.

The volume of data generated using the custom system necessitated the use of automated analysis techniques. Machine learning techniques were therefore developed in this work to analyse and classify the acoustic emission data generated during fatigue testing under laboratory conditions. Three signal processing techniques where tested; FFT, RMS and CWT with both shallow and deep neural networks developed for the FFT and RMS processing routes. High prediction accuracy was achieved using the custom system with the FFT shallow neural network achieving an accuracy of 87.8%.

Acknowledgments

I would like to thank the EPSRC for providing the financial support for this work.

I would also like to thank my Supervisor Dr Mayorkinos Papaelias for his guidance and support throughout the completion of this work. In addition, I am grateful to Dr Valter Jantara Junior and Dr Sanaz Roshanmanesh for their knowledge and guidance in support of this work as well as Manwika Kongpuang for her assistance in the completion of the laboratory fatigue tests.

I am grateful to Dr Patrick Vallely for his help in providing access to the rail network to allow field studies to take place alongside Network rail who also kindly proved the R220 and R260 grade steel samples used in this work.

Publications and Awards

- M.Papaelias, A. Amini, R. Culwick, J. Heesom, Z. Huang, V.L. Jantara Junior, S. Kaenwunruen, S. Kerkyras, M. Kongpuang, F.P. Garcia Marquez, S. She, A.Upton, P. Vallely "*Advanced remote condition monitoring of railway infrastructure and rolling stock".* In proceedings of the 1st International Conference on Welding & NDT, Athens, Eugenides Foundation, 22-23 October 2018.
- R. Culwick *"Remote condition monitoring of railway assets"* UKRRIN Annual conference, Birmingham, 21st November 2019, Poster.
- R. Culwick *"Acoustic emission monitoring of railway assets*" IOM3 Future Materials conference, Sheffield, 17th December 2019, Poster.
- R. Culwick *"Workstream 1: Fatigue assessment of steel bridges"* Assets4Rail International competition 2020.
    - Awarded first prize
- R. Culwick, B. Summers, M. Kongpuang, V.L. Jantara Junior, P. Vallely, M. Papaelias *"Advnaces in fatigue crack growth monitoring of R260 rail steel using acoustic emission monitoring"* Proceedings of the 10th International Conference on Structural Health Monitoring of Intelligent Infrastructure, SHMII10, Porto, Portugal, 30th June-2nd July 2021.
- R. Culwick *"Improving rail reliability using acoustic emission monitoring"* Royce summit 3rd March 2021, Online, Presentation and chair
- R. Culwick, V. Jantara Junior, M. Papaelias *"Real time monitoring of crack growth in rail using acoustic emission technique"* Condition Monitor, Issue 421, pp7-9 April 2022 Presented at the CM2021, the Seventeenth International Conference on Condition Monitoring and Asset management, 14-18th June 2021, Held Virtually.
- M. Kongpuang, R. Culwick, N. Cheputeh, A. Marsh, V.L. Jantara Junior, P. Vallely, S. Kaewunruen and M. Papaelias. *"Quantitative Analysis of the structural health of Railway Turnout using acoustic Emission technique"* Insight, Vol 64, No7 pp1-6 July 2022, DOI:10.1784/insi.2022.64.7.XXX Presented at the CM2021, the Seventeenth International Conference on Condition Monitoring and Asset management, 14-18th June 2021, Held Virtually

Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations and Symbols

φ^*	Conjugate of complex-valued function φ
µs	Microseconds
a	Scale parameter corresponding to the Pseudo-frequency
a_0	Initial crack length
AE	Acoustic emission
AI	Artificial intelligence
ASTM	American society of testing and materials
b	Translating parameter
bn	Billion
B_n	Frequency band
c	Acoustic Counts
CWT	Continuous wavelet transform
D	Acoustic Duration
DAQ	Digital acquisition
db	Decibels
DCPD	Direct Current Potential drop
E	Acoustic Energy
ERTMS	European Rail Traffic Management System
ESR	Emergency speed restriction
FEA	Finite element analysis
FFT	Fast Fourier transform
Fs	Sampling Frequency
GCC	Gauge Corner Cracking
HB	Brinell hardness
HCF	High cycle fatigue
HDD	Hit Driven Data
HPC	High performance computing
HRC	Rockwell hardness
HS2	High Speed 2
HV	Vickers hardness value
k	Thousand
kHZ	Kilohertz
km	Kilometre
kN	Kilo newtons
LCF	Low cycle fatigue
MFL	Magnetic flux leakage
MHz	Mega Hertz
mm	Millimetres
MnS	Manganese Sulphide
MPa	Mega pascals
MS	Million samples

n	Number of cycles
NDT	Non-destructive testing
Nm	Newton Meters
NMT	New measurement train
NN	Neural Network
PAC	Physical Acoustics – now Mistras
RCF	Rolling Contact Fatigue
ReLU	Rectified linear unit
RGB	Red Green Blue
RMS	Route Mean Square
RUL	Remaining useful life
SCG	Scaled conjugate gradient
SEM	Scanning electron microscope
SHM	Structural health monitoring
TDD	Time Driven Data
UK	United Kingdom
UT	Ultrasonic testing
V	Voltage
V_0	Initial Voltage
W	Width
WP	Wavelet Power
y	Half sensor distance
ΔK	Stress intensity factor

Chapter 1

Introduction

1.1 Challenges facing the rail industry

Rail networks provide an effective mass transport method for both goods and passengers with minimal environmental impact when compared with other transport modes (Asplund, 2016, Ekberg and Kabo, 2005, Hendy, 2015). In the year to March 2022 a total of 16.87bn tonne km of freight was moved a rise of 11.3% since 2021 and 1.8% from 2019 (Office of Rail, 2022). According to the Department for Transport Freight strategy, the transport of freight by rail emits 76% less CO_2 per tonne·km when compared with road transport. As recent events have shown, delays in supply chains can have far-reaching consequences across a large range of industries and everyday life activities. As a critical part of these supply chains and a key passenger transport mode, the reliability of the rail network must be optimized to ensure minimum disruption. This is becoming increasingly urgent as passengers and goods volumes continue to climb and are predicted to continue this growth until at least 2030 (H. Gunel et al., 2017, Yilmazer 2012, Vallely, 2015).

At just over two hundred years old, the UK rail network is one of the oldest networks in the world (Freeman and Shaw, 2000). It is also one of the busiest in Europe, operating at close to its maximum capacity (Hendy, 2015, Yilmazer, 2012). Ageing networks require maintenance and renewal, which is expensive costing Network Rail £4.645bn in the period 2019-20, or approximately 42% of its total budget for that year (NetworkRail, 2021). With the UK moving towards 24-hour rail network operations and with increasing axle loads, speeds and use of capacity, the already ageing network is under greater pressure with greater levels of wear and

reduced overnight downtime for routine maintenance. This with time will inevitably lead to greater failure rates, leading to increased costs to Network Rail in the form of fines and maintenance costs, economic impact due to delayed passengers and freight, and potentially could result in casualties and environmental damage. Any disruption due to failure in such a congested network will have knock on effects in the operations across the entire network, for example in 2015, 81,137 minutes in delays were caused on the London Northwestern south line due to the failure of rail crossings (Huang, 2017). Rail failure has also led to serious accidents such as the Hadfield rail accident in 2000 which led to the deaths of 4 people, the injury of over 70 others and an estimated economic cost of £1 billion due to the failure of rail infrastructure (Magel, 2011).

It is critical that solutions are found to increase capacity on the network whilst maintaining the reliability and safety of this transport mode. In reality two approaches must be pursued in parallel to meet the increasing rate in demand; firstly, the expansion of the physical infrastructure, for which work is already underway in the form of major infrastructure projects such as Crossrail and High Speed 2 (HS2). The second approach is to upgrade and enhance the existing infrastructure to maximise its capacity, for example with the introduction of the European Rail Traffic Management System (ERTMS) allowing more trains to run closer together and the upgrade and electrification of lines to increase speeds, for example the Great Western electrification programme.

Even with these two approaches, railway networks are still susceptible to failure of critical infrastructure. It is therefore vital that effective structural health monitoring, ideally in real time, is in place for evaluating key structural components. The monitoring of rail infrastructure has two key requirements; firstly, the accurate detection and location of faults before they become critical, and secondly to do so with minimal or no impact on the normal operation of said asset. This second requirement is becoming increasingly more important as rail networks

become more and more congested. Traditionally a period of downtime had been planned over night to allow access for monitoring and maintenance. This downtime is being steadily reduced in order to increase network capacity. With the time available for inspection and routine maintenance becoming gradually less and less, effective monitoring methods and maintenance schedules must be optimized to minimise the network access time required for inspection and maintenance activities.

1.2 Rail Defects

To monitor and effectively maintain the rail network the defects and failure mechanisms affecting the different structural components of the network must be known and understood. Rails experience high impact loads of up to $1200Nm^{-2}$ (Bonnett, 2005) and with increases in rolling stock axle loads this will lead to even greater wear rates of rails. The way in which these forces are applied to the rail depends on the section of the rail. In the case of straight track sections, the wheelsets have little to no lateral movement meaning that the force applied to the rail is primarily through rolling contact. Corners and crossings in the rail change the direction of the train leading to lateral forces being applied as well as impact forces particularly in the case of rail crossings in the area of the crossing nose.

Undetected cracks can lead to rail failure whilst wear can change the optimum railhead geometry. Whilst changes in geometry of the railhead may not necessarily lead to crack growth, they are not less serious since they can lead to derailment of rolling stock due to poor adhesion at the wheel-rail interface. They can also result in excessive loads being applied to both rails and wheels resulting in higher levels and rates of rolling contact fatigue damage (RCF).

Higher loads applied on the rail can result in accelerated damage evolution due to defects that may have already initiated or having been present due to manufacturing and installation processes. The problem of manufacturing defects used to be a more serious problem in old rail

steel grades due to lower manufacturing and control capabilities. However, improvements in the modern rail steelmaking coupled with stringent quality control, have almost entirely removed the occurrence of any manufacturing defects ensuring rails entering service are of much higher quality than in the past (Aglan & Fateh, 2007; Ph Papaelias et al., 2008). Nonetheless, modern rails entering service can be affected by a decarburised surface layer and the presence of MnS inclusions both of which will impact on the performance of the rail once in service (Fegredo et al., 1988; Garnham et al., 2010). This can be avoided through further improvements in the production of rail steel grades.

Following the improvements in manufacturing quality, many of the defects traditionally found in plain rails, such as hydrogen cracking, have become less common whilst the hardness of the steel has increased. However, this change in hardness has resulted in certain other types of defects becoming more commonplace, such as RCF cracking. RCF is now the primary failure mechanism detected on the rail network. The study of Perez-Unsueta and Beynon found that the introduction of harder rail steels is partially responsible for the predominance of RCF (Perez-Unzueta & Beynon, 1993). Harder rails were developed to reduce wear however this means that cracks forming at the surface of the rail are not removed due to natural wear and are therefore able to grow up to the point where they will lead to failure.

RCF is not associated with a single failure mechanism. It instead describes several different failure mechanisms, all associated with high cyclic loading of the surface of the rail and the wheel-rail interface leading to crack growth at the surface or below the surface of the rail (Grassie & Kalousek, 1997). Figure 1.1 shows examples of lipping, squats, gauge corner cracking (GCC) and shelling all of which are caused by RCF.

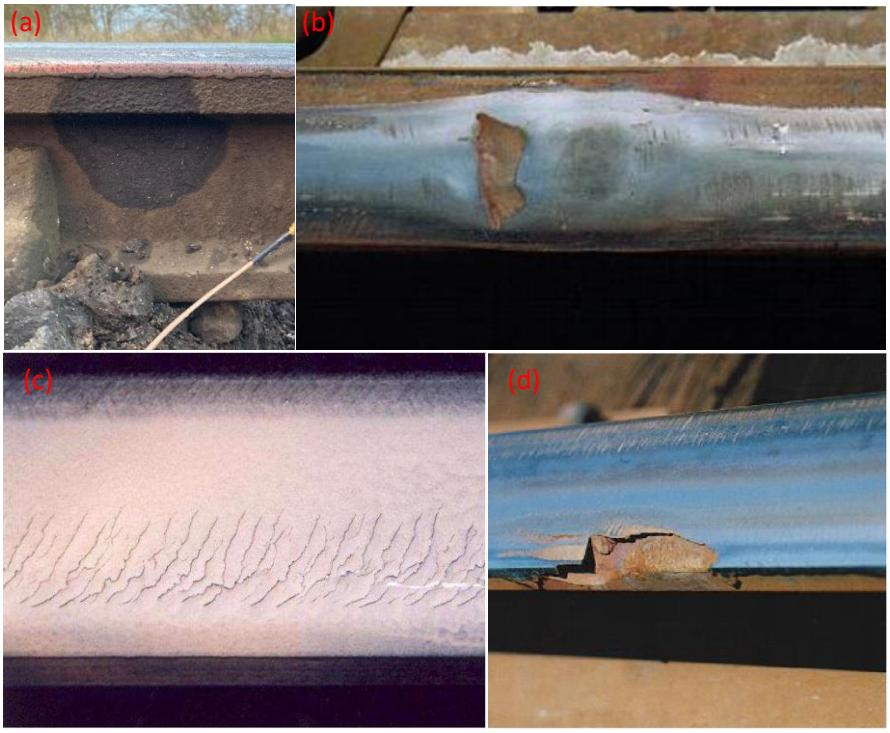


Figure 1.1 *Examples of defects formed due to RCF on plain track rail heads showing; (a) lipping (b) squats (c) GCC and (d) shelling* (Magel, 2011; NSWRailcop, 2019).

Whilst these defects are the visible damage to the surface of the rail, it is difficult to evaluate the severity of damage below the surface. This subsurface crack growth is particularly dangerous as it is challenging to detect using conventional structural health monitoring (SHM) techniques, which means cracks are potentially able to grow to failure with little or no indication on the surface of the rail. Figure 1.2 clearly illustrates this; a large subsurface crack has grown and yet there is only a very small surface breaking defect that can be observed.

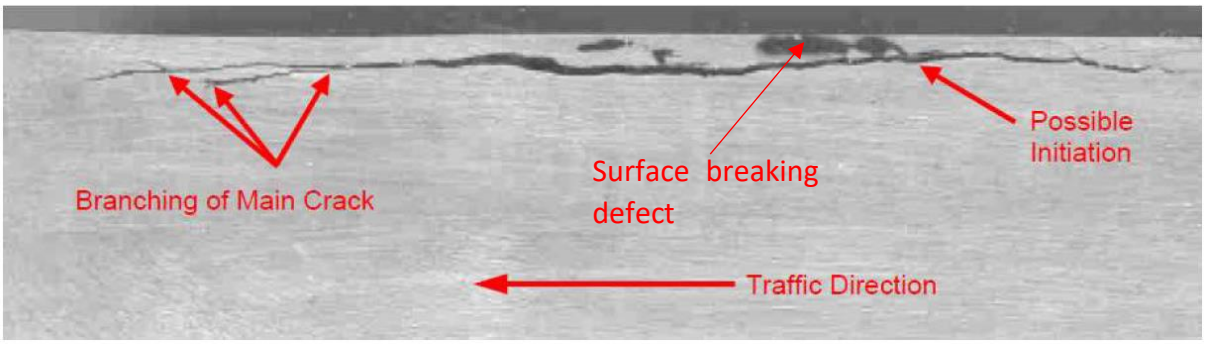


Figure 1.2 *Image of subsurface crack growth caused by RCF running parallel to the running surface of the rail head. The subsurface crack is significantly larger than the small surface breaking defect that would be observable on the rail head* (NSWRailcop, 2019).

In contrast to the plain rail steels cast manganese steel crossings are still highly susceptible to manufacturing defects which can eventually lead to structural failure due to accelerated damage initiation and propagation (Peters, 2005). These manufacturing defects are generally in the form of carbides and inclusions, such as MnS, acting as crack initiation sites below the rail surface (Alyaz, 2003; Smith & Mackay, 2003).

In its as cast state, cast manganese steel does not have the required hardness to withstand the high impact loads it undergoes. This problem can be rapidly overcome through the use of explosive hardening which rapidly work hardens the surface of the material. This reduces the wear damage to the head of these crossings. If explosive hardening is not applied lipping will occur until sufficient work hardening is achieved. This change in geometry requires subsequent corrective grinding. Even following explosive hardening, the high impact loads experienced by rail crossings lead to wear and changes in track geometry, an example of this can be seen in Figure 1.3.

Figure 1.3 *Image showing wear damage to a cast manganese crossing clearly showing a change in rail head geometry.*

As with plain track rail steels, sub-surface cracks are able to grow in the cast manganese crossings. These cracks initiate at subsurface carbides. Once initiated they run parallel to the running surface as shown in Figure 1.4. Due to the high impact loads experienced by these components undetected cracks are of particular concern as they are able to grow rapidly and may lead to sudden failure. Although partial failure of plain rails can be tolerated, breaks in crossings can lead to serious rail accidents as they are critical to guiding the rolling stock wheelsets to change direction and thus, can result in derailment.

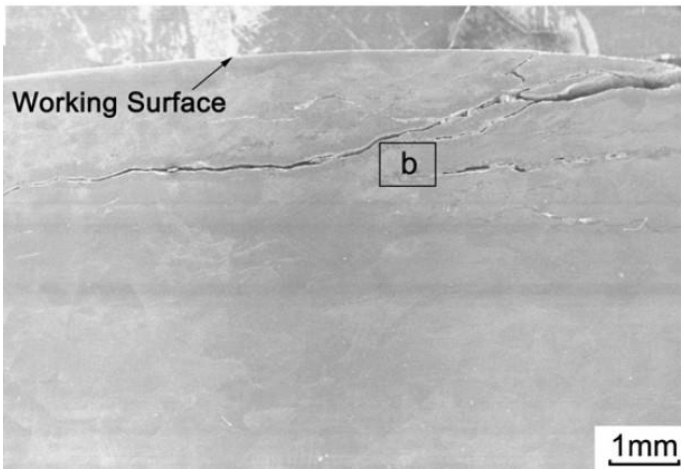


Figure 1.4 S*ubsurface crack (b) in a section of cast manages steel taken from the head of a rail crossing. A large crack has grown parallel to the running surface with only a small surface breaking section visible* (Lv et al., 2012).

The defects discussed herewith, give a short overview of the defects found in rails. It is important to understand these defects in order to develop effective monitoring and maintenance schedules ensuring a safe and reliable rail network. Defects can be broadly categorized in one of the following three categories (Ph Papaelias et al., 2008): 1. Manufacturing defects 2. Defects due to fatigue or corrosion of the rail 3. Improper usage. Many years of work by the steel manufacturing industry has led to a reduction in manufacturing defects entering service. Improper use and handling particularly during installation is arguably the hardest to monitor, however proper training should minimize the risk of this occurring. The focus for rail SHM is therefore on the defects caused by fatigue and corrosion in service.

1.3 Current structural health monitoring techniques

As with all critical infrastructure, rail infrastructure must be effectively monitored to maintain reliability and safety. There are many components encompassed within rail infrastructure each with different SHM techniques in use. In this section a brief overview of the current SHM techniques used on plain track and cast manganese crossings is given.

Traditional models of mechanical failure assume a relatively predictable loading cycle. In the case of rails and crossings this is not the case. Rails and crossings are stochastically loaded by a large variety of rolling stock operating in different lines. This leads to difficulties in predicting wear rates as traditional models are likely to overestimate the useful lifespan of the rails (Ding et al., 2016; Ringsberg & Bergkvist, 2003). This has meant that it is not uncommon for rails to fail before their predicted life span has been achieved (Zhu & Olofsson, 2014). That is not to say that models cannot be used to predict the current structural health and remaining useful life (RUL) of a component. Models must be developed that are adaptable with data of the real time structural health fed back into the model, updating it for example in the form of a digital twin. To achieve this effective SHM techniques must be developed that are able to accurately monitor infrastructure in real time.

Current track monitoring techniques can broadly be divided into three categories: 1. Track mounted 2. Rolling stock mounted 3. Mobile manual methods. As no single technique is yet able to effectively detect all rail defects and combination of techniques is deployed (Amini, 2016; Shi et al., 2017).

1.3.1 Track mounted SHM techniques

Track-mounted systems are widely used for the monitoring of critical rolling stock components, such as hotboxes for monitoring axle bearings. For this application wayside monitoring is ideal as a single hotbox is able to monitor the axle bearings of all the trains as they pass. Effective SHM, however becomes more challenging when rail track monitoring is required. As the track and the wayside monitoring units are fixed in place, each unit can cover a finite monitoring area. Therefore, a large number of units is required to cover the entire railway network.

The geometry of the rail and rolling stock wheel tread also presents a challenge for wayside monitoring. Any device mounted directly on to the rail or close to it must not interfere with any structure hanging from the rolling stock, this gives limits to the size and geometry of any attached device. Whilst plain track rail is relatively simple in geometry rail crossings are more complex with each one having a unique geometry. A further challenge to rail mounted systems is the type of rolling stock power units employed on the network. In some areas electrification is provided through rail. Therefore, any universal mounted system must be designed to operate on these powered rails without interference.

The impact of wheels on the rail, the deflection of rails under load, and the load applied to a rail can be directly monitored using contact strain gauges attached on the rail web. Rail mounted mechanical methods can also be used to detect rail wear. A simple pressure plate can be used to indicate a drop in wheel height relative to the rail foot indicating rail wear (Figure 1.5). This technique is however severely limited in the distance over which it can monitor, more advanced techniques using laser suggested by Barke and Chiu may have the capability to partially overcome this limitation (Barke & Chiu, 2005).

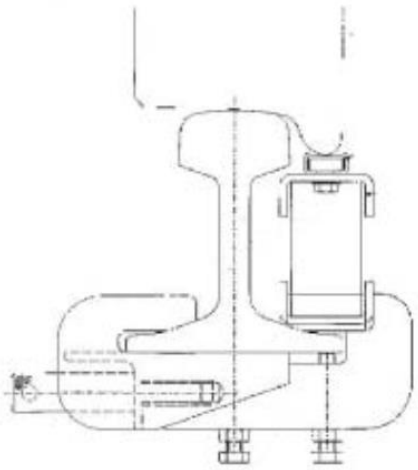


Figure 1.5 S*chematic diagram of a pressure plate to monitor the relative distance between the head and foot of the rail using a passing wheel* (Barke & Chiu, 2005).

The main advantage of these track-mounted techniques is that once installed they can operate continuously without impacting on the running of the rail network. If a track-mounted technique can be developed that can effectively monitor the structural health of rails and crossings, then this would result in minimum impact on the rail network operation. Currently however, there is not an effective universal monitoring method of this type avaible so other monitoring methods are required for effective SHM.

1.3.2 Rolling stock mounted SHM techniques

Track-mounted sensors are limited to a particular area requiring a large quantity of sensors to monitor the entire network. One solution to this is mounting the sensors directly on to rolling stock. This allows a single sensor array to monitor a larger section of the network. As discussed in section 1.2 the primary damage mechanism to rail failure currently is RCF caused by rolling stock. It is therefore reasonable to assume that the structural health of the rail can be ascertained by monitoring crack growth only when rolling stock is passing over a particular area. If this is the case continuous monitoring of a rail section when no rolling stock is passing over is not required making the sensors on areas of the network where rolling stock is not currently passing at any given time effectively redundant. It may therefore be more cost effective to install

sensors directly onto rolling stock. However, this method does have drawbacks, since, for these methods to be effective they must have the capability to operate at high speeds and effectively monitor all defects. In some cases, rolling stock mounted SHM techniques require specialized trains to operate from. These reduce the capacity of the network for freight and passenger rolling stock.

In section 1.3.1 it is mentioned that the geometry and location of the sensors was important from track-mounted sensors to ensure no interaction or collision with the passing rolling stock. The same principle applies to rolling stock mounted SHM systems. The running surface of the rail head should be uniform allowing sensors to potentially run along this surface and non-contact systems such as cameras can be mounted above this surface with to interaction issues. However, sensors that require access to the side of the rail are heavily limited. Cables bolts, fish plates and other rail infrastructure are commonly attached the side of the web of the rail causing interaction issues for any sensors running long side the rail. Side-monitoring systems would be further complicated when passing over crossing as the rail is no longer a uniform surface. Rolling stock-mounted sensors are not able to monitor the foot of the rail unless they can do so from the head of the rail.

Due to these challenges rolling stock SHM systems have been limited to those monitoring the head of the rail as well as monitoring changes in track orientation, although ultra-sonic testing (UT) systems can provide information about defects in the rail web and rail foot, not however 100% in the case of the latter.

Sensors arrays have been deployed on a range of rolling stock, both freight and passenger, the flagship of rolling stock mounted SHM techniques used in the UK is Network Rail's new measurement train (NMT) shown in Figure 1.6. As an adapted intercity high-speed train, the NMT is able to operate on high-speed networks with minimal impact on the operational

capacity by maintaining the network speed. This is a limitation of many rolling stock-mounted sensors, which require low speed operation showing a marked reduction in resolution and capability with increasing speed.

Figure 1.6 *Image of the flagship new measurement train used by Network rail to monitor the structural health of the rail network* (*New Measurement Train (NMT) - Network Rail*, 2022).

Figure 1.7 shows the effect of increasing speed on the resolution of a multi-sensory array developed by Aharoni *et al* (Aharoni et al., 2002). Figure 1.7 clearly shows the drop in monitoring resolution with increasing speed. Even where rolling stock-mounted systems require lower speeds than the normal operational speed of a line, they still have a use in the multi-approach method currently used for SHM. Where a high-speed system such as the NMT have detected a fault, a lower speed and higher resolution system can subsequently be deployed to confirm this.

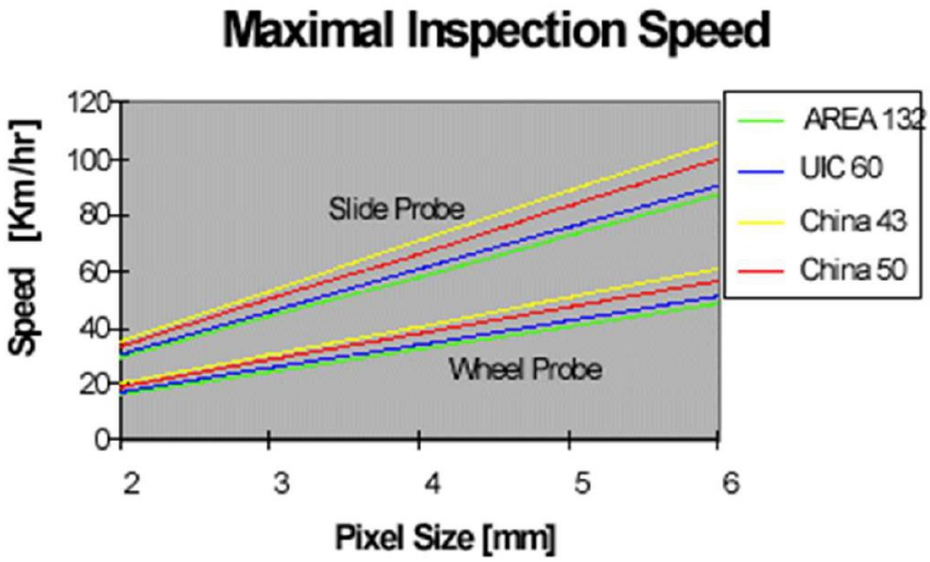


Figure 1.7 *Plot of the pixel size (resolution) vs the operational speed of multisensory array system mounted on either a rail surface slide array or a wheel monitoring array* (Aharoni et al., 2002).

Where a severe fault has been identified an emergency speed restriction (ESR) is likely to be applied to the section requiring monitoring and therefore the slower speed of these arrays will have minimal impact on the operational speed. In addition to this a slow speed monitoring asset will still cause less disruption than the track closure required for manual inspection of the rail which would otherwise be required. These manual methods and their impacts are discussed in section 1.3.3.

The NMT is able operate at higher speeds than the arrayed system to which Figure 1.7 refers to, due the choice of systems it operates. The NMT's primary focus is the detection of rail twist, changes in the rail head causing the wheel to bounce and the gauge of the rail. These faults are of relatively large and easy to detect using accelerometers measuring the rolling stock response to passing over the fault. Laser and image analysis are also used to detect damaged or missing assets. Whilst this system provides a good overview and detection of many faults it is limited in the detection of cracks and the subsequent classification of their severity. For example, a

RCF crack growing parallel to the running surface may become very severe before changes in the overlaying track geometry become detectable.

Rolling stock mounted monitoring systems clearly have a place in rapidly monitoring large areas of track and for multi-asset monitoring of these areas, the NMT alone covers 115,000 miles every year. These systems are however limited in their detection resolution and the types of failure they can detect. Greater resolution can be achieved by operating at lower speed to allow techniques such as ultrasonic arrays to be deployed. This however presents issues particularly on high-speed lines. The primary use of these systems is therefore for initial detection of faults, but they must still be used in conjunction with other SHM techniques to detect all critical faults.

1.3.3 Mobile manual SHM techniques

Mobile manual techniques in this study refer to those techniques which are not permanently mounted on to rail assets and instead require maintenance personal to apply and operate them. This work is normally carried out on temporarily closed sections of a network to ensure the safety of the operatives, therefore these techniques have a large impact on the operation of the network when operating. Despite this clear disadvantage in comparison with line or rolling stock-mounted SHM techniques these methods are widely used. As many of the constraints that apply to permanent SHM assets do not apply. It is therefore possible to use a wider variety of SHM techniques, many of which are highly accurate, as they will not be limited in size and operational speed. Within the rail industry these techniques are therefore called in to confirm and classify a defect detected via other methods.

As has been discussed the growth of subsurface cracks is of concern and challenging to monitor as they are able to grow with no clear surface indication of the severity. These subsurface cracks are by no means impossible to monitor with techniques such as ultrasonics, magnetic flux leakage, x-rays and eddy current testing used widely across a number of industries to monitor subsurface crack growth. Both ultrasonic testing (Figure 1.8), eddy current testing and magnetic flux leakage are currently used in the rail industry to confirm and classify rail defects (Magel, 2011). Table 1.1 shows a more complete list of techniques applied by Network Rail to the monitoring of the UK network. There are however certain situations where these methods are not suitable, and it is these which present particular risk to reliable rail monitoring as they are usually also challenging for other monitoring methods.

Figure 1.8 *Track maintenance operative investigating the structural health of a rail section using ultra sonic techniques* (*Ultrasonic Rail Testing - Pandrol,* 2022).

Table 1.1 *Table listing a number of manual non-destructive techniques used by network rail to monitor the structural health of the UK rail network, and the advantages and disadvantages of each technique.*

Technique	**Advantages**	**Disadvantages**
Dye penetration inspection	- Highly sensitive to small surface defects - Not material limited. - Usable on complex and large geometries. - Cheap.	- Surface preparation required. - Requires direct access. - Only surface breaking defects detectable. - Sensitive to surface finish.
Magnetic Flux leakage	- Does not require surface cleaning. - Affective for corrosion detection.	- Material must be ferromagnetic. - Probe distance must remain constant. - Detection speed fixed. - Cannot detect gradual changes
Magnetic particle inspection	- Detects surface and near-surface defects. - Under certain conditions can be detected visually.	- May require surface preparation. - Only applicable to ferrous materials. - Direct access required.

Radiography	- Not limited by material type - Can inspect assembled parts. - Subsurface and surface defects detectable. - Permanent record provided. - Minimal surface preparation required.	- Requires access to both sides of component. - Time consuming. - Safe handling of radiation required. - Sensitive to applied direction. - Expert user required. - Expensive.
Visual Inspection	- Easy to carry out. - Cheap (no specialised equipment).	- Expert required. - Influenced by environmental factors. - Direct access required. - Subjective. - Clean surface required
Ultrasonic testing	- Can detect non surface breaking defects. - Can be guided along structures.	- Requires direct access. - Absorption limits penetration depth. - Changes in material will affect propagation.
Automated computer vision inspection	- Less subjective than visual inspection. - Easy to carry out. - Can be operated at higher speeds than visual inspection. - No expert operator required.	- Sensitive to novel defects. - Direct access required. - Clean surface required.

An example of this is ultrasonic testing. Ultrasonic testing can be both rolling stock-mounted and deployed manually. The advantage of deploying the system manually is that a specific area can be inspected in detail and high resolution is feasible. The probe can be handled in different orientations as required, which is one of the advantages of manual techniques. Many of these techniques are highly orientation-dependent meaning that crack in certain orientations may be missed if only rolling stock methods are used.

1.4 Acoustic emission (AE) monitoring

No single current monitoring system is available to rail SHM personnel that is capable of detecting all defects effectively (Amini, 2016) and with no impact of the operational speed and capacity of the network, of particular concern are cast manganese crossings for which many current non-destructive testing (NDT) techniques cannot be applied due to its non-magnetic properties (Huang, 2016). To address this an array of NDT systems is deployed. To allow the rail network to achieve 24-hour operation without delays and cancellations due to monitoring and unexpected failure effective real time SHM techniques must be developed.

AE monitoring has been widely used across different industries to monitor cracks within structures and interest has been growing in the use of this technique to monitor crack growth in real time in a variety of settings including the rail industry (Amini, 2016; Bhuiyan, n.d.; Roberts & Talebzadeh, 2003; Yu et al., 2011; X. Zhang et al., 2014, 2015).

1.4.1 AE monitoring principles

AE monitoring is the monitoring of changes in a material by detection and analysis of elastic strain waves traveling through a material generated by the rapid release of strain energy within the material (Bruzelius & Mba, 2004).

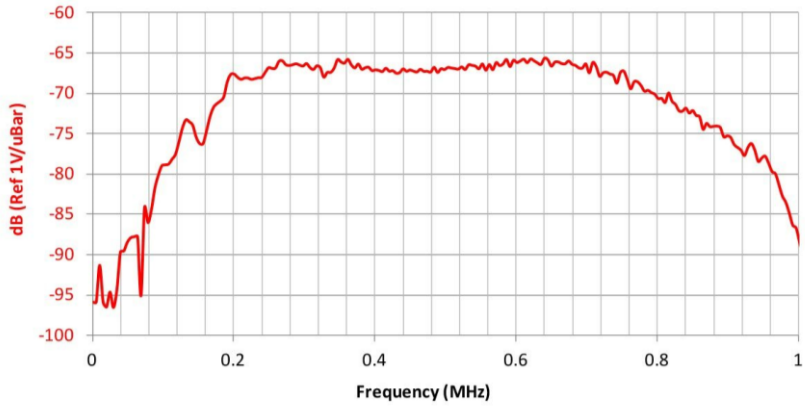
AE monitoring is by no means a new technique with the underlying concept being understood since antiquity, that is that acoustic waves propagating through a material, historically detected as audible sound, indicate the presence of a crack growing within the material. Whilst in this form it is not possible to quantify the level of damage it is possible to guess at the severity of crack growth events by the amplitude (volume) of the audible sound. This principle still applies today with crack growth visible on AE waveforms as high amplitude spikes (Sikorski, 2013). Dunegan *et al* found that high-frequency low amplitude AE signals indicated a healthy system

whereas high amplitude spikes indicated a damaged sample and were detected just before failure on a unflawed sample (Dunegan et al., 1968).

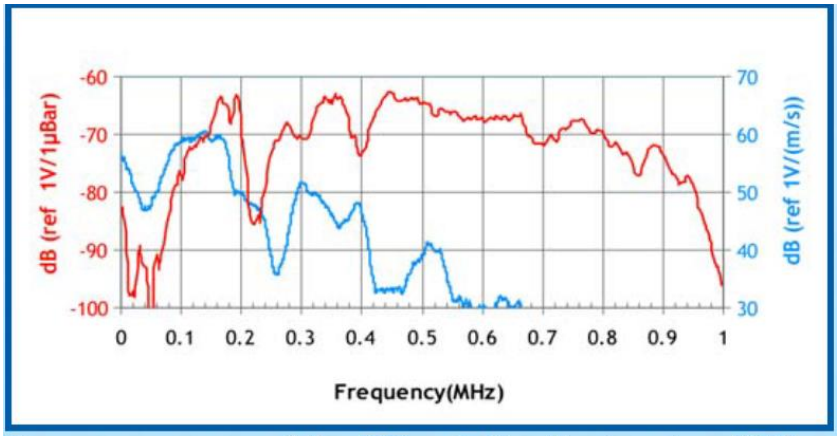
AE differs from other NDT techniques as it utilises data generated directly by the source allowing continuous real time monitoring of the structural health of an asset and without the need to have prior knowledge of the location of a defect as the location can be ascertained directly through the analysis of these signals (Swindlehurst, 1973). The scale of the sources of AE signals can vary from the monitoring of earthquakes down to micron level crack growth, the crystallisation of polymers and changes in microstructure during manufacture such a twinning (Bassim et al., 1994; Han et al., 2014; Yu et al., 2011). During fatigue crack growth in rail steel Shi *et al* reported detection of crack growth rates as small as $2.5X10^{-6}$mm/cycle (Shi et al., 201804). This accuracy ensures that models based on the AE data are highly accurate and are capable of detecting not just the presence of crack growing but also provide information of changes in crack growth due to the presence of defects. This level of detail allows that any models utilising this technique can go beyond simple crack growth dynamics but also adjust for the presence of inclusions which may affect the total life of the rail asset.

To detect the AE signals at the surface of the material piezoelectric sensors are commonly used, although other methods such as fibre optic sensors can also be used (Willberry et al., 2020). Piezoelectric signals can either be wideband or resonant sensors differing in their frequency response to a signal this difference is shown in Figure 1.9. Wideband filters show a flat response across all frequencies to which they respond giving no distortion of the signal. In contrast resonant sensors give a higher response to signals as particular frequencies. Resonant sensors in this way distort the signal. Whilst on the face of it this seems counter-productive a correctly selected resonant frequency allows key frequencies to be amplified differentiating them from the background noise.

(a)



(b)



Frequency response of the R50α. Calibration based on ASTM E1106; Calibration based on ASTM E976.

Figure 1.9 *Frequency responses of (a) a wideband piezoelectric sensor* (*F50α - 200-800 KHz High-Sensitivity Flat Frequency Response AE Sensor, by Physical Acoustics*, n.d.) *(b) a resonant piezoelectric sensor, showing the difference in the frequency response to an AE signal* (PAC, 2005).

Piezoelectric sensors are mounted directly onto the surface on the material being monitored and convert the AE waves reaching the surface of the material into electric signals. These signals have a low amplitude and so must be amplified. This amplification is done using a pre-amplifier and an amplifier. This ensures the complete AE signal is recorded with minimal loss. The generated electric signals are passed through a digital acquisition (DAQ) recorder to convert the analogue signals to a digital signal before being recorded by a monitoring computer. This set up is shown in Figure 1.10. By using two or more sensors the location of the source of the defect can be ascertained by comparing the time difference between the signal arriving at each sensor (Culwick, 2019). Recent work by Strantza *et* al and Zhang *et al* has sort to

overcome the need for more than one sensor in locating the source of the AE signal (Strantza et al., 2017; J. Zhang et al., 2016).

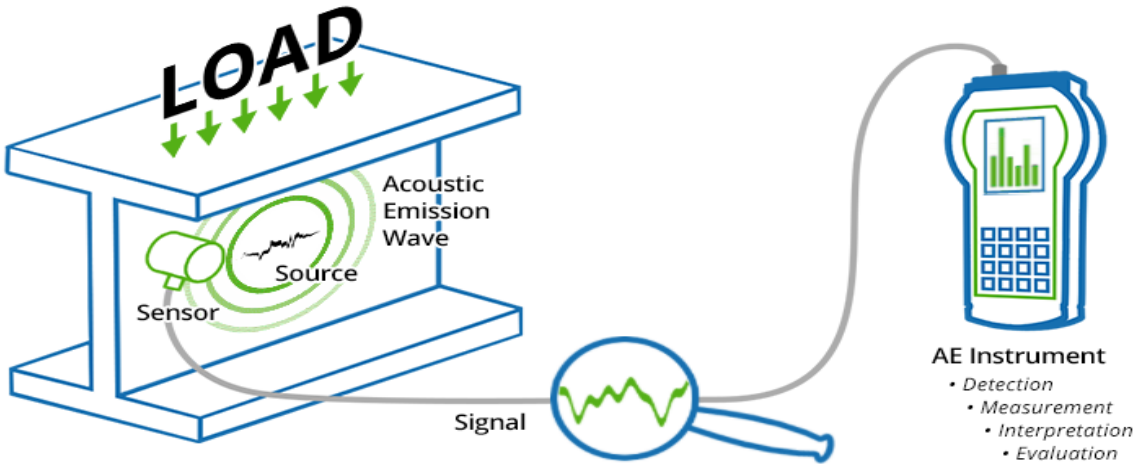


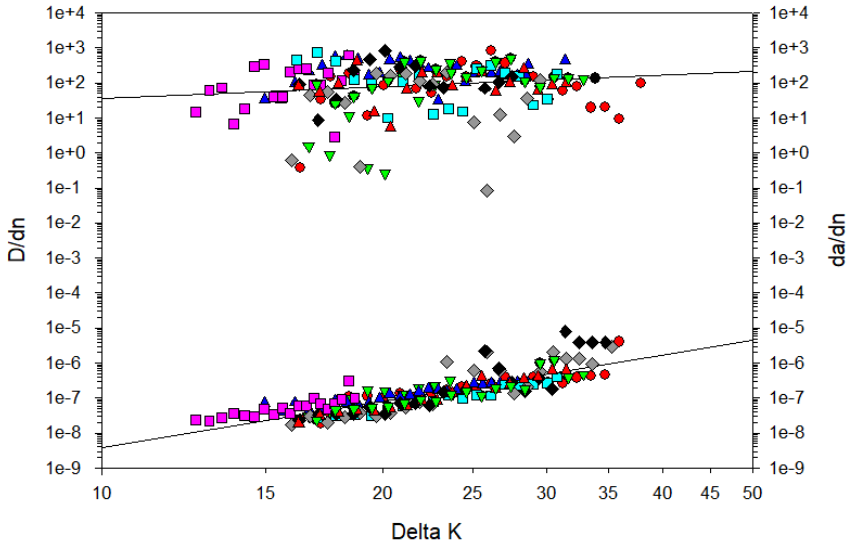
Figure 1.10 *Schematic diagram of AE monitoring equipment (adapted from animation by Physical acoustics* (PAC, 2022)*).*

Unlike conventional NDT techniques, which monitor the size and geometry of defects, AE is able to directly measure the stress intensity factor (Dunegan et al., 1968). This opens the potential for highly accurate models of the structural health of an asset through the development of a digital twin of that asset which is updated using the AE data giving a prediction for the remaining useful life of an asset (Roberts & Talebzadeh, 2003). Several authors have proposed that this may be possible without continuous monitoring instead using short bursts of AE signal (Ennaceur et al., 2006; Roberts & Talebzadeh, 2003). This work is focussed on the development of a continuous real time monitoring system that gives real time data to infrastructure managers. This ability of the technique to work with only short periods of data does however add redundancy to the system by ensuring the prediction model isn't negatively impacted should there be a failure in the system leading to a loss in data.

1.4.2 AE analysis

The way in which the AE signal is analysed depends on the way in which it is recorded. AE data can either be recorded as hit driven data (HDD) or time driven data (TDD), which in this case can also be referred to as parametric data and complete waveform data respectively. The difference in these approaches is shown in Figure 1.11.

(a)



(b)

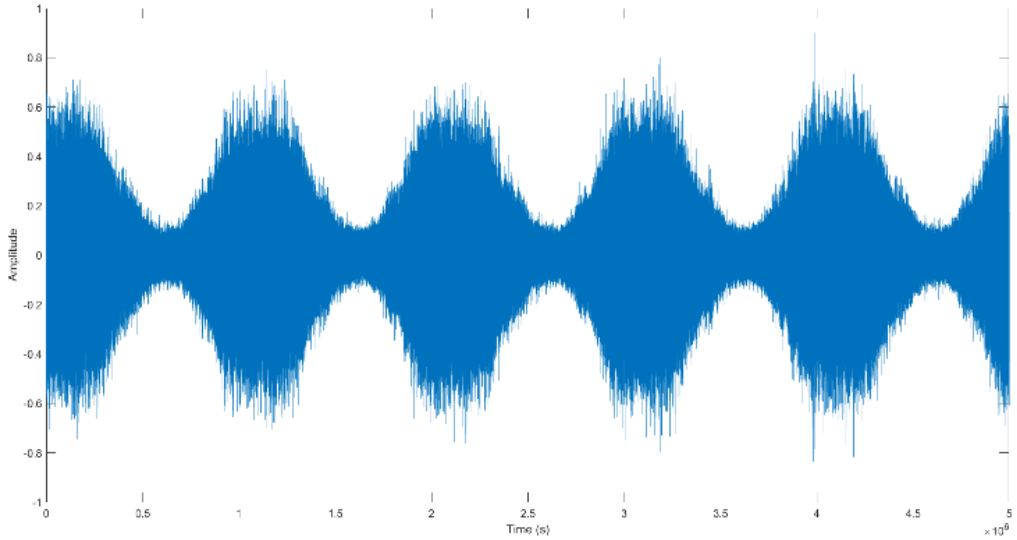


Figure 1.11 *Representative examples of (a) the HDD from the commercial system, (b) and the TDD from the customised system.*

HDD acquisition is both threshold and time controlled. A series of thresholds is set which the signal must be within for it to be recorded once these thresholds are met the signal is then time limited in terms of the signal acquisition time. This leads to a series of packaged data hits. Within these hits parameters can be recorded, commonly amplitude, counts, energy, duration and rise time. Figure 1.12 shows how AE amplitude, counts, energy and duration are calculated from an AE waveform using HDD acquisition. The most commonly used parameters are AE energy, Counts and Duration (Roberts & Talebzadeh, 2003).

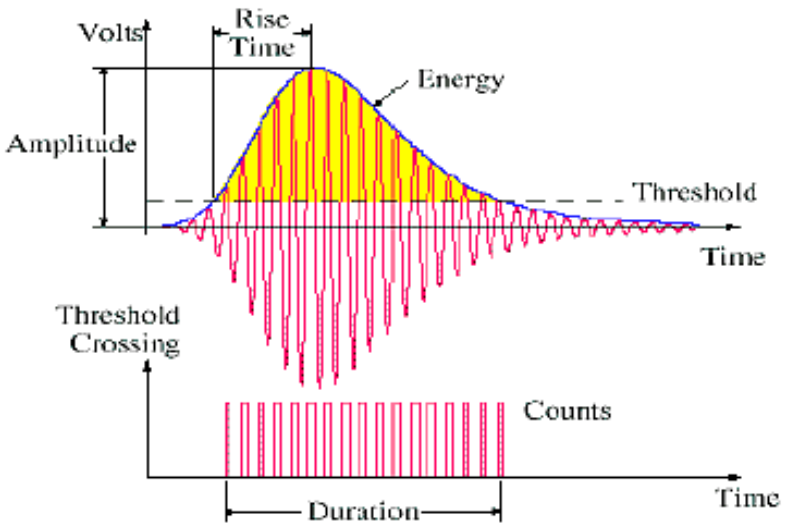


Figure 1.12 *Schematic showing how the AE amplitude, counts, energy and duration are calculated from an AE waveform using an HDD approach* (Uddin A.K.M, 2002).

The thresholds for use in HDD must be carefully selected to ensure that the AE parameters are representative of the received AE signal and capable of detecting all crack growth events (Gagar et al., 2015), whilst excluding any excess data. These parameters will be different for each material and in some cases sample geometry (Yu et al., 2011).

This approach has the advantage that it reduces the overall data recorded excluding large parts of the AE signal (Swindlehurst, 1973). These parameters are easily compared to crack growth and require little analysis to show the presence of a crack within a structure. Primarily, trend analysis is used for HDD and in many situations this is sufficient, to this end most commercially

available AE system make use of this approach (Sikorski, 2013). Current applications of AE use a HDD approach either continuously monitoring a system and triggering and alarm if there is an increase in any of the parameters, or where continuous monitoring is not utilized recordings are taken at intervals and compared to look for changes in the acoustic parameters which may indicate a change in material behaviour. These methods are in many applications effective and indicating the presence of a crack but not any indication of the nature of the crack. Further investigation would therefore be required using other monitoring methods to determine the structural health of the component.

The HDD approach is however limiting in that only a small part of the potential data received is utilised this has made the analysis of crack growth stages in steel has proved challenging with a multi parameter approach required (Ould Amer et al., 2013).

To set the frequency threshold the required frequencies must first be known. As with all thresholding the aim of frequency thresholding is to remove background noise whilst maintaining the relevant AE signal. The frequencies corresponding to crack growth differ between materials, with the frequency range for each material ascertained through the use finite element analysis (FEA) or through experimental data. Table 1.2 shows the AE frequency ranges obtained from studies into a range of materials. It can be observed in Table 1.2 that even for the same material the frequency ranges are not set with each report differing slightly in the range selected. It is generally agreed that the AE frequency band associated with defect growth is 20 kHz-1200 kHz (Vallen, 2002). Some studies have further refined this to identify specific frequency ranges associated with different crack behaviours. For example, under real world conditions Li *et al* found that AE signals generated during crack propagation had a frequency range of 400-650 kHz, whereas those generated during crack closure were in a lower frequency range of 100-350 kHz (D. Li et al., 2017).

Table 1.2 *Frequency ranges reported in previous studies for a range of materials.*

Material/Application	**Min (kHz)**	**Max (kHz)**	**Other info**	**Source**
Aluminium: Al-2024 T3 test tokens	250	400	Crack surface interaction: 30-100 kHz centre frequency 350 kHz frequencies associated with all sources up to 600 kHz	(Bhuiyan et al., 2018)
General	100	500	Can range 20 kHzto1 MHz changes in frequency 20-30kHz dur to changes in lubricant 20to 100 will	(Amini, 2016)
General	100	300		(Swindlehurst, 1973)
General	20	1 MHz		(Ono, 2011)
rail	20	1MHz	Most applications100-500	(Gunel et al., 2017)
Ferritic steel	50 200 200	200 250 500	-Sample deformation -Brittle crack growth -Transgranula cleavage of ferrite	(Kostryzhev et al., 2013)
Wheel defects	140 140 300	220 280 650	- Noise - Defects - Distinct defects	(X. Zhang et al., 2015)
	20	1MHz		(Bruzelius & Mba, 2004)
	100	150		(J. Zhang et al., 2016)
Axle bearings	150	250	Wayside detection	(Amini et al., 2016)
6061-T6 15%SiC particulates	125	250		(Shan & Nayeb-Hashemi, 1999)
Cast manganese	100	700		(Shi et al., 2017)
Rail steel	100	1000		(X. Zhang et al., 2014)
Rail steel	20	1000		(D. Li et al., 2017)
R260			Filtering 100-1000 and 200-1000 in a nosier environment	(Shi et al., 201804)
316 stainless	250	500	Applied as a filter	(Moorthy et al., 19941101)
Switch rail	100	150		(J. Zhang et al., 2016)
Full pearlitic	50 200	200 250	- deformation - brittle fracture	(Kostryzhev et al., 2013)

	0	50	- ductile fracture	
Fully ferritic	0	50		(Kostryzhev et al., 2013)
Dual phase	0-100, 150 550 100	175 600 120	- Deformation - Brittle crack growth - Ductile fracture	(Kostryzhev et al., 2013)
Roller bearings	50	450	- 175 key frequency for failure amplitude increases as the crack grows	(Eftekharnejad et al., 2011)

Relying on frequency thresholding alone to reduce background noise is only suitable where the background noise is constant. This is not the case under actual world conditions therefore, other approaches are required to remove background noise (Sikorski, 2013). TDD is only time controlled, although in reality some thresholds are applied such as minimum amplitude and frequency bands to remove back-ground noise, the TDD approach therefore records the complete AE waveform rather than AE parameters. By capturing the complete waveform less of the AE data is filtered out, thereby providing more data from which to ascertain the structural health of a material. More complex analysis is required for TDD compared to HDD requiring large data computational capability. This present study for example will utilise machine learning for the analysis of the TDD waveforms.

All methods used to analyse the acoustic waveform seek to extract key features that are indicative of the current structural health of the material and the crack growth behaviour. An important challenge to this is the extraction of those features only related to the behaviour being is monitored and not those associated with background noise. The way in which this is done various depending on the setting in which the data was collected. In a laboratory setting the loading conditions can be tightly controlled and background limited. When applied to in field operation in rail environment this is not possible with loading being stochastic and source of background noise uncontrollable and of varying amplitude.

Mathematical approaches are predominantly used to extract signal features. Spectral kurtosis is a widely studied mathematical approach. using the principle that AE signal are short bursts of energy spectral kurtosis has been successfully used to extract the AE signals from background noise under laboratory conditions (Amini et al., 2016; Antoni & Randall, 2006; B. Chen et al., 2014; Eftekharnejad et al., 2011; Sawalhi & Randall, 2004). The extracted signals can then be further analysed to predict the structural health condition. An alternative approach is to analyse the complete waveform and decompose that signal before extracting key features. Examples of this approach are to apply Fast Fourier transform (FFT), Route mean Square (RMS) or Continuous wavelet transform (CWT) to the complete signal and then analyse the resulting data to extract features (Amini, 2016; Vallely, 2015). These approaches have successfully been used to extract indicative features particularly in detecting the presence of defects (Papaelias et al., 2018).

All these mathematical approaches extract key signal features however require further analysis. These features still represent a large volume of data which must be analysed Siddique *et al* and Jayaswal *et al* both proposed the use of machine learning approaches to analyse these signals (Jayaswal et al., 2011; Siddique et al., 2003). Recent work has looked into this with both Suwansin and Phasukkit, and Chen *et al* developing machine learning methods for use in the rail industry (S.-X. Chen et al., 2021; Suwansin & Phasukkit, 2021). There has been a lot of interest in the use of these approaches for the analysis of failure in composites. These studies have successfully classified different failure methods in these materials (Muir et al., 2021). Some success has already been achieved in classifying crack growth in steel with Barile *et al* able to use AE to distinguish between three states; crack initiation, stable crack growth and unstable crack growth (Barile et al., 2015). Shan *et al* and Mohammad *et al* also separated the different stages of crack growth by applying Weibull based models to look at the distribution of signals at the different stages (Mohammad et al., 2014; Shan & Nayeb-Hashemi, 1999).

The prediction of the structural health of a material using AE monitoring relies on the elastic strain waves reaching the AE sensor being indicative of the signal generated at the source. Rails are ideally suited this with the elastic stress waves able to travel long distances with very little attenuation (J. Zhang et al., 2016). There will always however be some attenuation of the AE signal limiting the maximum distance a sensor can be from an AE source (Moore et al., 2005). Attenuation of the AE signal within a structure is caused by changes in geometry, dispersion of the signal, scattering and diffraction, and energy loss (Moore et al., 2005).

1.4.3 AE Structural health monitoring in the rail industry

With the growing interest in AE for SHM research has relatively recently turned to the application of this technique to the rail industry. In 2004, Bruzelius and Mba found no publication of the application the AE to rail defect detection although monitoring of crack growth in steels was reported by Bassim in 1984 (Bassim et al., 1994). Since then, the body of work has evolved with many aspects of this application now being explored (Bruzelius & Mba, 2004).

The testing and development of monitoring systems within the live rail environment is challenging. Firstly, access to the rail is tightly controlled due to safety concerns. Secondly the ability to control the conditions under which testing is carried out is limited. For monitoring systems to be effective especially those based off pattern recognition all crack growth conditions must be monitored. However, predicting where a failure may initiate is not possible. Therefore, short of continuously monitoring the entire network and then forensically assessing any failure it is not possible to train and test a system at least initially on a live system. This has led to most studies to date focussing on studying AE for rail assets under laboratory conditions.

In comparison with the actual rail environment, laboratory opens the opportunity for carefully controlled conditions, artificially induced defects, and effective control on the AE acquisition.

Lindley *et al* and Li *et al.* both utilised these control methods to separate the acquisition of signals during crack propagation and crack closure during fatigue crack growth allowing them to study the AE signals linked to these behaviours (D. Li et al., 2017; Lindley et al., 1978). Lindley *et al.* achieved this separation by the use of voltage gates recording AE signals only during crack propagation.

Only recording AE signals during crack propagation has been explored by a number of authors. Even with voltage gates defining when there is crack propagation as opposed to just loading of the sample is challenging. To address this, several studies have been carried out, as when the sample is in the top 20% of its loading cycle (Rabiei & Modarres, 2013; Roberts & Talebzadeh, 2003; Yu et al., 2011).

Li *et al* took a different approach (D. Li et al., 2017), to make the analysis capable of future practical implementation they used a wavelet power index to separate the crack propagation and crack closure signals and is defined as:

$$R = \frac{\max\{\mathrm{WP}_i | \mathrm{WP}_i \subset B_2\}}{\max\{\mathrm{WP}_i | \mathrm{WP}_i \subset B_1\}}$$

Equation 1.1 *Wavelet power index equation used by Li et al to sperate crack propagation and crack closure signals. Where B_1 and B_2 are the frequency bands 100-350 kHz and 400-650 kHz respectively. With the former relating both the crack propagation and closure, the later related to just crack closure AE events* (D. Li et al., 2017).

Once separated they developed a crack length classification method based on crack closure AE waves finding a positive correlation between the AE count rate and the crack length. Of significance to the development of a robust SHM system this method was found to be effective without prior knowledge of the structural health of the material. Li *et al.* also closed the window

of crack growth to the top 10% of loading and crack closure to the bottom 10%, this contrasts with 20% defined by the other studies discussed.

Despite advances in laboratory experimentation in the field of AE monitoring very few field trials have been carried out to verify the effectiveness of these approaches under real world conditions. It is clear that those that rely on voltage gates and prior knowledge of the structural health of the material are limited in their application. These studies do however provide data which is useful in understanding and development of practical systems. Recently studies have begun testing the use of AE systems on live networks including work by the author.

Both Huang and Grosse *et al.* tried to overcome attenuation limitations by characterising frequency patterns associated with defect growth (Grosse et al., 2004; Huang, 2016). By analysis using spectral coherence between pre-characterized patterns and the signals detecting the variation in the signal can be ascertained. This method proved effective under actual operational conditions proving that despite the increase in noise level when compared with the laboratory and greater distances between the source and the sensor AE monitoring can still be effective.

Vallely tested the effectiveness of AE system to detect crack growth in rail wheelsets and cast manganese crossings (Vallely, 2015). By making use of rail sidings, they were able to induce artificial defects into rolling stock wheels and bearings. An AE system was then able to detect the presence of these defects using waveform analysis. Instead of artificially inducing defects into cast manganese crossings a target approach was used to select crossings near the end of their life. Using this approach Vallely successfully classified the crossings as either healthy or damaged.

These advances in the use of AE on live rail combined with the laboratory experimentation have proven AE has the potential to be the basis for a real time structural heath monitoring

system. For implementation on a network work is still required, the system must go beyond the detection of a crack, the basis for current field trials, and classify the length and therefore severity of a defect. This has been successful under specific laboratory conditions, but these studies have struggled to make the step required to practical use. Work is particularly needed in the automation of the AE analysis removing the need for each signal to be analysed by an expert user. Machine learning approaches are providing the backbone, for example the work of Chen *et al* who used transfer learning to analyse AE waveforms for rail condition monitoring (S.-X. Chen et al., 2021).

1.5 Aims and Objectives

This study aims to further the development of a real-time structural health monitoring system for use on rail infrastructure, using Acoustic Emission (AE) techniques. To achieve this, this study will use two different AE monitoring systems. Firstly, a commercially avaible AE system procured from Physical Acoustics and secondly a custom system developed at the University of Birmingham.

The commercial system will be used to assess the overall capability of AE as a monitoring technique for three steels used for the rails, R220 and R260 used for plain track and cast manganese used for rail crossings. The custom AE system will then be used to record the complete AE waveform as opposed the packaged acoustic data recorded by the commercial system. Analysis and classification methods will then be investigated leading to the development of a machine learning algorithm capable of classifying the severity of a crack during fatigue crack growth. Finally, field tests will be completed to confirm the effectiveness of this system under real world conditions.

Chapter 2

Experiments

This work uses several different approaches for analysing the AE data generated during fatigue testing under laboratory conditions. In all cases, the same data collection methods and equipment were used. Two primary data collection systems were used, a commercial system provided by Physical Acoustics (PAC – now Mistras) and a custom system developed jointly by Krestos Limited and The University of Birmingham. Laboratory testing was completed at The University of Birmingham. Whilst there is greater background noise on the live rail network the AE signals generated from crack growth in both scenarios is expected to be the same allowing transfer of knowledge. Initial work on testing the system under real world conditions was carried out, the methods used for this are also described in this section.

2.1 Steel composition and properties

Understanding the properties and microstructure of the steels being investigated is important as the AE signals generated by each material will be different, as the features and defects which a growing fatigue crack will interact with are dependent on the microstructure of the material being monitored. As this work seeks to develop a system that can monitor crack growth and crack growth behaviour, the expected microstructure and defects must be understood. Whilst by no means exhaustive, particularly in the case of the cast manganese steel, below is detailed the composition, properties and commonly encountered defects in the steels investigated in this study, this is in addition to those defects more generally found in rails detailed in section 1.2.

Three different steel grades which are widely used across the UK rail network have been investigated in this study; cast manganese steel, used in manufacturing rail crossings, and R220

and R260 steel grades commonly used for plain track rails. Samples of cast manganese steel were taken from a hot-rolled 10 mm-thick plate procured from West Yorkshire Steel. For both R220 and R260 steels samples were cut from the web of rails removed from service and provided by Network Rail. The compositions of these steels are provided in Table 2.1.

Table 2.1 *Table showing the composition in weight percentage of three steels used in this study. These values are defined by EN13411 for the R220 and R260 steels and ASTM A128 for the cast manganese steel* (ASTM international, 2017; BritishSteel, 2018).

	Al	**C**	**Cr**	**Mn**	**N**	**P**	**S**	**Si**	**V**
R220	≤0.03	0.5-0.6	≤0.15	1-1.25	Not specified	≤0.025	0.008-0.035	0.2-0.6	≤0.03
R260	<0.004	0.6-0.82	<0.15	0.65-1.25	<0.008	<0.15	0.008-0.030	0.13-0.6	<0.030
Cast Manganese	Not specified	1.05-1.35	Not specified	≥11	Not specified	≤0.07	Not specified	≤1	Not specified

2.1.1 Cast manganese steel Crossings

In its hot-rolled plate form, cast manganese steel is soft and ductile but rapidly becomes work-hardened when loaded (Bal, 2018; Smith & Mackay, 2003). This leads to its superior hardness values when compared with plain track steels such as R220 and R260, giving the material high impact and wear resistance (Schilke et al., 2010). For this reason cast manganese steel is widely used across a range of industries for critical components. In the case of the rail industry, it is used for the construction of rail crossings as shown in Figure 2.1 (Rommelaere & Maujean, 2013) which are subject to high impact loads regularly. These impact loads can be up to 5 times that encountered in plain track rails, making the use of a high hardness material necessary (Guo et al., 20130730).

Figure 2.1 *In service cast manganese crossing showing high levels of wear*

The high Mn content in cast manganese steel leads to a predominantly austenitic microstructure as shown in Figure 2.2. This is due to the fact that Mn is austenite stabiliser all the way to room temperature. Depending on the installation process and operating conditions a wide range of mechanical properties can be achieved, examples of which are summarised in Table 2.2. Most notably is the high hardness value that can be achieved with hardness values up to 589 HV recorded (Alyaz, 2003).

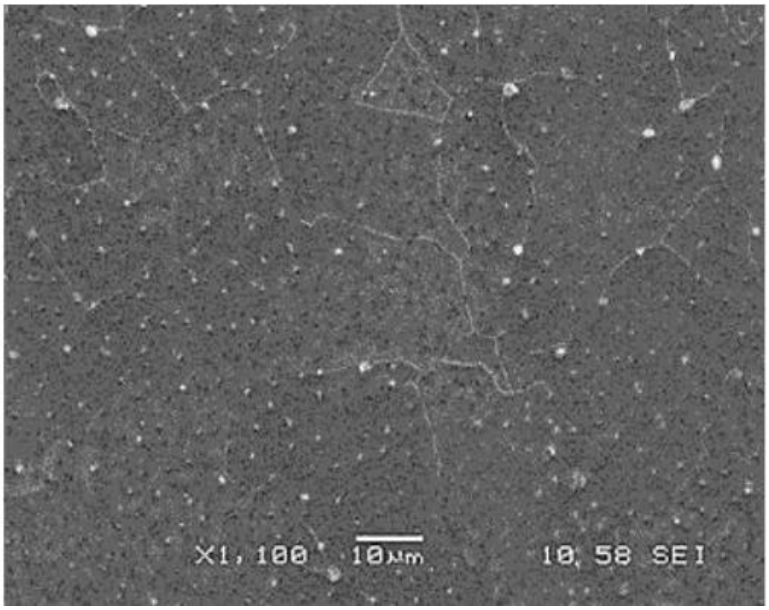


Figure 2.2 *Micrograph of the microstructure of cast manganese steel used in this study showing the fully austenitic microstructure* (Kongpuang et al., 2021).

Table 2.2 *Cast manganese steel displays a variety of mechanical properties depending on the manufacturing, pre-processing and loading conditions. This table summarises a range of these properties reported in the literature. As many of these studies have investigated work hardening primarily, a key mechanical property of this steel, they have only reported hardness values* (Culwick, 2019).

Source	**Surface Hardness**	**Yield strength (MPa)**	**Elastic Modulus (MPa)**	**Tensile strength (MPa)**	**Fatigue limit (MPa)**	**Elongation (%)**	**Toughens**	**Work hardening range yield to ultimate tensile (%)**	**Notes**
(Eck et al., 2014)	400 HV	384	$201x10^3$	724	141	0.07			Obtained from static and strain-controlled cyclic stress-strain curves at room temperature
(Peters, 2005)	370-385 HB	345-414	$186x10^3$	965	265	35-40	-	200	Hardness following explosive depth hardening on the network. All other properties are as manufactured
(Havel, 2017)	500 HB	303.37-393	-	551.58-999.74	-	15-40	-	-	Hardness maximum they were able to achieve. All other properties as manufactured given as an average from several sources.
(Chen et al., 2018)	471 HV	409	-	861	-	37.6	-	-	Results following work hardening

(Hadfield, 1914)	550 HB	-	-	-	-	30-50	-	-	Maximum hardness reported at that time all other properties in paper are as cast
(Rommelaere, 2013)	>352 HB	-	-	-	-	-	-	-	Explosive hardened
(Guo et al., 2013)	510 HV	1000	-	-	-	-	-	-	All properties in service following work hardening
(Harzallah et al., 2010)	Max 600-800 HV	400	-	1000	-	-	-	-	Hardness following work hardening all other properties as manufactured
(Lv et al., 2012)	Max 620 HV	-	-	-	-	-	-	-	In service
(Smith and Mackay, 2003)	350-400 HB 500 HB can be achieved	-	-	-	-	-	296-357	-	Hardness taken from explosive hardening in service
(Zhang et al., 2010)	395 HB	660	-	-	-	-	62 Jcm^{-2}	-	Gives varying range of hardness depending on the number of explosive shocks applied value given her is after three shocks. All other values given after hardening
(Zhang et al., 2012)	39 HRC	-	-	-	-	-	-	-	Hardness given following explosive hardening

(Zhang et al., 2008)	220 HB pre hardening 390 HB achievable with hardening Normally 360 HB	-	-	-	-	-	-	-	-
(Alyaz, 2003)	550 HB found after service to be 495-535 HB	345-393	-	690-1000	-	30-65	-	-	Hardness as in service all other values given as manufactured
(Norberg, 2010)	275-366 HV	-	-	755-859	-	-	-	-	All values following explosion hardening
(Shewmon and Zackay, 1961)	580 HV	1199.69	1544.43	-	-	19	-	-	Shock wave given by impact with a copper driver plate
(WestYorkshireSteel)	200 HB can be work hardened to 500 HB	320	-	880	-	40	-	-	As cast conditions
(So and Nagahiro, 1986)	>250HB	-	-	-	-	-	-	-	Patent for explosive hardening in crushing applications

The large variation in mechanical properties of the cast manganese steel shown in table 2.2 point the sensitivity of this steel to the conditions under which it is manufactured and installed, and the final use of the material. This inconsistency in properties is a benefit of this steel and is can be tailored to each use case, but also a problem in that the properties are easily changed and therefore is components manufactured from this steel are susceptible to sub optimum mechanical properties. The large variation in mechanical properties is mainly due to the rapid work hardening behaviour of this steel.

The exact mechanisms behind this rapid work hardening behaviour are not yet fully understood (Kang et al., 2014). However, a wealth of previous studies have shown that this is likely achieved through a combination of stacking fault formation, dislocation slip and twinning (Alyaz, 2003; Bal, 2018; Gumus et al., 2016; Karaman et al., 2000; Wen et al., 2014). It is important that the work-hardening behaviour is understood when developing a novel NDT method to monitor the structural health of rail crossings in real time, as it will have a direct effect on the materials fatigue crack growth mechanisms and wear resistance.

Kang et al. studied the effect of work-hardening on fatigue crack growth for cast manganese steel under fatigue loading (Kang et al., 2014). By showing the presence of dislocations on the fatigue fracture surface during both high and low cycle fatigue (LCF and HCF respectively), they found that crack growth was predominantly arising by plastic strain and dominated by the build-up of short-term stress fields around C-Mn clusters. This creates internal stress due to the interaction of dislocations with the C-Mn cluster resulting in a high density of tangled dislocations (Culwick, 2019). This in turn leads to rapid crack growth. Kang et al. further highlighted the observations of other studies such as that by Schilke et al. that cast manganese steel does not show normal work-hardening behaviour under cyclic loading. It instead initially hardens before softening to fracture (Schilke et al., 2010). This behaviour creates a challenge when attempting to model fatigue crack growth behaviour using traditional models.

Work-hardening of rail crossings can occur during service or through pre-loading methods such as explosive hardening, due to the rapid work-hardening capability of cast manganese steel (Rommelaere & Maujean, 2013). Pre-loading is often used in the case of rail crossings as the as cast manganese steel would wear at a greater than expected rate, due to its soft ductile properties, leading to a lower operational lifetime and potentially to premature failure of the crossing and increased maintenance requirements. There is also a significant financial burden linked to the failure of rail crossings with each crossing custom built to each location and costing up to £30k excluding installation costs which can run into the several hundred thousand pounds (Huang, 2017).

During the casting the cast manganese steel can interact with the mould reducing the Mn content at the surface. This both reduces the wear resistance and work-hardening capability of the affected surface region on the steel (Smith & Mackay, 2003). This initial wear leads to changes in geometry and lipping, Figure 2.3, requiring grinding of the crossings within the first month of service, correcting the geometry of the crossings and exposing the hardened steel below (Harzallah et al., 2010). Grinding is typically carried out according to the total tonnage passing over the crossing and whether the crossing has been explosive hardened prior to use (Rommelaere & Maujean, 2013). Pre-hardening the crossing reduces the corrective grinding required however does not remove the requirement as the Mn poor surface will not harden sufficiently.

Figure 2.3 *Example of tongue lipping on the contact surface of the rail head. The approximate original geometry is shown by the dashed line* (Schilke, 2013).

Unlike most steels cast manganese steel is non-magnetic since it is austenitic (Hadfield, 1914). This presents a challenge in detecting defects both before and during service and is further compounded by their complex geometries. Magnetic flux leakage (MFL) is commonly used to monitor internal crack growth. However, being non-magnetic this is not possible for cast manganese crossings. As noted previously whilst imaging can be used to detect surface breaking cracks, it is not able to give information on how far the crack propagates through the rail and clearly is unable to detect non surface breaking cracks. In cast manganese crossing cracks typically initiate internally running parallel to the load surface, this is shown in figure 2.4 (Lv et al., 2012). Without the effective monitoring of these internal defects, large and potentially critical cracks may occur without prior warning, leading to unexpected failure of the crossing (Lv et al., 2012). According to Peters the failure of these crossings is mainly due to manufacturing defects in the form of non-metallic inclusions and carbides at grain boundaries which act as initiation sites for crack growth (Alyaz, 2003; Peters, 2005; Smith & Mackay, 2003).The design life for crossings is typically 20 years. However, as quality control after installation and monitoring is limited during service, some crossings have been found to

fail in as little as 3 years from installation, possibly due to the presence of wet beds in the area of installation (Huang, 2017).

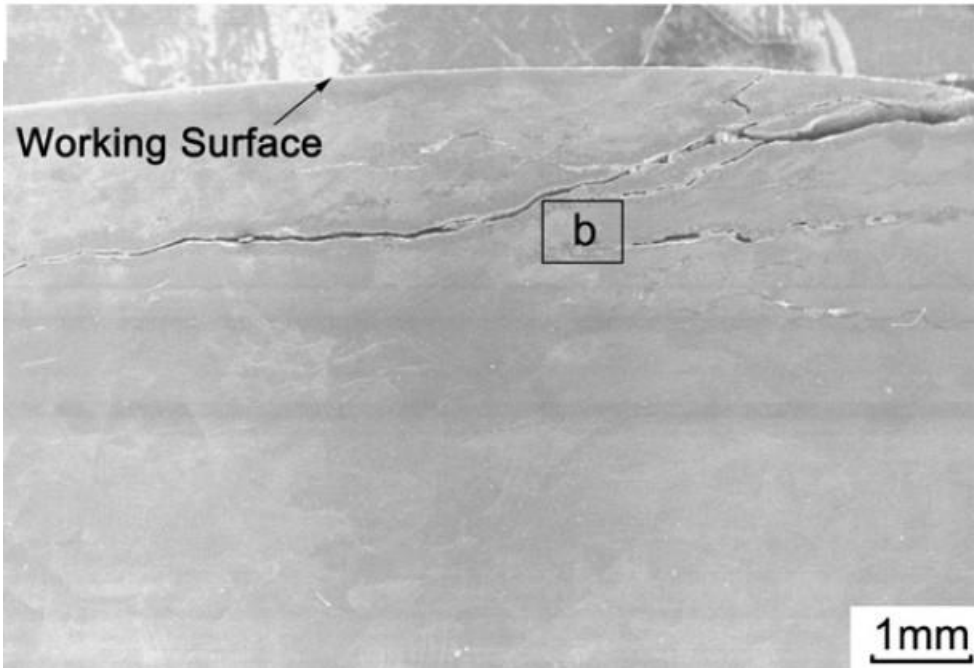


Figure 2.4 *Micrograph of a subsurface crack growing parallel to the loading surface of a cast manganese crossing nose due to RCF. Compared to the overall length of this crack only a small portion is visible on the surface making conventional monitoring challenging* (Lv et al., 2012).

Crossings allow trains to switch between lines and therefore, form bottlenecks on the network. Failure of a crossing has a significant impact on regular operations, resulting in high levels of disruption. Therefore, cast manganese crossing condition has wider reaching effects when failure occurs than in the case of normal plain rails. This combined with their high cost and no effective NDT method avaible makes the development of a novel real time monitoring method highly desirable.

2.1.2 R220 and R260 plain track

R220 and R260 grade steel are both used widely and interchangeably used for manufacturing plain track rails. These two steels have a very similar composition Table 2.1 and microstructure, primarily pearlitic with very fine lamella spacing (Yilmazer, 2013). According to the work by Perez-Unzueta and Beynon this fine lamella spacing is the key contributor to the wear resistance of rails manufactured from R220 and R260 steel (Perez-Unzueta & Beynon, 1993). They do however differ slightly in their mechanical properties as shown in Table 2.3.

Table 2.3 *Table giving the mechanical properties of the R220 and R260 steels as defined by EN14811* (BritishSteel, 2018)

	Vickers Hardness	**Tensile strength (MPa)**	**Elongation (%)**
R220	222-260	≥770	≥12
R260	260-303	≥880	≥10

As with the cast manganese steel grade, these steels are used for their work-hardening during service helping to increase their wear resistance, resulting in an increased service lifetime (Eden et al., 2005; Magel, 2011). The mechanisms of work-hardening in these materials is better understood than that of cast manganese steel. In these steels it is primarily due to the distortion of the pearlitic lamellae as shown by the work of Eden *et al* in Figure 2.5 (Eden et al., 2005).

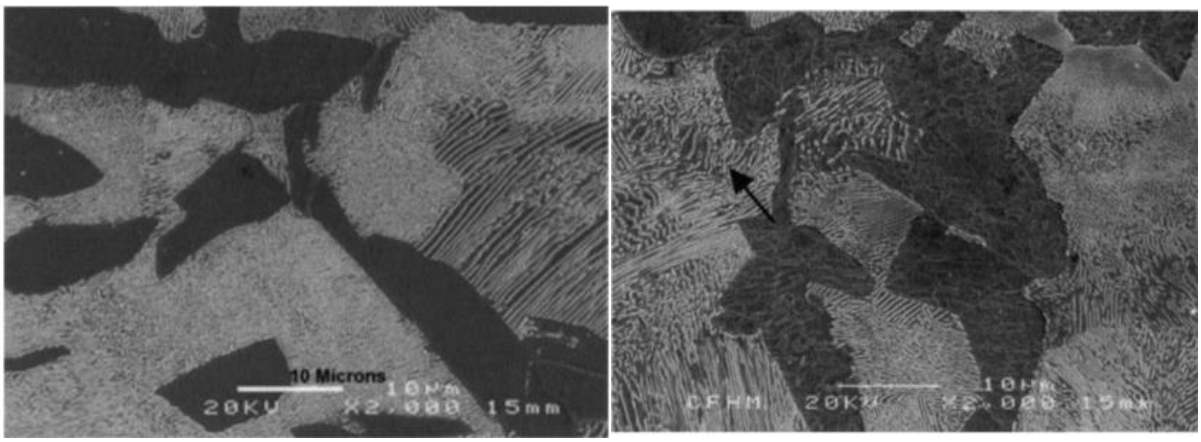


Figure 2.5 *SEM micrographs of an R220 steel sample, showing the typical pearlitic lamella microstructure of both R220 and R260 steel before service loading (a) and following distortion of the lamella structure under service loading (b) indicated by the red arrow* (Eden et al., 2005).

Much like the cast manganese steel these steels suffer from sub-optimal properties at the start of service leading to rapid wear, and it is due to a similar manufacturing fault that this occurs. During manufacture the outer layer of the steel interacts with the surrounding environment leading to decarburisation and therefore the formation of ferrite at the surface. This ferritic layer is softer than the expected pearlitic microstructure and leads to a lower-than-expected wear resistance. The ferritic layer is removed during routine grinding of the rail, used to correct geometry defects. It is possible to grind the rails before they enter service. However, this is not considered economically beneficial and is therefore not used routinely (Cannon & Pradier, 1996; Magel, 2011). Grinding must be carefully tailored to correct any defects whilst maintaining the work hardened surface layer which the work Garnham *et al* found to be to a depth of 5 mm (Garnham et al., 2010). As the samples used in this study were taken from the central part of the rail web this decarburised layer is not present and therefore the conclusions reached are based on that of a rail in service that has undergone routine grinding.

As discussed in section 1.2 a wealth of work has been carried out to reduce the presence of defects in rail steel. Despite this work defects remain present and lead to premature failure of the rail. In R220 and R260 steel the presence of MnS inclusions, shown in Figure 2.6, have been found by both Fegredo *et al* and Garnham *et al* to lead to the formation of cracks within the rail (Fegredo et al., 1988; Garnham et al., 2010). MnS inclusions are soft and ductile in comparison with the surrounding pearlitic microstructure and are elongated in the manufacturing rolling direction. This elongation is observed in the samples used in this study as shown in Figure 2.6. Like the pearlitic microstructure these inclusions become distorted

during service loading, leading to flattening and elongation of these inclusions along the loading direction and shown in Figure 2.7.

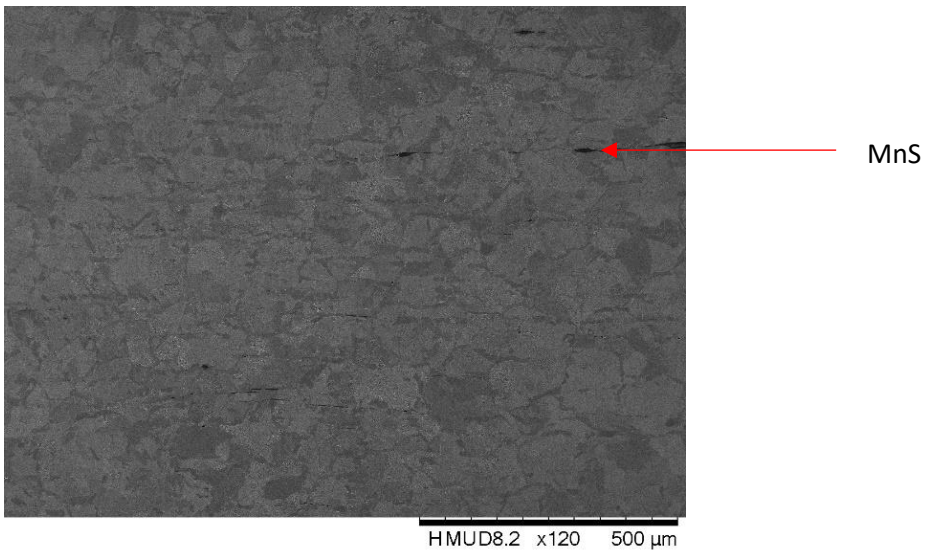


Figure 2.6 *SEM micrograph showing the MnS inclusions in a pearlitic rail sample that have been elongated in manufacturing rolling direction.*

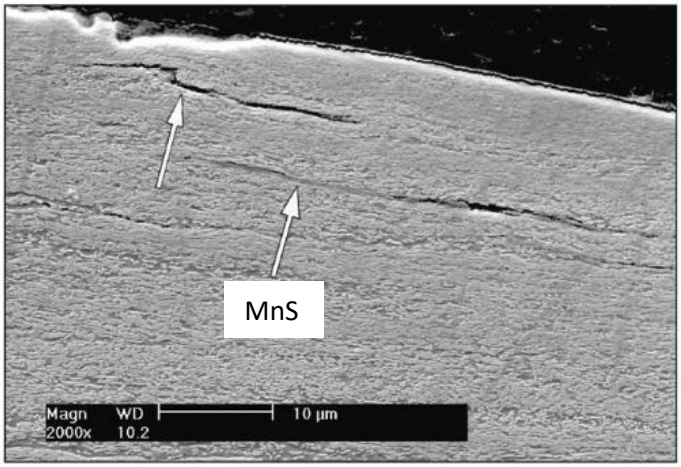


Figure 2.7 *SEM micrograph of elongated MnS inclusions near the rail head surface taken from an in-service rail* (Garnham et al., 2010).

As already highlighted one of the main issues for the NDT of cast manages steel is the lack of magnetic properties which makes many conventional NDT methods redundant. R220 and R260 are magnetic and therefore do not present this challenge. As discussed in section 1.3 there are many techniques currently used to inspect plain line rail however, these all have their own challenges making no one technique suitable. The directionality and non-continuous nature of

these techniques proves particularly challenging in the case of cracks propagating along MnS inclusions. These cracks are highly directional and propagate rapidly below the surface reducing the effectiveness of inspection systems such as the new measurement train (NMT). Novel NDT methods are therefore needed which can detect this crack growth effectively and in real time to reduce the incidents of unexpected rail failure.

2.2 Laboratory investigation

Under laboratory conditions, bending samples made from all three steels were subjected to the same testing regime and conditions. All samples were machined to dimensions 120x10x20 mm and spark wire eroded using a 0.1 mm diameter wire, to form a V-notch at their midpoint to a depth of 2 mm as shown in Figure 2.8.

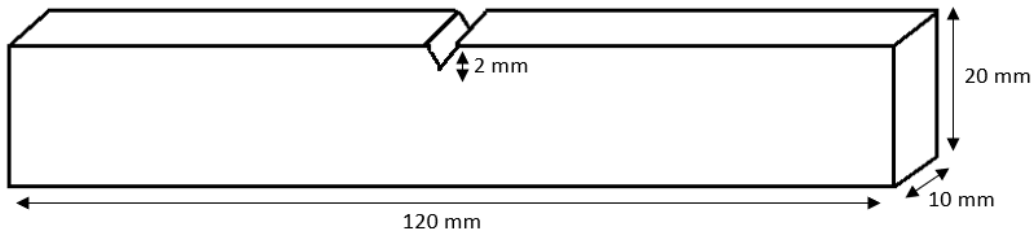


Figure 2.8 *Schematic representation of the sample geometry used for all steels under laboratory fatigue testing. Each sample has a geometry of 120x10x20 mm with a v-notch of depth 2 mm.*

2.2.1 Pre-cracking

To ensure the presence of crack to monitor all samples are pre-cracked using an Amsler 20 kN vibraphone. During this process a force of 0.85-8.5 kN was applied to the R220 and R260 steel samples, and a force of 0.65-6.5 kN was applied to the cast manganese samples both at a resonant frequency and under three-point bending conditions. The presence of an initiated crack is confirmed using replicas of the sample surface being observed at regular intervals. As soon as a crack has been detected this process is stopped.

Once complete, this initial crack length was measured optically from both sides of the sample with the final initial crack length taken as an average of these measurements. The initial crack length in this study is defined as the distance from the surface of the sample to the tip of the crack as observed from the surface as shown in Figure 2.9.

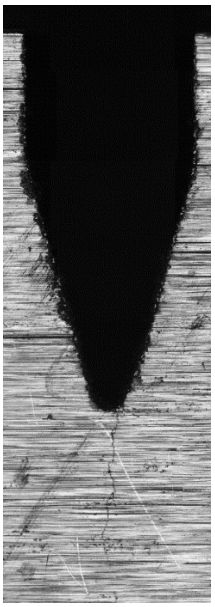


Figure 2.9 *Micrograph taken at a magnification of 100x showing the spark wire eroded V notch and the pre-crack. The initial crack length is taken as the distance from the surface of the sample to the pre-crack tip* (Culwick, 2019).

2.2.2 Fatigue testing

To simulate loading conditions experienced by in service rails samples were fatigue tested under three-point bending conditions. A force of 0.65-6.5 kN was applied under a loading frequency of 1 Hz for the R220 and R260 samples and a force of 0.45-4.5 kN at a loading frequency of 5 Hz for the Cast manganese steel samples, using a DARTEC 50 kN servo-hydraulic universal testing machine. Crack growth was monitored using Direct Current Potential drop (DCPD), with the actual crack length calculated using Johnson's Formula, Equation 2.1 (Johnson, 1965). All samples were taken to failure before being overloaded to expose the crack surface. The experimental set up is shown in Figure 2.10 complete with the AE monitoring system described in section 2.2.3.

$$a = \frac{2W}{\pi}\cos^{-1}\left(\frac{\cosh\left(\frac{\pi y}{2W}\right)}{\cosh\left(\left(\frac{V}{V_0}\right)\cosh^{-1}\left(\frac{\cosh\left(\frac{\pi y}{2W}\right)}{\cos\left(\frac{\pi a_0}{2W}\right)}\right)\right)}\right)$$

Equation 2.1 *Rearrangement of Johnson's formula (Johnson, 1965) giving the crack length a where; W is the width of the sample, y is the half sensor distance, V is the voltage, V_0 is the initial voltage and a_0 is the initial crack length.*

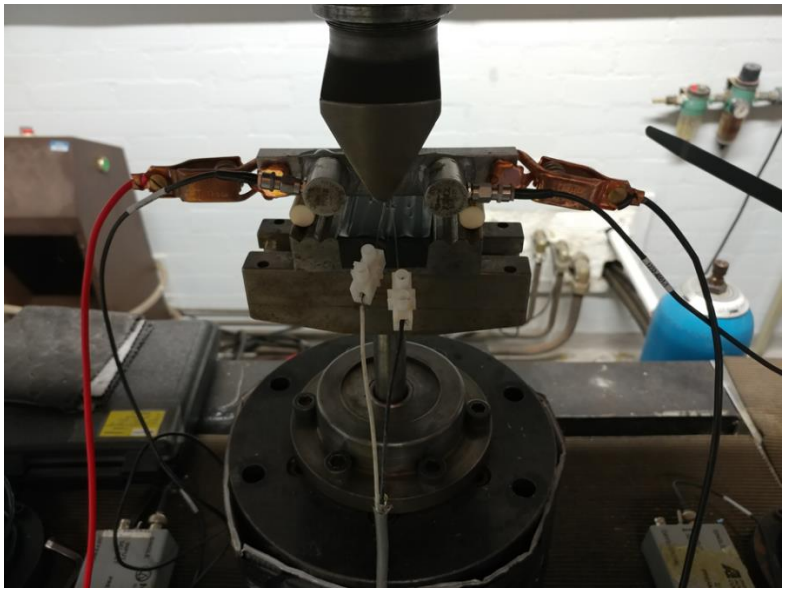


Figure 2.10 *Laboratory set up of fatigue test samples showing the arrangement of DCPD monitoring equipment and AE monitoring sensors.*

2.2.3 Acoustic emission monitoring

The AE signals generated during fatigue testing were detected and converted to an electrical signal by two R50α Piezoelectric sensors procured from Physical acoustics. These were surface-mounted and acoustically coupled to the sample using Araldite® ensuring no air bubbles were present between the sample surface and the sensors. The electrical signals generated by these sensors were passed through a pre-amp and amplifier set to 40 dB and 6dB respectively with a 28 V phantom voltage applied. This system was connected to the logging computer using an AGILANT 2541A USB DAQ. As already detailed, a custom and a commercial system are used to log the recorded AE signals. A set of universal filtering parameters was used for both systems and are detailed in Table 2.4.

Table 2.4 *AE parameters set for both the custom and commercial processing systems.*

Limiting Parameters	**Setting**
Hit definition time	1000 µs
Hit lockout time	2000 µs
Max duration of the signal	25 µs
Peak definition time	600 µs
Resonance Filter	100 kHz
Sampling rate	5 MS/s
Threshold	40 dB

The commercial system uses a hit driven data (HDD) approach recording three key AE parameters; counts, energy, and duration. The custom system uses Time driven data (TDD) approach capturing the complete waveform recorded after the parameters detailed in Table d, have been applied. The custom system acquisition is time limited to 5 s.

2.3 Field Studies

Field studies were used to collect data from a live rail environment to confirm the effectiveness of equipment and techniques developed during this project. Of particular interest are the levels of background noise recorded in the real-world environment, these are expected to be greater than that in the control laboratory-based experiments. Field trails took place at Bescot Yard depot.

2.3.1 Field trail equipment set up

The custom AE system used in the laboratory environment described in section 2.2 was used in the field trails. The R50α AE were attached to the web of the rail using Araldite® with the cables running between the sleepers to the pre-amplifies, amplifies, DAQ and finally the computer used to record the AE signals in the same set up as the laboratory environment described in section 2.2. The same settings were used as those described in Table 2.4. Figure 2.11 and Figure 2.12 show the field trial set up.

Figure 2.11 *Images of the AE acquisition equipment used during field trails. (a) R50α AE sensors attached to rail web (b) pre-amplifier in rail bed (c) Amplifiers (d) Recording computer.*

Figure 2.12 *Image of the AE sensors, pre-amplifiers, and amplifiers in situ.*

2.4 Summary

This chapter has covered the experimental set up of both the laboratory-based fatigue testing and the field trails completed at Bescot yard. Both a commercial AE system and a custom-built acquisition system were used in the Laboratory based experiments whilst only the custom AE system was used in the field trails.

This study will primarily focus on the data acquired from the laboratory-based experiments for both the parametric analysis, for the data from the commercial system, and machine leaning analysis of custom system data. Field trails were used to test the custom-built system under real world conditions.

Chapter 3

Commercial system: Initial comparison of key AE parameters

Commercially available systems for AE monitoring in general use a Hit Driven Data (HDD) approach for recording the received AE signals. The system procured from PAC (now MISTRAS) and used in this study is thus, no exception. Whilst it is possible to record a range of acoustic-related parameters using this approach, the most widely employed are typically AE signal amplitude, energy, duration, and counts (Roberts & Talebzadeh, 2003). It is these parameters which will be explored in this chapter.

As already discussed, the aim of the present study has been to explore the development of a real-time structural health monitoring system using AE monitoring techniques, focussing particularly on the quantitative evaluation of fatigue crack growth in critical railway components, such as rails and cast manganese crossings. To explore the initial feasibility of using AE monitoring in this way a relatively simplistic and computationally cheap approach was initially used. This initial approach compared the growth per fatigue cycle and each of the three chosen parameters per cycle against the stress intensity factor (ΔK). If one of these parameters is shown to follow a similar trend to the crack growth rate it may be possible to rapidly draw a comparison. It is worth noting that analysis of this data to quantitatively monitor the structural health of a material is complex.

For both the R220 and R260 steels, it is expected that they will follow the Paris-Erdogan law during fatigue crack growth . Therefore, the confirmation of a close correlation between these

parameters and the Paris-Erdogan law plot for these materials would allow the development of a fatigue model based on the Paris law plot which could be updated in real time using the AE data recorded. For this model to be effective it is vital that the AE data very closely follow the crack growth trend. This has been achieved previously by Hamel et al. who developed equations to link AE with Paris law in 1981 (Hamel et al., 1981). Even earlier than this Morton et al. quantitatively correlated AE with the crack growth rate and stress intensity factor (Morton et al., 1973). In the study carried out by Bassim et al. in 1984, it was possible to create an automated system for the calculation of ΔK from the AE using the Paris law (Bassim et al., 1994). Dunegan et al. found that the AE counts measured in tensile fatigue tests carried out on aluminium and beryllium samples were proportional to the fourth power of the stress intensity factor obtained from a sharp crack (Dunegan et al., 1968). They did however note that this is dependent on the geometry of the sample. This observation gets to the heart of the limitation of trying to apply generalized models to all materials and applications as it is also well understood that different materials will have a different response. Therefore, equations and trends seen in one material or application will not necessarily apply to another. For example, the correlation of AE data to crack growth for cast manganese steels is more complex than that of the R220 and R260 steels due to the strong effect of plasticity at the tip of the crack in the generation of AE signals during crack propagation. As shown in the work by Kang et al. these steels do not display typical fatigue crack growth behaviour and therefore, will not follow a typical Paris law plot making a comparative modelling approach more challenging. Nonetheless, this issue does not necessarily render this approach redundant. Instead, this approach may act as a feasibility study or as part of a wider multi-parameter approach.

3.1 Commercial system data

In the following data, the crack growth and AE parameters are grouped in data points of the average of that parameter per cycle between each ΔK value recorded. Whilst this allows for overall trends to be observed, much of the detail that is potentially retrievable from the AE data will be lost.

Figures 3.2-3.11 show the AE energy, duration and counts responses for each steel. The crack growth behaviour for both the R220 and R260 steels (Figures 3.2, 3.3, 3.6, 3.7, 3.9, 3.10) follow the expected Paris law trend, allowing for a comparison to be made between the AE data and the Paris Law. It is noted that some samples show some deviation from this plot, with spikes in crack growth. This indicates that the crack growth behaviour is not stable. The crack growth behaviour for the cast manganese steel samples (Figures 3.4, 3.8, 3.11) also in general follows the trend of increasing crack growth with ΔK as it would be expected. However, a greater level of variation in crack growth is observed when compared with the R220 and R260 steels. This is made clearer where samples are plotted individually as in Figure 3.1. Herewith, it can be observed that the R220 and R260 follow a relatively linear crack growth compared with that for cast manganese samples.

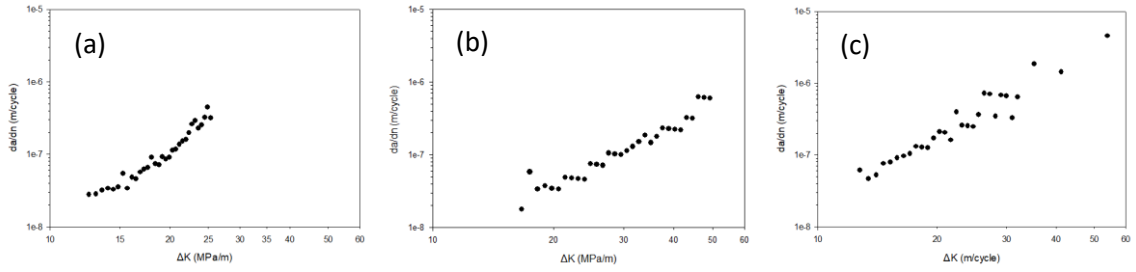


Figure 3.1 *Plots of the crack growth per cycle (da/dn) against ΔK for the R220 (a), R260 (b) and cast manganese steels (c). Each plot represents a single sample as opposed the grouping of samples shown in Figures 3.2-3.11. By showing individually the greater variation in crack growth displayed by the cast manganese can be observed compared to the more linear behaviour shown by the R220 and R260 samples. These Plots also show the R220 steel failing at a stress intensity factor significantly lower than that of the R260 and cast manganese steels.*

3.1.1 AE Energy

The acoustic emission energy is the enveloped area under the AE signal as shown in Figure 1.12. As the ΔK increases during fatigue crack growth the crack becomes more unstable. This leads to an increase in the crack growth per cycle. As the AE signals detected are the elastic stress waves generated as the result of the strain energy released from the crack tip as the fatigue crack grows. It follows that an increased crack growth as the crack growth accelerates, and the cracks gets nearer to the point of instability will lead to a greater release of AE energy. It could therefore be expected that the AE energy plots will follow a smooth close correlation with the crack growth rate. This however is not expected to be the case, as discussed in Chapter 1 the microstructure of these steels is not homogenous with defects and inclusions present which will affect the crack growth rate locally and therefore, the energy released. It is hence, expected that for the R220 and R260 steels the general trend will be that as the ΔK increases so will the energy however there will be variation in this energy. The cast manganese steel is expected to show greater energy variation with crack growth due to its fatigue crack growth behaviour.

3.1.1.1 R220 AE Energy

In Figure 3.2 it is observed that as the crack growth per cycle increase with increasing ΔK the Energy per cycle also shows an increasing trend with most of the samples following to the approximated trend line (black line). The similarity in the AE energy and da/dn plot is expected with an increasing stress intensity factor with increasing crack growth leading to an increase in the energy released. The obvious exception to this is sample 9 which sits as an outlier. Ideally for the creation of a monitoring based exclusively on this comparison all samples would fit the general trend observed. In reality, however, this would not be likely to occur, with many factors effecting the energy released and recorded.

As discussed in Chapter 2 R220 steel contains defects such as Iron carbide and MnS inclusions. These inclusions will affect the local crack growth behaviour leading to a change in the energy

released. It is well understood that changes in material lead to different AE responses (Yu et al., 2011). Thus, it follows that the presence of defects within a material will also have an effect on the AE behaviour. These changes in crack growth behaviour are not observed in the crack growth plot. This hints towards the sensitivity of AE monitoring and its potential use to give a highly accurate in depth understanding of real time crack growth behaviour.

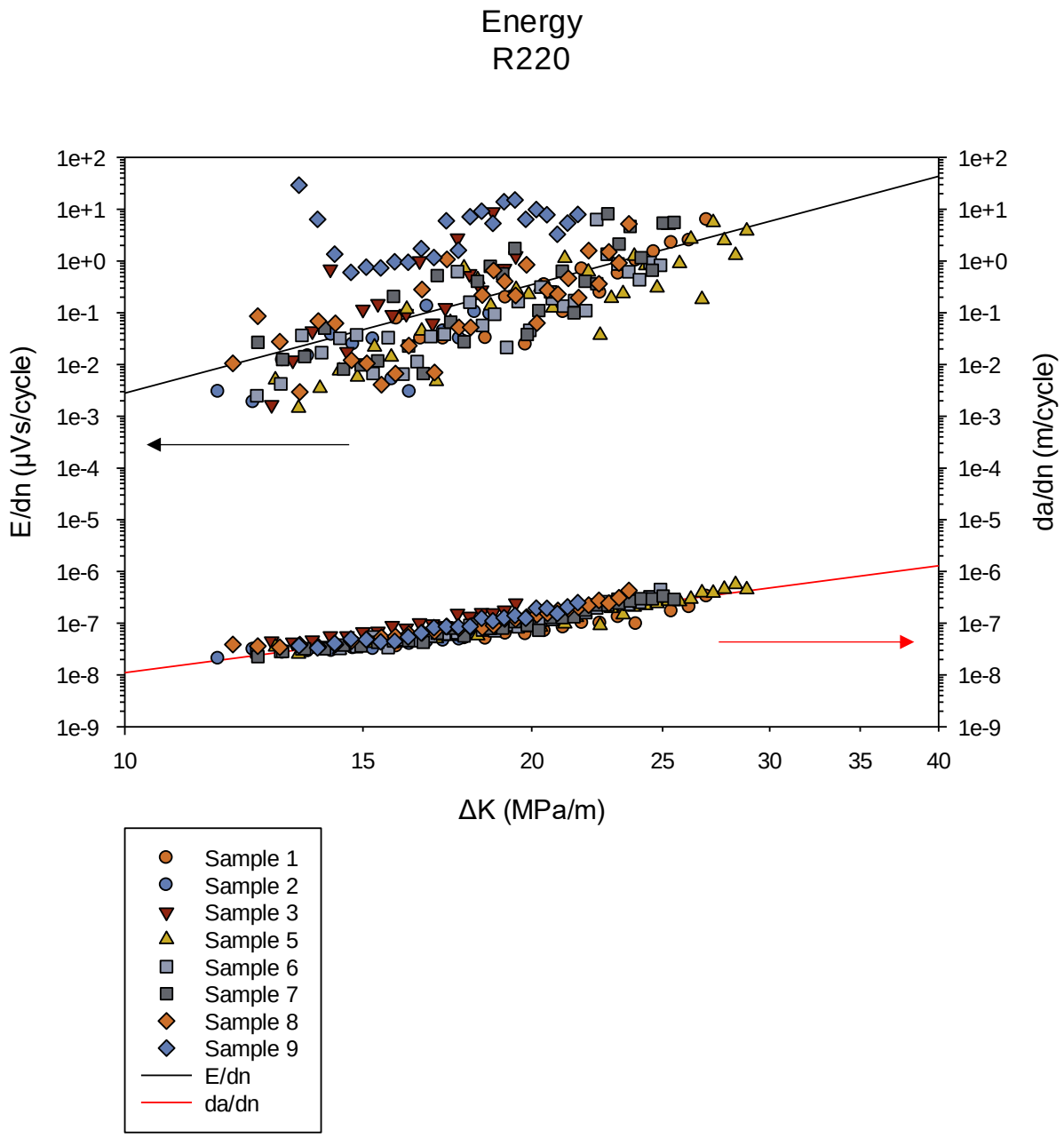


Figure 3.2 *AE energy per cycle (E/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the R220 steel samples. The trend lines for E/dn and da/dn are shown as black and red lines respectively. The red and black arrows indicate the axis to which the data is plotted with the E/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

Considered on its own, it is not possible to determine whether this increase in AE energy per cycle is due to an increase in the number of AE signals per cycle, or a similar number of signals but with a greater energy. Hence, other AE parameters must be considered in conjunction with the AE energy for this material to gain a complete picture of the crack behaviour and structural health of the component being monitored.

3.1.1.2 R260 AE Energy

The AE signals generated during crack growth are expected to be different for each material, even in the case of R220 and R260 which are very similar in composition and somewhat limited variations in the mechanical properties are observed. Figure 3.3 shows the AE energy recorded for the R260 steel. As with the AE energy response observed for the R220 steel in Figure 3.2 the da/dn plot follows the expected increase with increasing ΔK. The trend is however, not so clear with the E/dn plot. in Figure 3.3 a trend line has been added to fit all the AE energy ample results. This trend line does follow a similar trend to that shown on the da/dn but it is clear that the data is well spread around this trend line. Any model based solely on this data would thus lead to a high level of error.

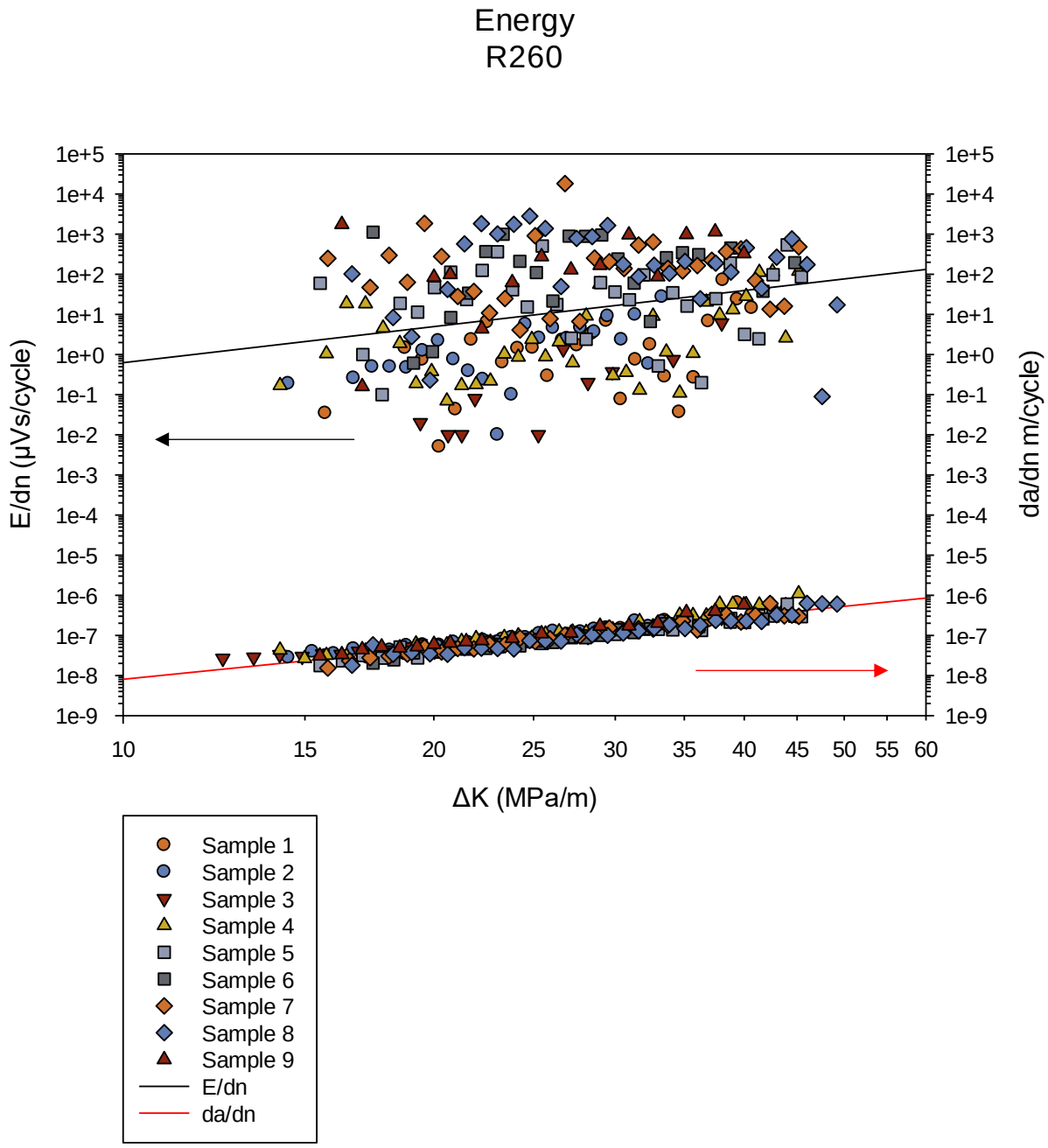


Figure 3.3 *AE energy per cycle (E/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the R260 steel samples. The trend lines for E/dn and da/dn are shown as black and red lines respectively. The red and black arrows indicate the axis to which the data is plotted with the E/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

Whilst on its own, this approach is not useful in the construction of a structural health monitoring model, it can be argued that there is an overall trend towards increasing AE energy with increasing ΔK although with relatively low confidence of individual results following this trend. This overall upward trend in AE data does however, pave the way to look at other AE

data approaches to develop a system for monitoring crack growth, these approaches will be explored in greater detail in later chapters of this thesis. In principle they are able to create a more accurate model by exploiting the wealth of data that is recorded by AE monitoring and does not lose this detail through the grouping of the data which has been used in this case.

3.1.1.3 Cast Manganese steel AE Energy

In Figures 3.2 and 3.3 a trend line was added to the AE energy data with Figure 3.2 for the R220 steel showing a greater correlation than the R260 data shown in Figure 3.3 which presents a much greater spread. No trend line has been added to the AE energy in Figure 3.4 which shows the AE energy recorded for the cast manganese steel samples. It can be clearly seen in Figure 3.4 that the AE energy data does not show a clear trend to which it would be appropriate to add a trend line to. The da/dn data does nonetheless, follow a trend in the same way as that shown in Figures 3.2 and 3.3.

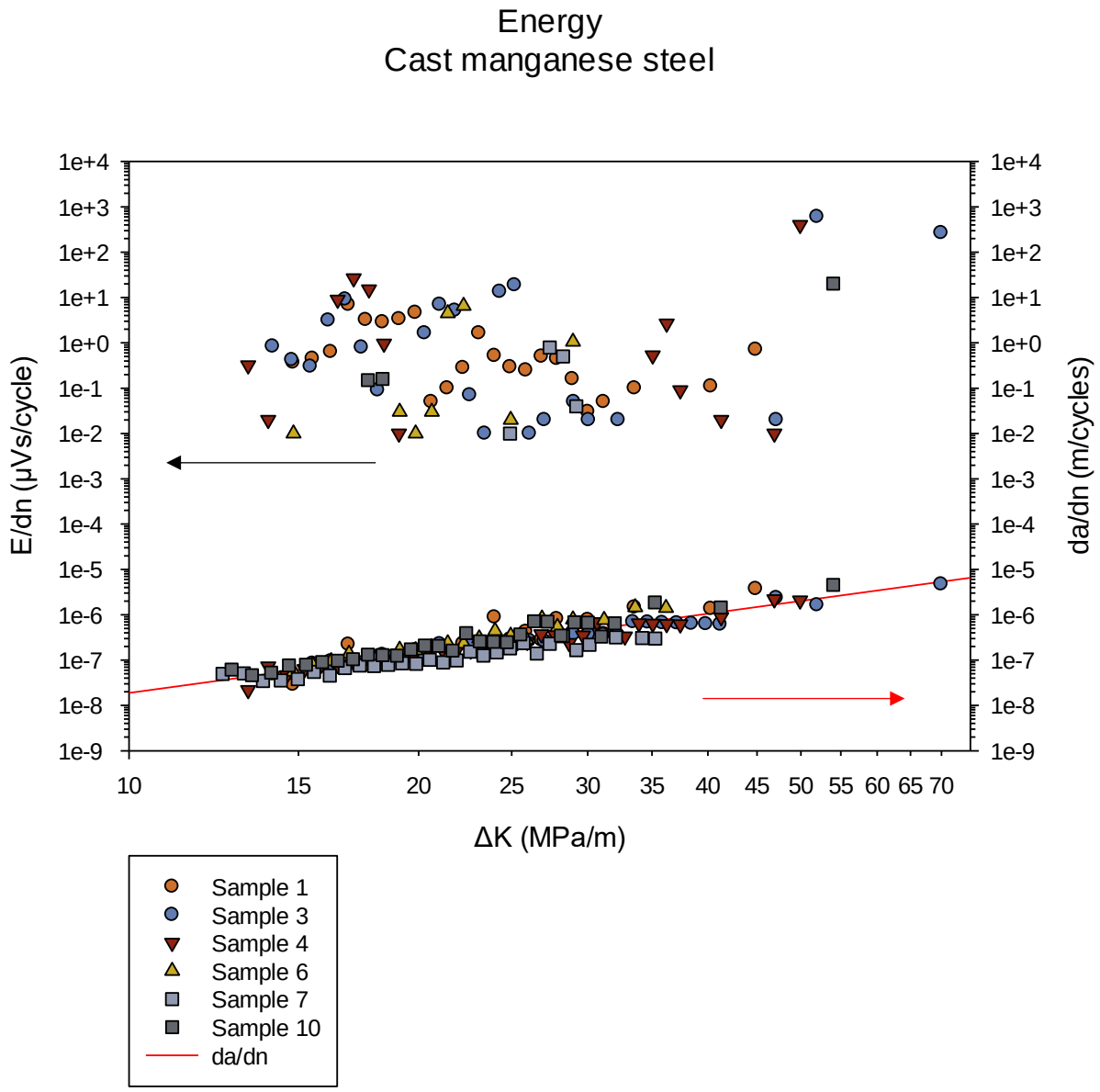


Figure 3.4 *AE energy per cycle (E/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the cast manganese steel samples. A red trend line is shown for the crack growth per cycle however, due to the clear spread in data no trend line is added to the AE energy response. The red and black arrows indicate the axis to which the data is plotted with the E/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

As discussed earlier, previous work has found that cast manganese steel do not follow the normal fatigue crack growth behaviour displayed by other steels. Instead, they initially become work-hardened before finally softening to failure. Therefore, if the AE energy is indicative of the crack growth behaviour it would be expected that the AE energy recorded would not show

the same behaviour as that of the R220 and R260 steel. Looking at Figure 3.4 this is certainly the case, there is a large spread of AE energy data throughout the crack growth cycle.

Looking at the AE data grouped as a whole, as in Figure 3.4, the establishment of a pattern or trend in the data is not simplistic. If each sample is plotted separately a pattern does start to emerge. This can be observed in Figure 3.5. Towards the start of the fatigue crack growth cycle there is a high level of AE energy activity recorded this then drops to only a small before rising again before final failure. This trough like behaviour of the AE energy has been in other studies such as the work by Kongpuang et al. and Kang et al. (Kang et al., 2014; Kongpuang et al., 2021). A single sample was shown here in Figure Y as a representative sample with a similar behaviour observed across all the samples in this dataset.

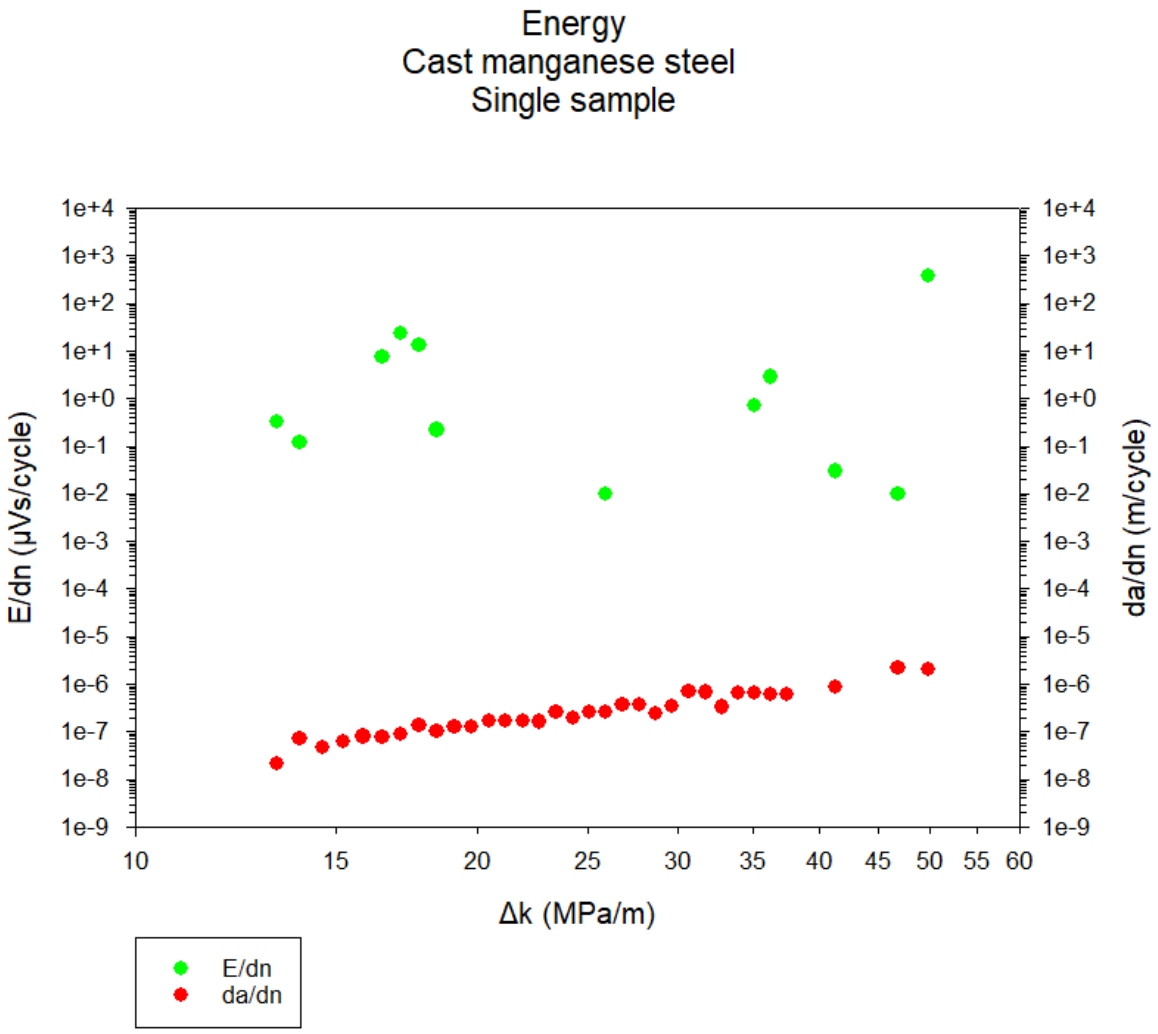


Figure 3.5 *A plot of the AE energy per cycle (E/dn) and crack growth per cycle (da/dn) against the stress intensity factor (ΔK) for a single sample of cast manganese steel. Plotted as a single sample it is possible to observe the grouping of the AE energy at the start and end of the fatigue crack growth period monitored. This has previously been described as a trough like behaviour and attributed to the work hardening and softening behaviour displayed by cast manganese steel during fatigue loading.*

The trough-like nature of the AE energy observed in the cast manganese steel means that a direct comparison with the stress intensity factor is not possible with a more complex analysis approach required. Used in isolation this data could be used to develop a system that is not able to always provide the current structural health of the steel but could instead provide an early warning system of a crack being present and a warning prior to final failure. This system would not perform adequately satisfactorily and would be susceptible to loss of or missing signals,

since without continuous monitoring it would not be possible to distinguish between the initial signals generated in the early stages of crack growth and the subsequent AE energy data points prior to final failure.

3.1.1.4 AE Energy Discussion

As it has been discussed, ideally for a simple comparison to be made between an AE parameter and the Paris law plot the AE parameter of choice should follow the same trend as the Paris Law plot. This has not been the case for all the steels investigated here and shown in Figures 3.2, 3.3 and 3.4. The AE energy plot for the R220 steel (Figure 3.2) followed the da/dn plot more closely with most of the data falling along a similar trend to the da/dn plot. There are outliers to this trend, but this is to be expected, since the microstructure of these steels is not homogenous with defects known to be present. Any variation from the base microstructure will lead to variation in the AE energy released.

Even more variation from a clear trend in the AE energy data is observed for the R260 steel (Figure 3.3) with a large spread of data, although a trend line fitted to the entire data set for all samples does follow a similar trend to the da/dn plot. As with the R220 data, some variation is expected. However, the greater spread of the AE data in this case would make a direct comparison model difficult to develop.

The cast manganese steel was not expected to show exactly the same trend as that of the R220 and R260 steels and this is certainly the case with now a clear trend shown in Figure 3.4. Nonetheless, as discussed earlier when the data is looked at a sample level (Figure 3.5) a trough like pattern emerges. It is suggested that this could be used to develop a system that provides a warning of failure rather than a real time structural health monitoring system. A system of this sort may be useful to the rail industry and the monitoring of crack growth in cast manganese steels. It is a particular challenge, however, to convince the industry that a system is viable to

implement since it would likely need to be more reliable and accurate than a system based on this approach alone could achieve.

It is clear from looking at the AE energy for these three steels that more information is needed to develop a system that can monitor in real time their structural health under fatigue loading. Used in isolation and with the data grouped as it is in this case it is not possible to tell the nature of the AE signals. That is whether for example there are a small number of high energy signals or many lower energy peaks. These scenarios would indicate very different crack growth behaviour which cannot be identified by the analysis of the plots shown in Figures 3.2, 3.3 and 3.4. Therefore, it is important to look at other AE parameters which will be able to differentiate between these scenarios.

3.1.2 AE Duration

The AE duration is the length of time from a detected signal first passing any threshold value set to the detect signal to it dropping back below this threshold value as shown in Figure 1.12. The duration gives an indication on the nature of the AE signal detected in terms of whether they are short burst signals or longer period signals. This may give an indication of the nature of a crack growth event as well as the crack severity.

When the AE duration is used in conjunction with the AE energy, a more detailed picture of the nature of the received AE signals can be determined. For example, in the case of AE energy the value recorded gives an indication of the overall energy released but not the source of the energy released. A high energy value could be due to short-duration high amplitude signal or due to a more long-duration but lower amplitude signal. By using these parameters in conjunction, a more detailed picture of the crack growth mechanisms may be gained. The energy also gives a greater understanding of the signal compared to just the duration being used as the duration gives little indication of the amplitude of the AE signal, it simply gives the length of time over which the signals are above a threshold.

3.1.2.1 R220 AE Duration

The comparison of the AE energy per cycle, Figure 3.2, showed an overall increase in the energy per cycle with increasing ΔK. It was however noted that there was a spread in this data making a direct comparison between the AE response and the current crack growth. Figure 3.6 shows for the same set of samples the change in duration per cycle (D/dn) compared to the crack growth per cycle (da/dn) as in Figure 3.2. As with the E/dn plot increasing ΔK leads to an increase in the duration of the AE signal per fatigue sample. Unlike Figure 3.2, Figure 3.6 shows significantly less spread in the AE data around the plotted trend line. This stronger fit to the ΔK vs da/dn plot may indicate that the AE duration could more effectively be used to create a monitoring system for fatigue growth by simply comparing the duration of the AE signals received with a dataset which includes the ΔK vs da/dn plot giving a value for the structural component being monitored.

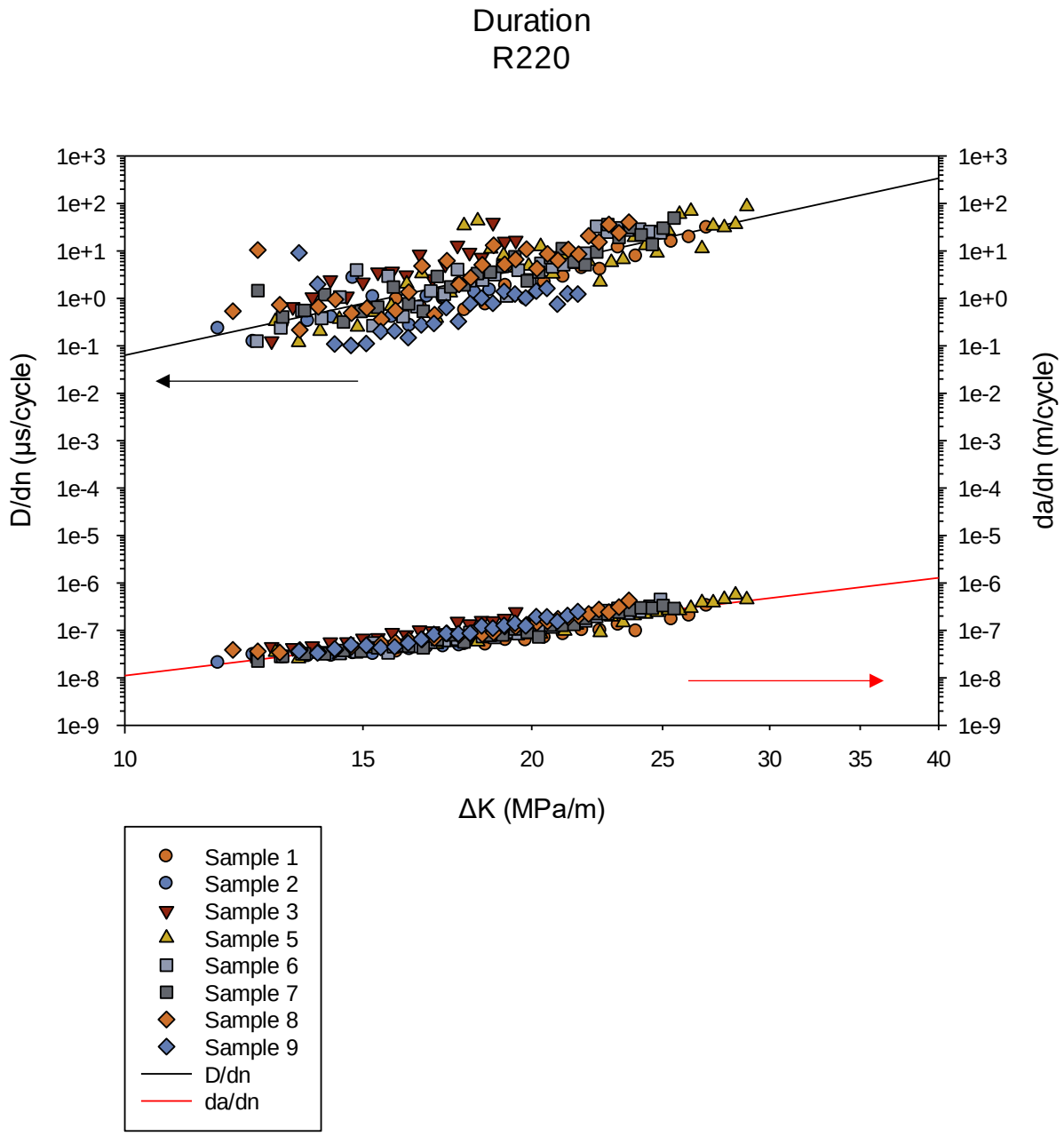


Figure 3.6 *AE duration per cycle (D/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the R220 steel samples. The trend lines for D/dn and da/dn are shown as black and red lines respectively. The red and black arrows indicate the axis to which the data is plotted with the D/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

The trend of increasing AE duration with increasing ΔK is not unexpected. As the length by which the crack grows increases, the time taken for this to occur would also increase unless the nature of the crack growth was changing. For example, rapid brittle fracture may lead to a sudden increase in crack with only a short-duration burst AE signal.

3.1.2.2 R260 AE Duration

As with the change in energy per cycle shown in Figure 3.3 for the R260 steel the plot showing the change in duration per cycle, Figure 3.7, the overall trend in the AE data displays an upward trend similar to that of the change in crack growth per cycle. Figure 3.7 also shows the same spread in this data as in Figure 3.3. The spread in E/dn data in Figure 3.3 is uniform across the plot, in contrast the D/dn plot in Figure 3.7 which initially shows a wider spread in data for all sample plots which then coalesce to a more uniform evolution in the late stages of crack growth. The overall upward trend in D/dn which follows that of the overall increase in E/dn in Figure 3.3 implies that the increase in energy is due at least in part to an increase in duration of the signal rather than an increase in amplitude of the AE signal.

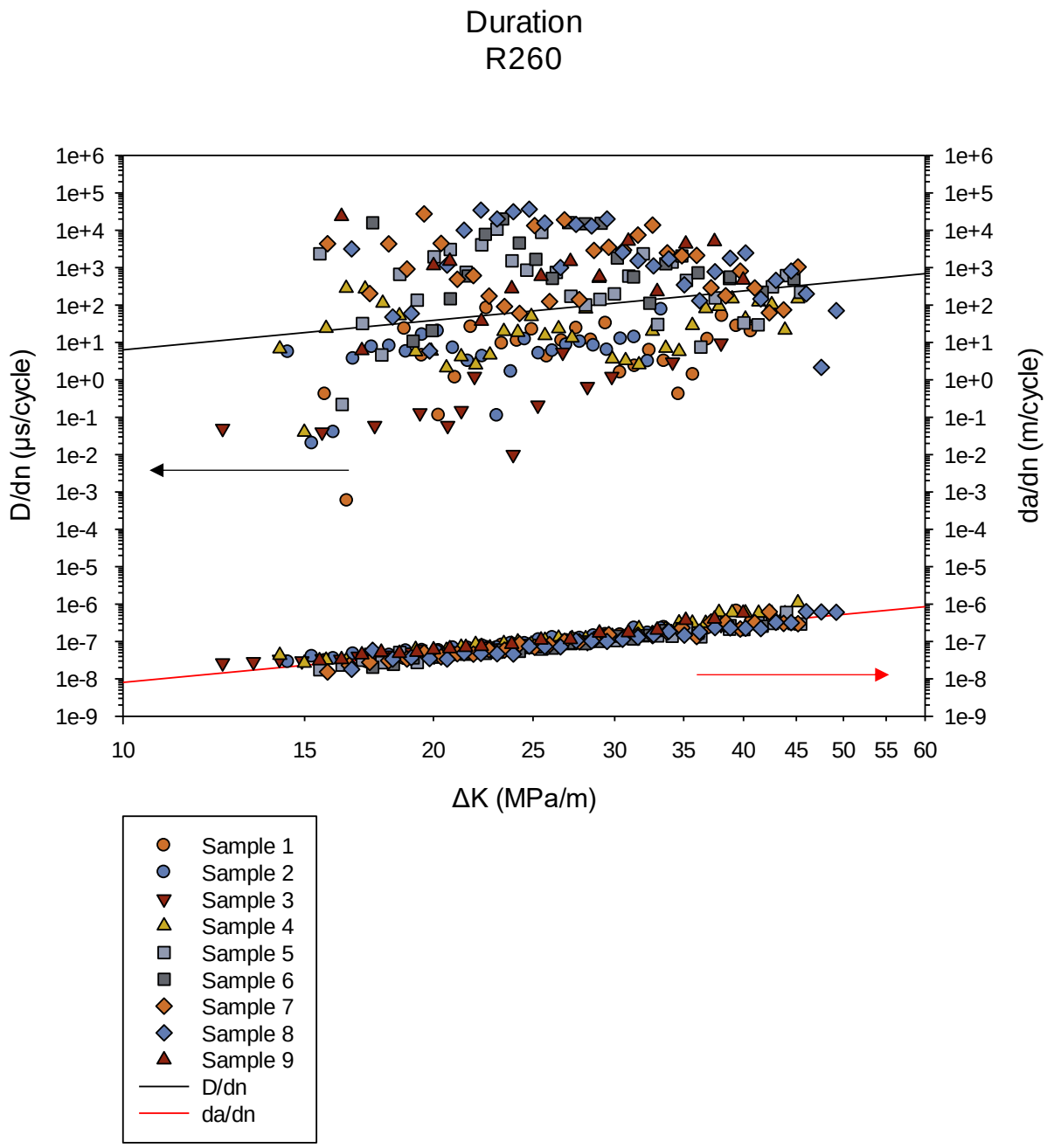


Figure 3.7 *AE duration per cycle (D/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the R260 steel samples. The trend lines for D/dn and da/dn are shown as black and red lines respectively. The red and black arrows indicate the axis to which the data is plotted with the D/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

As with the E/dn plot for this steel the spread in data in the D/dn plot limits the ability to create an effective model from this data. However, the trend seen in the case of Figure 3.7 may offer an insight into the crack growth behaviour. In the early stages of crack growth, the wide spread

of data is a challenge to interpret. It is possible this wide spread of data is due to the presence of defects within the microstructure leading to variation in the AE response. If this is the case, then another factor must become dominant in later stages of crack growth that lead to a consolidation of the data. One possibility of this is that in the early stages of crack growth, when the stress intensity factor is low the presence of defects has a greater effect on the crack growth behaviour and hence, the AE response, than when the crack becomes more unstable in the later stage of the fatigue crack growth cycle. Whilst this would render the effective modelling of current structural health challenging if using just this approach, it does indicate that AE monitoring may be able to provide an insight into the current crack growth behaviour and the presence of defects in the crack growth path. This could be built into a multi-factorial modelling system that is able to not only monitor the current crack growth rate or length but also the current crack growth behaviour and the presence of defects that may affect the fatigue life of the asset.

3.1.2.3 Cast Manganese steel AE Duration

Both R220 and R260 showed a change in the AE response when the plotted parameter was changed from E/dn to D/dn. Beyond a slight shifting in value the D/dn plot for cast manganese steel shown in Figure 3.8 largely follows the same pattern shown in the E/dn plot (Figure Q). That is displaying a trough like pattern in the AE plot. This implies that any model developed using the E/dn plot could also be adapted to use D/dn plot.

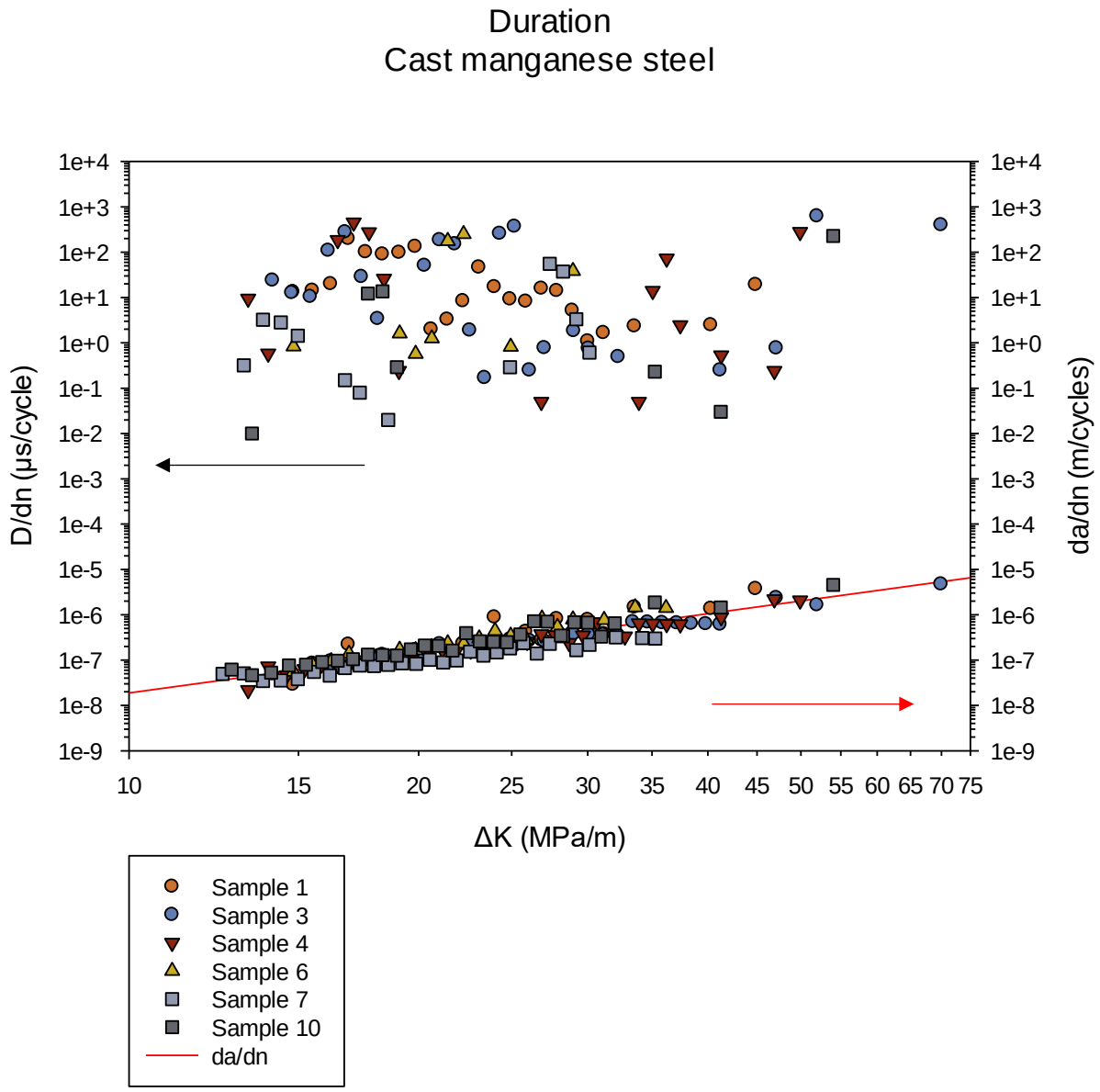


Figure 3.8 *AE duration per cycle (D/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the cast manganese steel samples. A red trend line is shown for the crack growth per cycle however, due to the clear spread in data no trend line is added to the AE duration response. The red and black arrows indicate the axis to which the data is plotted with the D/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

As with the plots for R220 and R260 the similar pattern shown by the D/dn plot compared with the E/dn plot implies that the increase in energy is at least in part due to the increase in duration of the signal rather than the presence of higher amplitude signals.

3.1.2.4 AE Duration Discussion

Both the R220 and R260 steels showed less variation in the AE duration data in comparison with the AE energy, which may potentially make it a more useful AE parameter in the development of a monitoring system. In contrast the cast manganese steel showed very little variation between the E/dn plot and the D/dn plot implying that both parameters would be of similar use.

The similar increase in AE duration in comparison with the AE energy in both R220 and R260 steels implies that the increase in the AE energy is at least partially attributable to increasing signal duration rather than an increase in amplitude of the signal. Whilst showing a different pattern in the AE response to that of R220 and R260 steels, the cast manganese steel does show a similar trend in that the E/dn plot and D/dn plot follow the same trend. As with the other steels this would imply that for the cast manganese steel the changes in AE energy are, at least in part, due to the increase in duration of the signal.

3.1.3 AE Counts

The AE counts are the number of peaks recorded from when the set threshold value is exceeded to the signal dropping below the threshold value as shown in Figure 1.12. The number of counts may indicate a number of crack growth scenarios. A small number of counts may indicate a short burst signal for example a single high amplitude peak, however, it could alternatively indicate a very low amplitude signal where only a limited number of the peaks are over the threshold value. This and the wide variety of circumstances which could lead to changes in the number of AE counts could make the use of AE counts challenging as an indicative parameter of the current structural health.

Through the use of correctly selected threshold values the number of different scenarios leading to the same AE counts can be restricted. With these thresholds in place, it is expected that as the crack grows and becomes more unstable, and therefore more rapid crack growth, the

number of AE counts recorded per hit will increase. This trend is expected because as the crack growth per cycle increases the strain energy released will increase. Assuming that the threshold parameters are correctly selected to exclude background noise but maintain the AE signal corresponding the fatigue crack growth, it follows that this would lead to a greater number of peaks detected. It is clear however that even with correct threshold values set there is lightly to be a lot of variation in this AE parameter. This variation may in itself prove useful in the classification of the crack growth behaviour, particularly when used in conjunction with other parameters and monitoring methods which may be able to detect the crack length but not provide detailed information on the crack growth behaviour which may be determined from the AE counts.

3.1.3.1 R220 AE Counts

Continuing the trend observed in both the AE energy and duration (Figures 3.2 and 3.6 respectively) for the R220 steel the AE counts, shown in Figure 3.9, follow a clear increase in counts per cycle (c/dn) with increasing ΔK. In particular, the c/dn plot in Figure 3.9 follows a very similar trend to the D/dn plot in Figure 3.6, that is a clear upwards trend with very little spread in contrast to the E/dn plot in Figure 3.2 where outliers in the results were clearly observed.

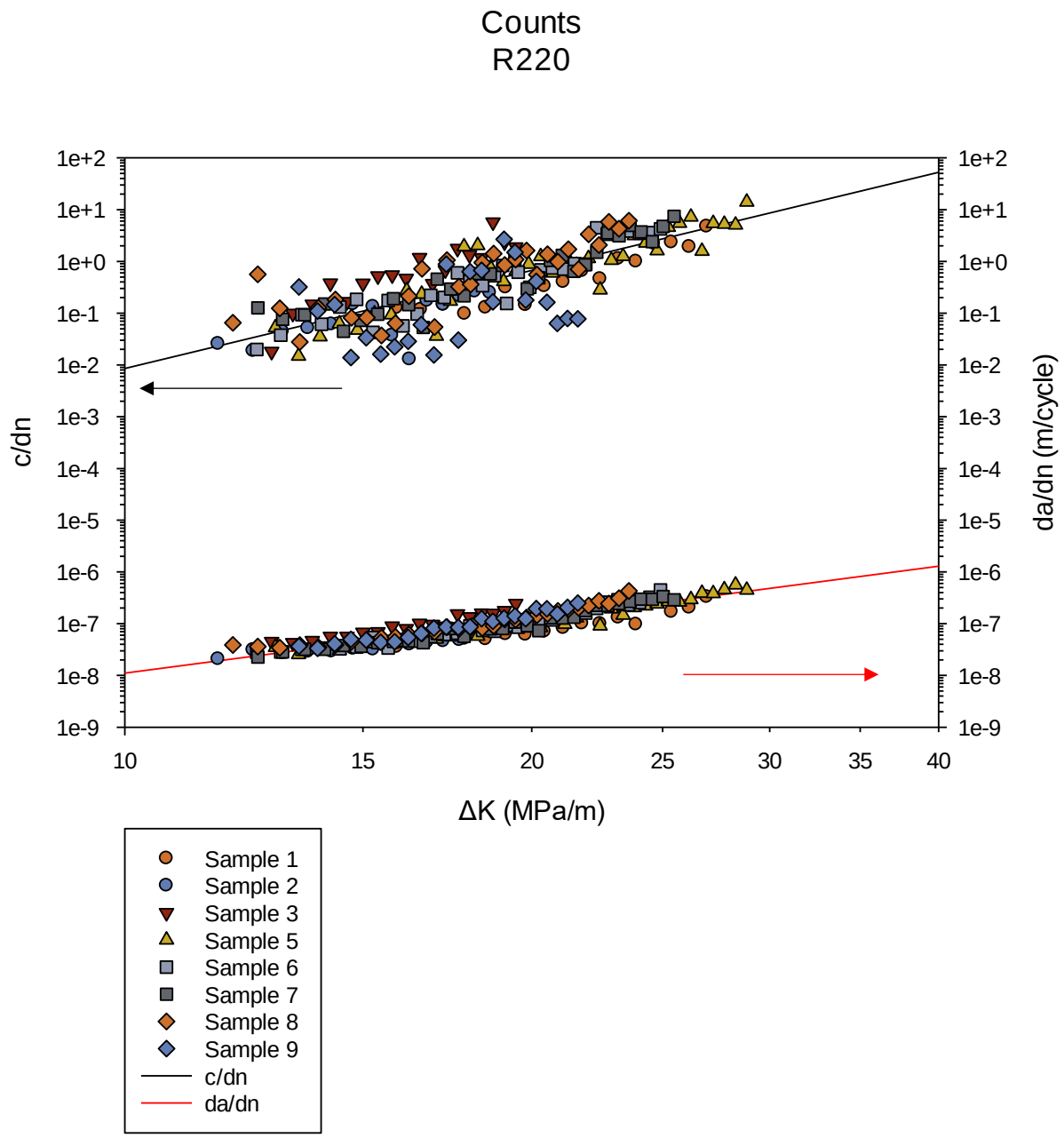


Figure 3.9 *AE counts per cycle (c/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the R220 steel samples. The trend lines for c/dn and da/dn are shown as black and red lines respectively. The red and black arrows indicate the axis to which the data is plotted with the c/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

The similarity between the AE counts and duration is expected as a greater time length (duration) over which the AE hit is recorded the more AE peaks are likely to occur. Any variation in the AE counts in comparison to the AE duration would imply changes in the frequency of the AE signal recorded. The fact that very little difference is seen in these datasets

is expected as the data are filtered by a set of frequency bands which limits the frequency that will be recorded as peaks and hence, the AE counts. If no filtering was applied, it would be expected that there would be more variation between the results obtained. The nature of this data analysis approach averaging out the results at each point to give a counts per cycle single data point means that detail of local changes in frequency will be lost.

3.1.3.2 R260 AE Counts

The AE counts for the R260 steel samples, shown in Figure 3.10, show a similar result in comparison to the AE energy and duration as displayed by the R220 steel samples. Like the AE duration plot in Figure 3.7 the c/dn plot in Figure 3.10 shows a relatively wider spread at the start of the data set before coalescing at higher values of ΔK. In the case of the AE duration data this was suggested to be due to the presence of defects in the microstructure, if this was the case then the presence of defects would have a similar effect on the AE counts. As with the other AE parameters studied here this wide variation in data would make the development of a model using this approach challenging.

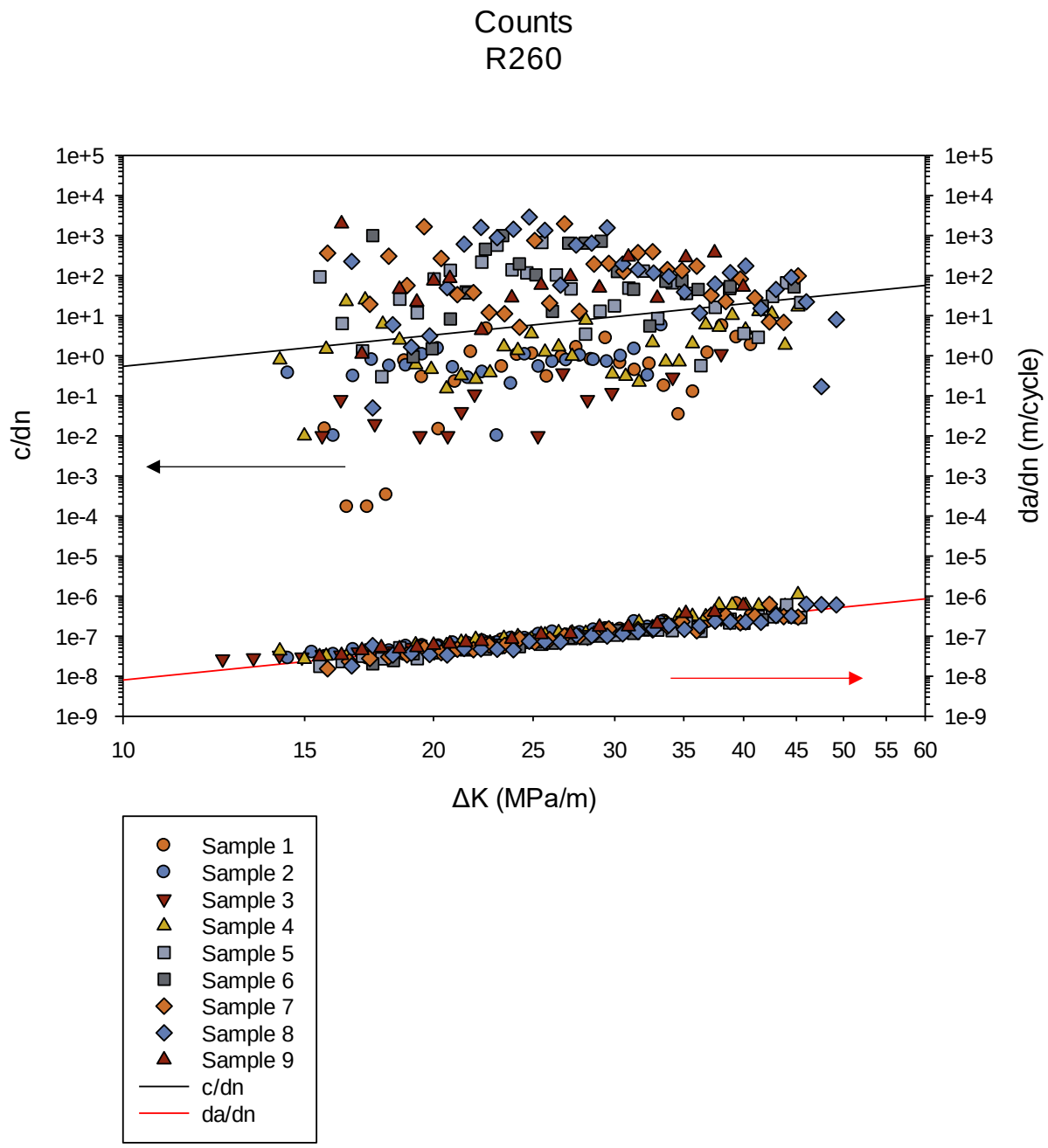


Figure 3.10 *AE counts per cycle (c/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the R260 steel samples. The trend lines for c/dn and da/dn are shown as black and red lines respectively. The red and black arrows indicate the axis to which the data is plotted with the c/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

3.1.3.3 Cast Manganese steel AE Counts

Following the same pattern observed for both R220 and R260 steels the cast manganese steel shows a similar pattern in the AE counts plot (Figure 3.11) compared to the AE duration plot (Figure 3.8). As discussed with the R220 and R260 steels it is expected that the AE counts does follow the AE duration data. As with the other AE parameters studied here the AE counts for the cast manganese steel does not follow a similar trend to the da/dn plot, instead forming a trough shaped plot.

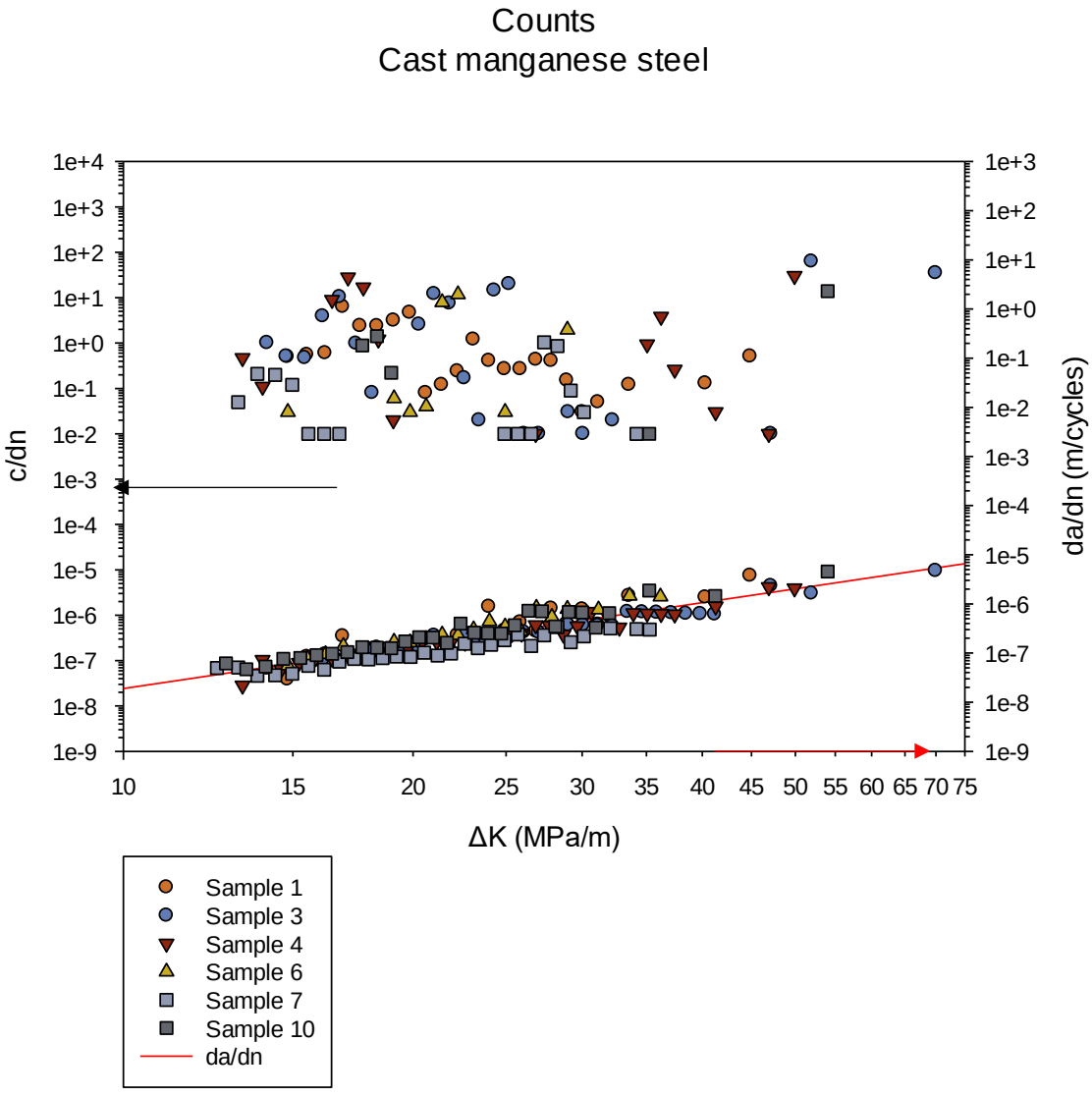


Figure 3.11 *AE counts per cycle (c/dn) and the crack growth per cycle (da/dn) plotted against ΔK for the cast manganese steel samples. A red trend line is shown for the crack growth per cycle however, due to the clear spread in data no trend line is added to the AE count response. The red and black arrows indicate the axis to which the data is plotted with the c/dn data being the data clustered at the top of the plot and the da/dn the data at the bottom.*

3.1.3.4 AE Counts Discussion

For all three steels investigated here the AE counts plots (Figures 3.9, 3.10, 3.11) follow very closely the trend of the plots of the AE duration (Figures 3.6, 3.7, 3.8). The close relationship of the trend between the AE counts and AE duration plots is not unexpected. Frequency filtering is applied to all of the datasets, thus, restricting the number of different frequencies recorded to a finite set. Due to this filtering combined with the averaging of the data, which is carried out to create individual data points in these plots, very little variation of the frequencies recorded is expected. The uniformity in the frequency of the AE signals recorded means that if the duration increases more peaks and AE counts will be recorded.

The AE counts for the R220 steels does show a similar trend to that shown by the da/dn plot, compared to the greater spread and less clear trends displayed by the R260 and cast manganese steels. This close trend displayed by the R220 steels may allow for a model to be developed using this AE parameter through simple comparison with the da/dn plot. A simple model of this nature would not be possible with the R260 and cast manganese steels and they do not have such a strong trend in the AE counts data.

3.2 Conclusions

The cast manganese steel shows a similar pattern in data across all the AE parameters investigated herewith (Figures 3.4, 3.8, 3.11) with all displaying a trough like pattern, i.e. that is an initial rise in the AE parameters followed by a drop before rising to final failure. This trough-like behaviour is likely due to the properties of this steel under fatigue loading, initially work hardening before softening to failure (Kang et al., 2014). In terms of the development of a system based on the comparison between the AE signal received and the Δk vs da/dn plot this trough-like behaviour is a problem with no easy comparison possible. A system could be developed that provides an alarm system of failure, initially indicating the presence of a crack from the initial rise in AE signals and then a final failure alarm when the data starts to rise again. This approach would, however, be highly susceptible to failure as it would be very dependent on no loss in data during monitoring. If there was a drop in data it would not be possible using only this approach to tell the difference between a rise in data due to the start of a fatigue crack or a rise due to the approach of failure.

In contrast to the cast manganese steel the R220 and R260 steels showed an upward trend in all AE parameters with increasing ΔK (Figures 3.2, 3.6, 3.9 and 3.3, 3.7, 3.10 respectively). This is particularly true in the case of R220 steels were the AE data closely follows the trend of the ΔK vs da/dn plot, although the AE energy plot (Figure 3.2) does show some variation in the data. Whilst an overall trend is observed in all the R260 steel AE parameters there is a much larger spread in data making a direct comparison with the ΔK vs da/dn plot challenging.

One of the limiting factors in this approach to using the AE data collected is loss in detail caused by the way the data is combined to form these plots. AE data points are created for each point on the ΔK vs da/dn plot to allow a direct comparison. Overall, there are far more AE hits than there are ΔK vs da/dn points. Therefore, the AE data between these points is averaged to form an average AE data point per cycle. Whilst this still allows any overall trends to be

observed, it is clear that a lot of detail that could be used to develop an effective monitoring system is lost, limiting the use and accuracy of any system developed using this approach alone. This approach does however act as an effective feasibility study. Trends have been identified with all steels and AE parameters investigated in this study, these trends differ between the steels but this is expected as the AE response is unique to each material.

To develop an effective real-time monitoring system it is clear that more detail and information is needed and that entirety of the AE signals recorded must be exploited to avoid any loss in data. As the AE data will vary for each material it is expected that the AE data should be able to identify the presence of defects in the crack growth path. This is not born out in the approach used in the chapter except for potentially being the cause of some the variation in AE data, further analysis of the AE data is needed to confirm this. Even using a relatively simplistic and generalised approach used in this chapter patterns and trends in the data have been identified. The subsequent chapters in the present thesis will seek to build on this by further exploiting the AE data obtained using the same approach as that in the present chapter as well as looking at other novel techniques to exploit the AE signals generated during fatigue crack growth.

3.3 Summary

This chapter has investigated the use of a commercially available AE monitoring system to monitor fatigue crack growth in three steels, R220, R260 grade and cast manganese steel. Three parameters were investigated: acoustic counts, energy and duration. An upward trend in all three parameters was shown by both R220 and R260 grade steel in line with increasing crack severity. From this trend it may be possible to assess the structural health of an asset using this commercial system. The data was more complex for the cast manganese steel due to a non continuous trend making any comparison challenging.

It is noted that the data is processed into packages or hits of data by the commercial AE system. This packaging of data simplifies the output data potentially reducing the accuracy of the system. It is proposed instead to use the complete AE waveform captured using a custom AE system. The remained of this work will focus on the development of a custom AE system for this purpose.

Chapter 4

AE Waveform Analysis

Until relatively recently, research on AE condition monitoring has primarily focused on the use of key parameters taken from the data recorded. For example, parameters which are considered to be more useful include AE energy, signal duration and counts and have been studied herewith. Whilst the use of these parameters has proven to be successful for many applications, it does not necessarily capture all the information related to the size and severity of the propagating crack. Analysis of the complete AE waveform would instead allow for the development of a highly accurate and sensitive SHM system.

The problem facing the development of these systems is a big data problem (Malekzadeh et al., 2015). Even using key parameters, as this study has done so far, large volumes of data can be recorded and are required to create an accurate picture of the current structural health of an asset. Adding in more parameters to address the shortcomings of these key parameters does not solve the problem and results in even larger and more complex datasets. Until recently this has limited the development of SHM systems using AE techniques, as the computational power required to process the data was not readily available or low cost.

The recent development and availability of high-performance computers at much lower cost and hardware size to researchers has meant that the processing of data is no longer limited to small datasets or simplistic signal processing techniques (Bacioiu, 2019). In recent years there has been a boom in the use of high-performance computing in the development of novel data processing techniques, in particular machine learning capabilities, to rapidly process big data sets (J. J. Lee et al., 2005).

For AE monitoring this removes the limitation of traditional signal processing methods and approaches to this problem in handling large datasets. In Chapter 3 this work argued that the complete AE waveform could be used to develop a highly accurate automated SHM system overcoming the limitation of using the key acoustic parameters investigated so far in the thesis. This would of course generate a large quantity of data which it would not be possible to analyse in an effective way in real time using manual methods.

Therefore, the present study has looked into the use of machine learning techniques for analysing these complete waveforms and developing an automated SHM system. The present chapter covers the principles behind machine learning techniques for signal pattern recognition and the use of signal processing to extract key features for AE waveforms in the time, frequency and time-frequency domains.

4.1 Machine learning

4.1.1 Machine learning approach

Machine learning techniques sit within the umbrella of artificial intelligence (AI), which covers a wide range of high-performance computing applications including computer vision and natural language processing (Bacioiu, 2019). El Naqa and Murphy describe machine learning techniques as "computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment" (el Naqa & Murphy, 2015). These machine learning techniques allow a computer to gain experience of a task and progressively improve its execution of this task using the acquired experience. The way in which these algorithms learn from data can be split into many subdivisions defining a wide range of possible algorithmic approaches. Each combination of machine learning approaches used will be tailored specifically to the problem being solved by optimisation to a particular dataset or the level of pre-existing knowledge about the data being studied.

The machine learning algorithm selected has a large impact on the accuracy of the data analysis and the computing capability required to run it. The selection of the correct algorithm and the parameters that define its architecture is not necessarily a tick box exercise instead it could more closely be described as informed trial and error to reach optimization.

To optimise the selection of the parameters and to reduce the number of times the algorithm must be run on the training data. In order to find the optimal solution cross-validation can be used to test multiple parameters and compare the accuracy of the resulting algorithms. Cross-validation works by training several modals on separate subsets of the training data which are then validated on other subsets of the training data. The outputs from each modal are compared allowing the effect of each parameter to be compared increasing accuracy and for overfitting of the modal to the detected and reduced.

In all cases, the available data to which the machine learning technique is applied to should be split into a training and a test dataset. Ideally, the bulk of the data should be used to train the system with a small representative batch of data set aside to act as an unseen dataset to confirm the accuracy of what the system has learnt. This is necessary because of the way in which machine learning operates. In machine learning computers simulate learning by applying a series of statistical operations to a dataset until some distinct characteristic is found, so that when it is applied to the data, it maximises the number of times the system can correctly characterise the data of that sort. Whilst this is a powerful way of extracting features from large datasets that would otherwise be impossible to find it does present some limitations.

When data is inputted into a machine learning algorithm that algorithm is programmed to continue applying to the data until a solution is found or a limiting factor is reached (these will be discussed later in the present study). Therefore, even if there is not a definite feature to extract it is possible that the algorithm will find a solution. This is described in Watanabe's

ugly duckling theorem which finds that a similarity can be found between any two objects unless some control bias is added (Watanabe, 1985). These cases can often be identified as low levels of accuracy on both the known and unknown data. In other cases, a high level of accuracy can be achieved overall however the features extracted are not those relevant to the current task, for example classifying animals by their colour rather than by animal species (Muir et al., 2021). It is also possible to reach a high level of accuracy on the training data set yet when this is tested on the unknown set the accuracy of the system drops dramatically, this is known as overfitting (Bacioiu, 2019). This can be caused by many factors but is a particular problem where the training data set is either two small or not representative. Accuracy, in this case, can be increased by increasing the data available and ensuring that the data is effectively randomised between the test and training sets.

Whilst machine learning techniques are powerful tools for the extraction of features from data and the subsequent application of these learnt features to classify unknown data they are limited in their inherent rigid analysis of that data. That is, they are unable to adapt to scenarios that they have not seen before. As an example, if a machine learning algorithm that has been trained to classify apples and oranges is presented instead with a banana it will not be able to classify it as that as it has not seen a banana before and therefore it will classify it as the nearest known category. In any case, the error would be obvious to an operator monitoring the system as they would be able to identify the difference. Now apply this to the scenario studied in this project. If a machine learning algorithm is monitoring crack growth and the crack growth behaviour differs from that which it has been trained on the algorithm will classify the data however this will not be representative of the actual scenario present. Therefore it is important that the training data used to train a machine learning algorithm realistically and accurately represents all scenarios it is expected to encounter (Chang et al., 2000). Where this is not possible novel analysis should be applied to all received signals identifying those that differ from any

previously seen signal. These signals can then either be rejected as outliers or passed to an expert user to be analysed rather than passed through the classifier which would mis-classify these signals.

4.1.2 Supervised vs unsupervised learning

When selecting a machine learning algorithm first one must consider how much information and knowledge exists in the data being investigated. There are two main scenarios to which machine learning is applied in data. Firstly, the dataset is a labelled dataset, that is the input data has a known output. In this case, the machine learning is applied to train algorithms that can classify data or predict outcomes from a data input. The second scenario is used where the training data to which the algorithm is applied is unlabelled. That is the output of a particular input data is not known. This approach is used to find data groupings and hidden patterns in data. It is of course, possible to use a combination of both approaches which is useful in scenarios, where only a limited proportion of the data is labelled. Whilst these approaches differ in the data to which they are applied, the same underlying algorithms can be applied in all three scenarios, for instance, the use of neural networks.

In this study, the AE waveform data collected using the custom system is time-stamped when it is recorded. This timestamp can be compared with the DCPD plot that is recorded in parallel during fatigue loading to give the length of the fatigue crack when the AE data are correlated. This comparison of the AE data and the crack length allows the data to be labelled. Therefore, these data lend themselves towards the use of supervised learning characterising the data with respect to the crack length. Thus, correctly applied machine learning algorithms should then be able to identify key features in training data corresponding to crack length. By doing so it could then be applied to unknown data to indicate the current crack length also.

In Chapter 3 it was discussed that it would be beneficial for an SHM system to be able to detect defects within a structure where a crack is propagating, as this would change the crack growth

behaviour and therefore affect the remaining useful life of the material. For the development of this system supervised learning could again be used with signals attributed to a defect where they are recorded in the same location. Another approach would be to use either unsupervised or a combination of both techniques to extract features from the signal that otherwise may be missed by labelling data on only observed defects.

4.1.3 Neural networks

Neural networks (NN) are a form of machine learning that seek to emulate the way in which a brain operates through a series of interconnected neurons to achieve human-like problem solving performance (Bishop, 1995; Lippmann, 1987). Whilst the recent advances in computer technology have driven the extensive use of NN, they have been proposed for use as far back as in the 1940s in the pioneering work of McCulloch and Pitts (Mcculloch & Pitts, 1990). The neurons which they proposed are still used as logic circuits today (J. A. (James A. Anderson & Rosenfeld, 1988). These neurons work by each neuron performing logic functions with a set threshold. If that threshold is exceeded the neuron fires. Logic functions of these sort can form simple task, such as deciding if something is on or off. In their work Hebb (Hebb, 1949) presented the first network that was able to learn. This was achieved through adapting the strength of the connection between two neurons if they were activated at the same time. The increased connection strength between two neurons that are inactive has since been added (Fausett, 1994). These have led to the subsequent development of computer simulations (Rochester et al., 1956).

A big step in the development of the NN that resemble those now used today was the development of Perceptron's by Rosenblatt (Rosenblatt, 1958) and subsequently by others. These are made up of input neurons (the input layer) connected by fixed waited connections to other neurons. The weights of these connections can then be adjusted iteratively to allow them to adapt. This was further improved by Widrow and Hoff (Widrow & Hoff, 1960). In their

proposed model, the connection weights are adapted when the incorrect classification is made to output the lowest mean squared error. By doing so the NN produced using this learning function can adapt the data that is not identical to the training data. This makes them of more practical use in real-world applications, where they may encounter non 'clean' data. It must be recognised that while these networks show great potential through the use of iterative learning, they are not suitable for all situations as there is a limit to what they can learn. These limitations were studied mathematically in 1969 by Minsky and Papert (Minsky & Papert, 1988).

Modern NNs use backpropagation to increase the accuracy by feeding back through the hidden layers of the NN information regarding errors in the output layer. The weights of the connections can then be adjusted and the data run back through. By continuing this back and forward of information, the NN is able to reach higher levels of accuracy. This method was developed independently by Bryson and Ho (Bryson & Ho, 1969), Werbos (Werbos, 1974), Parker (Parker, 1985) and LeCun (le Cun, 1986)

Due to the wide range of potential parameters sizes and layers that can be added to define the architecture of a neural network there are a huge number of different NNs that have been described in the literature with a wide range of uses (Masters, 1994). Whilst the selection of the NN has an impact on the resulting accuracy there is often more than one acceptable solution. The use of transfer learning has proven an effective way of utilising previously developed NN and adapting them to a new problem without the need for the development of an entirely new network and the resulting volume of accurate data that would otherwise be required.

4.1.3.1 Neural network architecture

NNs vary in complexity from the McCulloch-Pitts neuron to the complex multilayer NN in use today to the ultimate goal of complete replication of the complex NNs that are our brains (J. A. (James A. Anderson & Rosenfeld, 1988). In its simplest form a NN consists of an input and a neuron which carries out a mathematical operation on this input and produces a scaler as an

output using an activation function (sometimes referred to as a transfer functions)(Hagan, 1996) this can be seen in Figure 4.1.

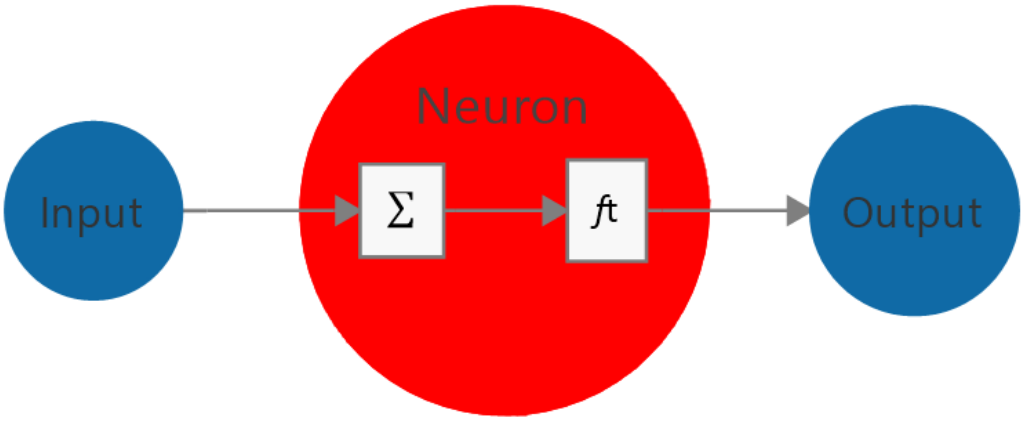


Figure 4.1 *Diagram single input Neuron showing the flow of data from the input node through the neuron to the output node. Within the Neuron ∑ is the mathematical operation applied to the input data and ft is the activation function which produces the scaler output to the output node.*

The single node neuron shown in Figure 4.1 is the building block of a NN that completes highly complex tasks. However, on its own is only capable of yes/no solutions to a simple task in a similar way a single biological neuron cannot solve a problem or learn but combined with many others, hugely complex tasks can be solved. As the tasks become more complex so does the NN required to solve them with a NN usually consisting of multiple inputs connected to multiple neurons. NNs generally have several neurons working in parallel. These parallel working neurons are said to be in a layer (Figure 4.2).

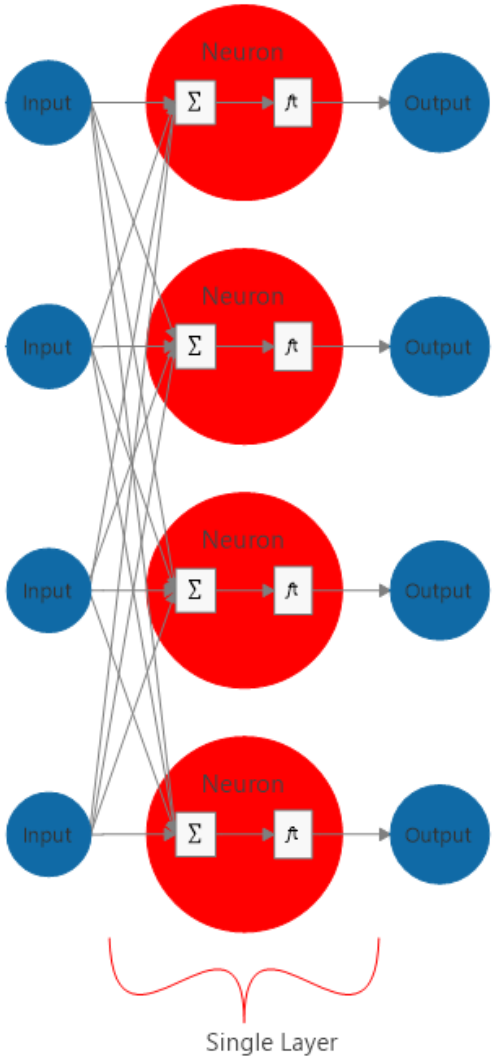


Figure 4.2 *Schematic diagram of single layer NN. As these neurons are all working in parallel they are referred to as being in a layer.*

Single layer NNs, often referred to as perceptron's from the work of Rosenblatt (Rosenblatt, 1958), are able to effectively find a solution if the data being analysed is linearly separable as described by the perceptron convergence theorem, a simple proof of which can be seen in the work of Bishop (Bishop, 1995). Where data is non-linearly separable further layers are required to achieve a solution, the limitations of perception were extensively studied mathematically in the work of Minsky and Papert (Minsky & Papert, 1988) early on in the study of these networks. This need not however, lead to a large increase in the number of layers. In fact, due to the increased complexity and hence computing requirement linked to increasing the number of

layers a network should be designed with the minimum number of layers capable of reaching and acceptable approximation of the solution.

There are many studies looking to prove the minimum number of layers required for a given problem. For example the addition of one more layer, forming a two-layer NN, providing the number of hidden units is large enough, can arbitrarily well map from one finite-dimensional space to another (Bishop, 1995). Adding a third layer adds an advance on this and is able to achieve arbitrary accuracy for any smooth mapping (Lapedes & Farber, 1988) and represent any arbitrary decision boundary to this level of accuracy (Bishop, 1995).

With accuracy achievable with only a few layers the adding of further layers to form deep-learning networks in comparison with these shallow networks often becomes unnecessary and computationally expensive. They are thus, rarely used in practice (Hagan, 1996).

Having introduced multilayer NNs it is worth adding in some further definitions. Up until this point the neurons used to classify data have been referred to as layers. However, these can be further refined. The final layer in the network providing the network output is referred to as the output layer. Hidden layers output into all layers other than the final layer (the output layer). So a two-layer network would have a single hidden layer and one output layer as shown in Figure 4.3.

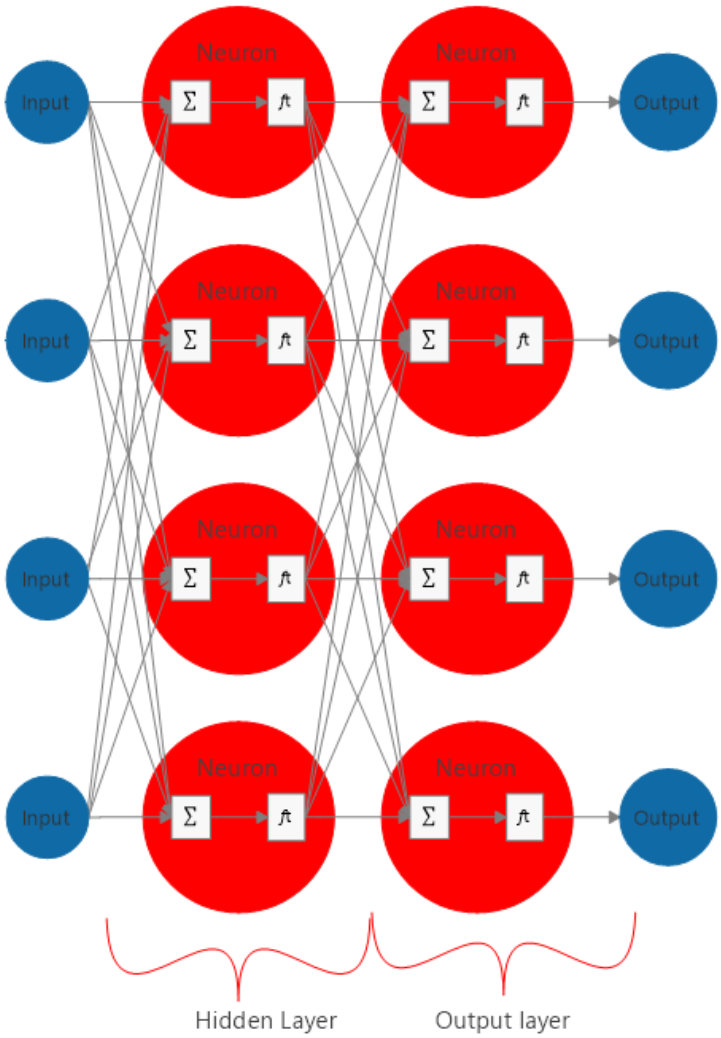


Figure 4.3 *Schematic diagram of a two-layer neural network showing the difference between the hidden layer and output layer.*

Figures 4.1-4.3 show a simplistic representation of the overall architecture of a NN. Whilst these representations are useful in understanding the overall connected nature of a NN the real workings of the NN are found in the algorithms, weights and biases that make up the neurons and connections of the NN. The selection of these is determined by the problem being solved. In the case of the AE data studied herewith, the aim is to categorize each input into set severity levels using the NN techniques to find the characteristic pattern for each severity level. If the AE signals can be characterised in this way, then this approach will produce high levels of accuracy. In terms of NN computational complexity, pattern recognition networks are

relatively simple (Fausett, 1994). Some of the earliest NNs were used for pattern classification. For example, the work of Specht, which analysed electrocardiogram data using NNs to detect heart abnormalities (Specht, 1967).

4.1.3.2 Pattern recognition Neural Networks

Pattern recognition NNs seek to characterise data according to characteristic patterns in the data. Pattern recognition NNs are widely used across a range of applications with great success. NNs are powerful tools in the extraction of these features often finding patterns that would not be identifiable by a human operator. The use of these approaches must however be carefully controlled and understood to ensure the desired classification is achieved. NNs can be described as black boxes in terms of how they identify these patterns. That is, it is challenging and often impossible to establish exactly what features the NN has learnt (Lin et al., 2017).

This has proven to be the case in several studies that have developed machine learning algorithms using unsupervised learning techniques. These studies have been able to separate clusters of AE data into groups, however what these groupings represent is not understood. Whilst this is promising for the AE machine learning field as a whole on that the data is separable, for an end user this does not produce a usable model. In an attempt to effectively classify these groupings several studies have investigated the corresponding fracture surface using this information to label the learnt groupings.

In some cases, misclassification would be clear to an operator. For example, a pattern recognition NN designed to classify types of transport may instead classify them by colour. This misclassification would become clear during through testing of the network and the hyperparameters adjusted accordingly. In cases like this the problem would be largely solved by using supervised learning rather than unsupervised. However, in more complex cases such as waveform analysis this does present a problem.

This problem is the NN is identifying patterns that even an expert user may not be able to identify, thus, making it also challenging to identify when the data has been misclassified. This is often solved through thorough testing of the network using representative test data. Researchers have sort ways to analyse the features that a NN extracts. Erhan *et al* sort to visualise each hidden layer by generating synthetic signals that maximised the output from each hidden neuron therefore identifying the features or features identified by it (Erhan et al., 2009). Using this method, Lin *et al* found that hierarchical features are learned layer by layer (Lin et al., 2017). Despite these challenges, pattern recognition networks are widely used with great success. To ensure their success, a certain level of understanding is required with reference to the data being analysed so as to select the correct learning algorithms and hyperparameters.

There are several different approaches to the way in which pattern recognition NNs learn. However, they are all based around the use of statistical operations applied to a dataset, essentially learning by finding features that are statistically more likely to be present in each classification. Traditionally, AE researchers have used the K-means algorithm to classify AE waveforms (de Oliveira & Marques, 2008; Godin et al., 2004; L. Li et al., 2014; Ramirez-Jimenez et al., 2004). Whilst these have proved effective at classifying several damage states there has been limited evaluation of the accuracy of the characterisation of AE data using this method (Muir et al., 2021).

The classification of damage is not a one size fits all model. That is for each material studied the classification network used must be adjusted. Many of the studies that have looked to develop pattern recognition and machine learning techniques for AE have focused on their application to composite structures. The studies on K-means algorithms are an example of this. This is partially due to the failure modes of interest in these materials lending themselves to classification. These studies have looked at the classification of failure methods, such as fibre

debonding and fibre pull-out. These failure modes are distinct but also give an indication of the damage level in the material being monitored. This distinct nature of these categories has led to the successes of grouping algorithms such as the K-means algorithm.

In this study, the author seeks to classify the severity of damage in steels with the categories defined by the length of a fatigue crack. This classification system has less well-defined boundaries as it is not defined by failure mode. Therefore, the clustering of data is less distinct with overlap between the boundaries. Therefore, other pattern recognition approaches are needed. Pattern recognition NNs lend themselves to this less easily definable data finding complex underlying patterns within the data that can accurately predict these arbitrary severity categories.

Pattern recognition systems have been applied to other aspects of AE analysis beyond the classification of severity studied in this study. Suwansin and Phasukkit used these techniques to identify, whether the AE source originated in the head, web or foot of the rail (Suwansin & Phasukkit, 2021). Traditional location methods using multiple sensors have been able to accurately locate AE sources already. However, this work used machine learning techniques to improve on this by requiring only a single sensor to locate the AE source. Ebrahimkhanlou and Salamone, also sought to use machine learning techniques to improve the accuracy of AE source location (Ebrahimkhanlou & Salamone, 2018). They utilised the pattern recognition capability of machine learning techniques to identify the characteristics of signals that are caused by reflection and other interference to the AE signal to reduce the number of false positive AE events detected. Like Suwansin and Phasukkit, Ebrahimkhanlou and Salamone, sought to accurately locate an AE source using a single sensor. Both studies used a convolutional neural network approach.

4.2 Signal Processing

4.2.1 Background noise removal

In the development of an effective AE monitoring technique we are presented with two problems in terms of the waveform analysis; firstly the removal of background noise (Moore et al., 2005) and secondly the extraction of key indicator features from the waveform. The removal of background noise is considered key to the effective evaluation of AE signals in that it removes any unwanted signals leaving only the proportion of the data which is useful to the classification of the structural health of an asset.

Background noise can however refer to much more than environmental or external noise. AE signals are generated from a variety of sources such as elastic deformation of the material by temperature or load which would not be an indicator of defect growth but would be detected as an AE signal (ISUNDE-ed, 2022). It is clear therefore that filtering alone cannot be used to isolate the key AE parameters for damage detection and other analysis techniques are required to extract these.

Filtering is commonly used to remove background noise from the raw AE signal. Filtering of the signal is simply applying upper and/or lower limits to the signal that will be recorded. In Chapter 1 of this thesis the different capturing methods HDD and TDD were discussed, these are forms of filtering. In the case of the complete waveform captured in this study a TDD approach was used. This sets a limited time over which a signal is captured. Whilst this will not remove background noise from the captured signal, the volume of data captured is limited to a finite quantity.

If the frequency range of the AE signals is known all frequency outside of this range can be removed. The frequency ranges used to filter AE signals have been widely studied across a range of materials.

Most approaches to removal of background noise have been based on data with randomly artificially-induced noise and/or noise that follows a particular distribution. Under real world conditions, this is often not the case with background noise being randomised. As the traditional mathematical approaches for removing background noise have been based on predictable data and are effective at this, they fail when applied to unusual noise distributions. NNs on the other hand are able to cope with this variation in background noise distribution quite successfully still extracting the relevant key data for classification (Masters, 1994).

4.2.2 Waveform analysis

Before being passed through a pattern recognition network, the recorded AE waveform must first be analysed to consistently extract key features from the waveform under all conditions. There are a variety of techniques used for this operating in the time, time-frequency, and frequency domains with no one approach yet used predominantly. This study will investigate the effectiveness of a selected technique from each of these domains and compare their effectiveness when applied to a custom-built machine learning algorithm.

4.2.2.1 Time domain analysis

Historically research has focused on the analysis and classification of damage by using specific AE parameters in the time domain as in Chapter 3 of this report (Barré & Benzeggagh, 1994; de Groot et al., 1995; Morscher, 1999), focussing on AE energy, counts and duration. Broad trends in the data can be determined, however the potential highly dimensional data of the complete waveform is lost (Muir et al., 2021). Moving beyond these parameters was, until relatively recently, limited by both recording equipment and data storage and computational power. Several of the parameters analysed are also known to be susceptible to interference from experimental factors and the AE propagation pathway (Muir et al., 2021). In addition, the use of FEA and other simulations have cast doubt on the repeatability of the models developed from these parameters outside closely controlled environments (Muir et al., 2021). The work of Ould *et al* found that these parameter-based techniques were not sufficient to discriminate between fatigue crack growth stages (Ould Amer et al., 2013). Despite these doubts researchers have successfully proven a link between many of these parameters and crack growth events and severity (Mohammad et al., 2014; Papaelias et al., 2014, 2018; Sikorski, 2013).

It is possible to analysis the signal in the time domain without the need to focus on these parameters, thereby utilising the entire recorded AE waveform. Statistical parameters including Crest Factor, Skewness, Kurtosis and root-mean-square (RMS) have been used to do this

(Amini, 2016; Vallely, 2015). In this study RMS analysis will be applied to the recorded RMS data and the results fed to the custom machine learning algorithm.

Figure 4.4 shows an example of a recorded AE waveform showing several strong signal peaks. A solution to analysing these waveforms could be to look for these peaks in the time domain and categorize the signal for example by peak-peak amplitude analysis. This technique has a strong draw back in that it does not consider the energy contained within the signal. It is possible that a peak may appear strong and therefore significant but contain no energy. Peaks of this nature are associated with background noise and therefore, would generate error in any subsequent analysis of the signal.

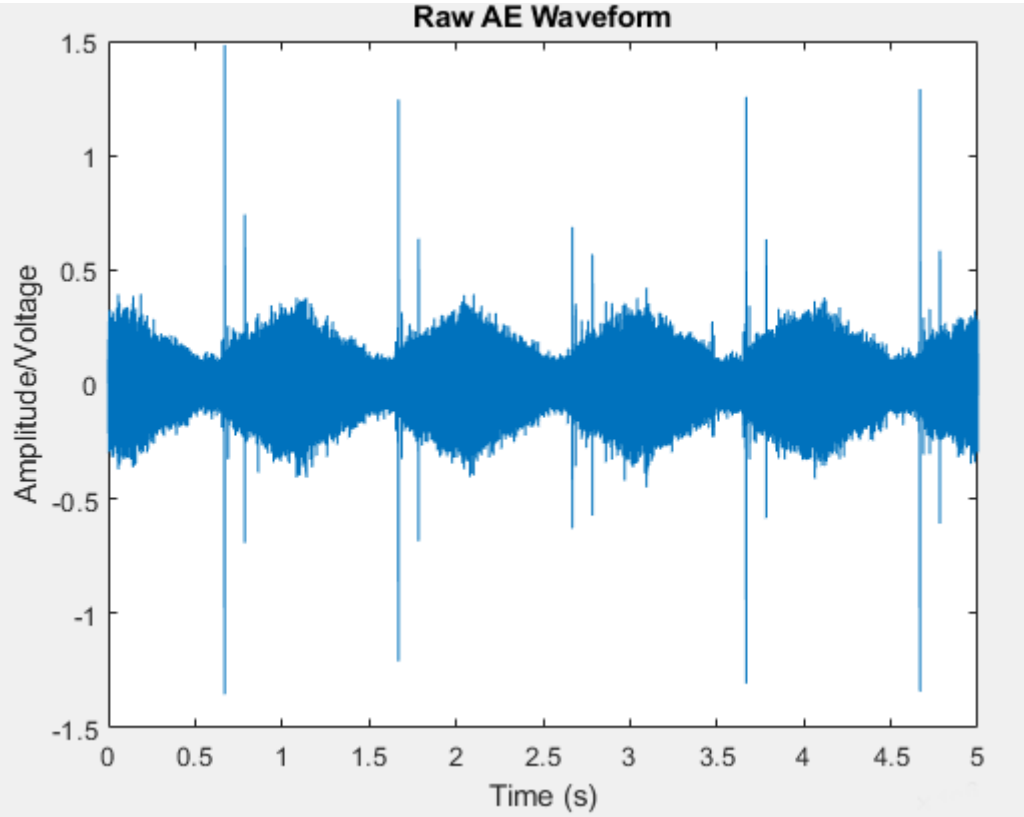


Figure 4.4 *A typical AE signal recorded during fatigue testing of a cracked steel sample.*

To overcome this the energy contained within a signal can be analysed using RMS. The RMS is calculated using Equation 4.1.

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2}$$

Equation 4.1 *Equation used to calculate the RMS of a signal* (Lebold et al., 1985)

The moving RMS extends this principle allowing the energy of a complete signal to be analysed by dividing the signals into windows of a set length and the RMS calculated across them. The length of the window used in this approach is important. However, there is no standard on window length selection, so trial and error must be used to find the optimum window length. The optimum window length ensures that excessive background noise is filtered out whilst maintaining all the data required for accurate analysis. Figure 4.5 shows the moving RMS plot of the waveform shown in Figure 4.4. To obtain this a window of size 100 and a sigmoidal correction factor are used. This window slides through the data calculating the RMS of the data enclosed by the window at each data point.

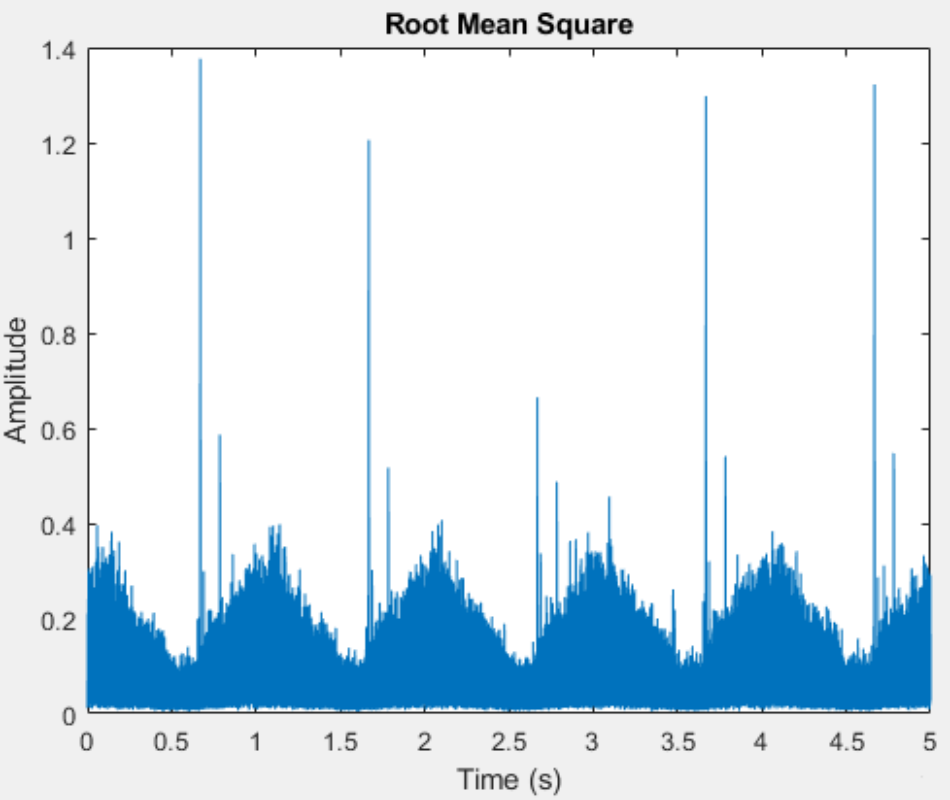


Figure 4.5 *The moving RMS plot calculated from the AE waveform shown in Figure 4.4*

By extracting the high energy signals from the waveform much of the background noise can be removed and the energy of the signal can be analysed. RMS however, becomes limited where the background noise signals also contain high energy level peaks which would then not be filtered out. RMS has proved effective at detecting the presence of crack growth events which are indicated by the high energy signals detected (Vallely, 2015).

4.2.2.2 Frequency domain analysis

Interest in the frequency domain is born out of the great depth of analysis that can be applied, extracting indications of the severity of damage as opposed to just the presence of a crack. Frequency domain analysis is widely used for reciprocating or rotating machinery such a gear boxes and bearings (Amini, 2016).

As is the case with much of the research into AE frequency and time-frequency analysis techniques have only recently been made feasible by the improvement in equipment storage

capability and processing power. Despite these previous processing limitations a wide range of studies have looked at the use of frequency domain analysis to extract features from an AE waveform for example Fourier transforms, Spectral Kurtosis, Moving Cepstrum, high and low Frequency resonance. This study will focus on the most widely used and simplest of these the Fast Fourier Transform (FFT), which is an implementation of the Discrete Fourier Transfrom (DFT) that reduces the number of computations required therefore speeding up the processing of the signal. The DFT is shown in Equation 4.2.

$$Y(k) = \sum_{j=1}^{n} X(j) W_n^{(j-1)(k-1)}$$

Where

$$W_n = e^{\frac{(-2\pi i)}{n}}$$

Equation 4.2 *Equation used to calculate the FFT of a signal. Where x is the input signal in the time domain and n is the signal length.*

The FFT breaks down a signal representing it as the frequencies that make up the signal and their corresponding amplitudes (NationalInstruments, n.d.). That is taking the signal from the time domain to the frequency domain this is shown in Figure 4.6 which is the FFT plot of the AE waveform shown in Figure 4.4.

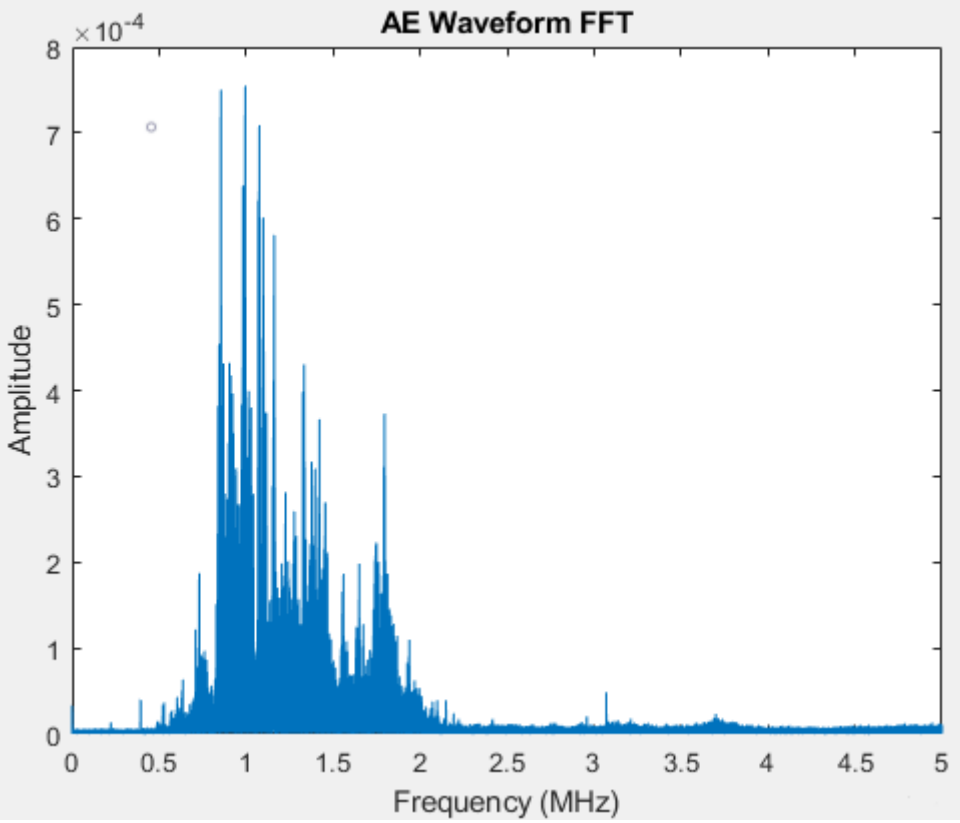


Figure 4.6 *The FFT plot of the AE waveform shown in Figure 4.4 with a Hann filter applied.*

The FFT approach assumes that the signal covers a complete period of a repeating periodic signal. In an experimental setting it is possible to control the signal acquisition so that signals are recorded for a complete cycle of the fatigue testing sequence. However, this does not necessarily match the corresponding AE signal meaning that the signal recorded is not periodic in nature. This is exasperated in real world conditions where the fatigue loading is stochastic and therefore challenging to precisely tailor the AE acquisition to ensure a periodic signal. Where the signals being analysed using FFT are non-periodic spectra, leakage occurs with high frequency's displayed that are not part of the original signal (NationalInstruments, n.d.).

To overcome this, windowing is used to smooth the start and end of the signal to create a periodic signal. There are a range of different windowing functions that can be applied to a signal each of a slightly different shape (NationalInstruments, n.d.). In this study a Hann window function is used. A Hann window shown in Figure 4.7 has a sinusoidal shape which

starts and ends at zero thereby artificially creating a periodic signal by suppressing the edges of the recording so that the start and end of the signal are at the same origin. As the signals recorded in this thesis are taken at random time intervals the signal is unlikely to start and end and the same amplitude and therefore not periodic. It is therefore important that a window is applied to the signal to ensure that when the FFT is calculated edge effects are not observed.

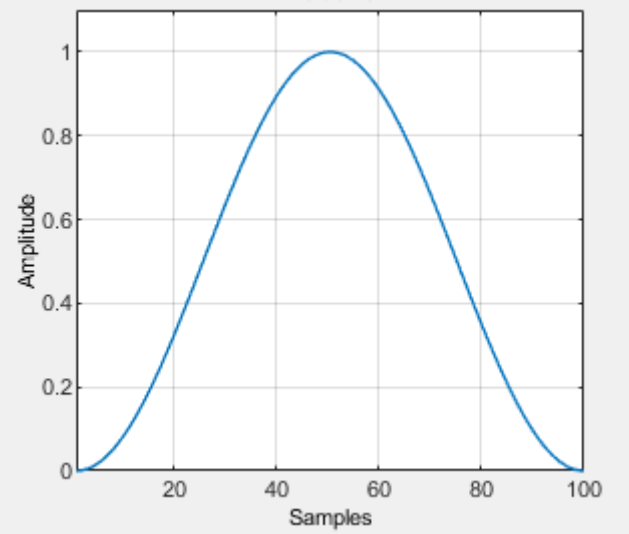


Figure 4.7 Graphical representation of the Hann window used to create a periodic signal to reduce spectral leakage when FFT is applied.

The constituent frequencies of the AE signal can now be analysed and changes in frequencies can be compared to crack growth and severity. The frequency of the AE signal released during fatigue crack growth is expected to change with different crack growth events, for example the presence of inclusions, as well as the severity of the crack. Analysis in the frequency domain should therefore allow for the severity classification of a signal and the development of an automated classification system.

As with all analysis technique there are limitations to this technique and frequency domain analysis. Frequency domain techniques are often based on the assumption that the signal is stationary or repeating as in the case of gear sets. In the situation of rail defects this is not the case.

4.2.2.3 Time frequency analysis

To overcome this, researchers have investigated time-frequency analysis techniques that are able to analyse nonstationary signals. Wavelet transform methods have been investigated to overcome the need for a stationary signal, this makes it potentially more useful for use in the case investigated in this study. Whilst FFT and RMS are relatively simple to implement, wavelet transforms are more complex to both calculate and interpret.

The Continuous Wavelet Transform (CWT) can be used for non-stationary signals where the signal is non-repeating, that is not a repeating sigmoidal signal. The CWT equation is given in Equation 4.3 where; a is the scale parameter corresponding to the pseudo-frequency, b a translation parameter and φ^* is the conjugate of the complex-valued function φ.(Muir et al., 2021).

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\varphi^*\left(\frac{t-b}{a}\right) dt$$

Equation 4.3 *Definition of the CWT where a is the scale parameter corresponding to the pseudo-frequency, b a translation parameter and* φ^* *is the conjugate of the complex-valued function* φ.(Muir et al., 2021)

This technique has been successfully used to characterise AE signals across a number of applications from the detection of drilling burr formations (S. H. Lee & Lee, 2008) to the detection of leaks in water pipes (Ahadi & Bakhtiar, 2010). This technique has advanced the way in which AE signals are analysed, for example traditionally multiple sensors were needed to locate the source of the AE signal, however this technique has allowed signal locations to be classified for example the work of Hanstad *et al* who qualitatively located AE sources using wavelet transforms (Hamstad et al., 2002). One of the main advantages of this technique is that the time data is not lost from the analysed data allowing changes in the signal frequency to be used to classify the AE signal.

4.3 Summary

This chapter gives a brief overview of the principles and algorithms that will form the basis of the machine learning technique developed in the following chapter. An overview is given of machine learning techniques and their development followed by a focus on Neural Networks used for their powerful processing capability achieved by emulating the way in which a biological brain functions.

For use in any machine learning technique pre-processing must be applied to the raw AE data to remove background noise and extract key features. This chapter discusses both steps and techniques that are commonly used. The different processing domains (time, frequency, and time-frequency) are discussed with the techniques that will be used to pre-process that data in each domain later in this thesis highlighted. They are RMS, FFT and CWT.

Chapter 5

Machine learning: Development and Results

The development of the machine learning algorithm considered in this study seeks firstly to build and train a neural network using labelled acoustic data. Secondly, it aims to deploy this trained network to analyse unknown acoustic data generated by the propagation of a crack and characterise it according to its severity. The code is written in MATLAB®, making use of the available signal processing and machine learning toolboxes built in already. The present chapter will describe the general development of the algorithm presented herewith as well as the results produced. The relative benefits of the different algorithms selected will be discussed. The code developed in the present study, generates a neural network which is trained, tested and then can be applied to unknown AE data. It was noted in chapter 4 that neural networks are a subset of machine learning therefore these terms will be used interchangeably to describe the code developed here.

5.1 Raw data processing

Before a neural network can be designed and implemented the data that it will analyse must be processed to allow representative features to be extracted. The processing techniques that can be applied to the raw acoustic data are detailed in chapter 4. In this study three algorithms have been explored; Fast Fourier Transform (FFT), Root Mean Square (RMS) and Continuous Wavelet Transform (CWT). These three techniques cover three processing domains, including time, frequency, and time-frequency, and are discussed in detail in chapter 4. Whilst these three techniques are different in the way they are used to process the data and the outputs they

generate; the overall processing route can be summarised simplistically by the pseudo code shown in Figure 5.1.

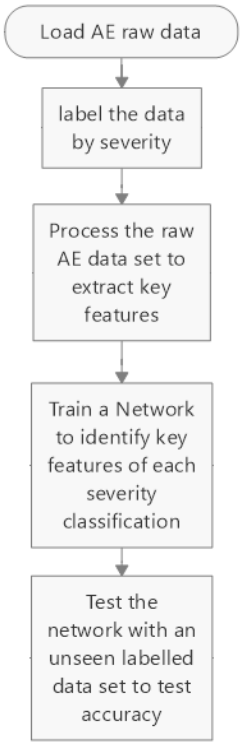


Figure 5.1 *Pseudo code for the raw waveform processing prior to analysis by the machine learning algorithm. Represented visually here it shows the steps required for all three of the analysis techniques used in this study.*

Analysis of a single signal using the selected signal processing techniques is relatively simple requiring only a few lines of code to be written in order to import the raw data, process it, and then output it numerically of visually. For example, the code shown in Figure 5.2, which applies FFT to a raw data signal, outputs both a numerical and visual representation of that data.

```
[spec, freq] = barefft(data,Fs);
figure();
plot(freq, spec);
xlabel('f (Hz)');
ylabel('Amplitude');
```

Figure 5.2 *MATLAB code required to take a numerical raw data signal, apply FFT and output the result both numerically and visually.*

This code, however, only represents part of step 3 of the sudo code in Figure 5.1. Steps 1-3 must be completed before the data can be passed to the machine learning algorithm. That is the data must be loaded from the recording system, the required data must be extracted, the data processed using signal processing and finally saved in a form usable by the machine learning algorithm. Due to the large volume of data required of machine learning, as described in Chapter 4, this process must be automated for the processing of the training and test data. To manually load, process and save the data would take a prohibitively long time to be practical even with the relatively small dataset used in the present study.

Even using automated processing of the data the present study has made use of the Blue BEAR high performance computing (HPC) facility at the University of Birmingham to carry out this initial processing. In the present study, only a relatively small dataset is used in comparison with that which will be required for the training of a complete system. Any processing system developed must also be adaptable such that it can be later applied to a single data recording for implementation as a real time monitoring system. It also must be capable to complete this process without the need for supercomputing capability. For practical use, it is envisaged that this process will also be automated. It is this approach that has been built into the code in the present study. To automate the process, a loop system is utilised, which both process the AE data and labels the data.

Whilst three different signal processing techniques are used to analyse the data, they all have processing steps in common. In all cases a frequency bandpass filter is applied to the raw dataset. This frequency bandpass filter filters out all frequencies outside of the range of 250-400 kHz. The remaining frequencies are those that correspond to acoustic emission signals recorded.

5.1.1 RMS processing

There are several parameters that can be adjusted to optimise the RMS processing to specific datasets. In the system developed herewith, the three parameters that are user-definable are the window size, the dimension along which the RMS is completed and a correction factor. Several window sizes were considered with a window size of 100 points being selected from these trials. This value gave the best results in removing the background noise, whilst maintaining the peaks of interest. Due to the differences in frequency response of different materials it is likely that this would need to be tailored to each application.

The RMS was run along the time dimension. It is not envisaged that there would be any circumstances where this would be adjusted by a user for the application that this algorithm is applied to. A correction factor of 1 is applied to the RMS.

Once the RMS processing has been applied a findpeaks function is used to select the characteristic peaks. The number of peaks is set by a user. The more peaks the greater the data the machine learning algorithm must work with. It is, however, not a case of just selecting the highest number of peaks that is expected. As each signal differs in the number of AE peaks detected the number of peaks selected must be as high as possible, whilst ensuring that the quietest signal still has more peaks that can be selected. This is because the input size of the data must be the same for all samples in order of the machine learning algorithm to analyse it. Once the peaks have been selected the amplitude of these peaks is recorded. In this study 114,600 peaks were selected from each dataset.

For the training and test dataset this amplitude data is added to a matrix where each sample is represented by a row. In the real-world test data, the sample data is recorded as a single row vector.

5.1.2 FFT processing

Fourier transforms are used to find the constituent frequencies making up a signal. There are several implementations of this transform, as discussed in Chapter 4 this study uses the FFT with a Hann filter applied.

The same extraction of key features is used as that of the RMS processing, peak analysis. As with the RMS, the number of peaks selected is taken from analysis of the output data to find the highest number of peaks selectable from the dataset. In this case 740,000 peaks were selected.

In terms of implementation through coding, this method is the simplest of the three techniques used in the present study. Whereas the other two techniques have parameters that are adjusted to achieve the best results, for FFT only the number of peaks selected from the output can be adjusted.

5.1.3 CWT processing

The pre-processing for the data differs from that of the FFT and RMS generating a spectrogram. This is treated as an image. Hence, an image recognition network is applied as opposed to the pattern recognition network used in the other two cases. The same bandpass filter is applied to the raw data. The output from this is then analysed using a wavelet transform with the outputted spectrogram converted to a RGB image. The generated images are saved as an image datastore.

5.2 Data labelling

Two different labelling methods are required as the data from the RMS and FFT data are numerical, and the CWT data are treated as an image. For the machine learning networks developed and discussed, this difference in data requires variation in the way the labels and data are presented to these networks.

This study has used a classification system informed by the empirical and qualitative classification system currently used for crack severity in RCF scenarios by Network Rail (Network Rail, 2018). This classification system is shown in Table 5.1. In the subsequent coding the label given to each data set is taken from the file name of that dataset, which corresponds to the crack severity as listed in Table 5.1 at the time that the AE data was recorded. Whilst this requires prior knowledge of the severity state the final trained algorithm outputs the predicted severity label requiring no prior knowledge of the severity state.

Table 5.1 *Table of the severity classifications used in this study. It lists three different classifications; 1. The crack length range 2. A descriptive label 3. A numerical classification label* (Network Rail, 2018). *Severity classifications of crack length and the descriptive label are based off*

Crack length	**Descriptive label**	**Numerical classification**
0<0.1 mm	Healthy	1
0.1≤1.5 mm	Light	2
1.6≤3 mm	Moderate	3
3.1≤4 mm	Heavy	4
4.1≤4.9 mm	Severe	5
>5 mm	Very Severe	6

5.2.1 RMS and FFT Data labelling

The machine learning algorithm developed here requires the training and the test data to be labelled using a numerical classification system. When the trained network is applied to real world data these numerical outputs can be outputted to an operator in a more user-friendly format such as colour coding or descriptive.

To label each data set with the corresponding numerical classification the code reads each file name, which contains a descriptive label. It then assigns the appropriate numerical classification and subsequently saves this in a column vector, where the row number containing the classification represents the AE data sample. An example of this is shown in Figure 5.3. This column vector will not be required for any data subsequently analysed using a trained network as this is the purpose of the network.

2208x1 double

	1
1	4
2	3
3	1
4	4
5	2
6	5
7	3
8	2
9	5
10	1
11	4
12	3
13	2
14	2
15	4
16	5
17	3
18	2
19	1
20	4

Figure 5.3 *Example of the column vector created containing the numerical classification for each data sample in either the training or test data set.*

5.2.2 CWT data labelling

For the RMS and FFT data a separate column vector of labels is generated. The images from the CWT processing are instead labelled by the folder in which they are saved within the image data store. For example, all images that correspond to a severity level of 3 (Moderate) will be saved in the folder named 3. As with the loading and labelling of the RMS and FFT data this process is coded to be completed as a loop running through all the files available.

5.3 Machine learning development

Now that the data is labelled and processed, an algorithm must be applied that can link the labels to the data and then extract representative features from the processed data for each classification. As has been discussed this is done using machine learning techniques. As the outputs from the processing using RMS and FFT are similar the same machine learning approach can be used from both. The CWT outputs differ and therefore a different analysis method is required.

5.3.1 Shallow Pattern recognition neural network

For both the RMS and FFT data a pattern recognition network is utilised, in this case a patternet. The patternet algorithm is a feed-forward network with a single hidden layer. The size of this hidden layer, the training function and the performance function can all be adjusted to the application. Due to having a single hidden layer this is a shallow network. The effectiveness of a deep learning variation is also tested on both the RMS and FFT data.

The same network structure was applied to both the RMS and FFT data set. A single hidden layer neural network was used with 28 neurons in the hidden layer, creating a 2-layer NN. The number of neurons in the hidden layer can be adjusted to achieve optimum results. As with many of the parameters used to define the machine learning algorithm, the optimisation of the number of these neurons is found through experimentation of different values. In this study 28 neurons were found to give the best results.

5.3.1.1 Training function

The training function is the part of the machine learning process that learns features from the dataset. That is, it seeks to find the optimum weights and biases to reduce the loss. The loss is how different the prediction is from the truth. If the loss is 0 then the weights and biases are ideal from there the higher the loss the worse the prediction.

There are different ways to analyse this loss and for the algorithm to decide what adjustments should be made to reduce that loss. In the present study, the scaled conjugate gradient (SCG) approach is used. This approach analyses the loss gradient, whether the loss is increasing or decreasing with the change in weights, with the aim to find a minimum value. An idealised loss curve is shown in Figure 5.4. In this case the minimum loss can be clearly identified.

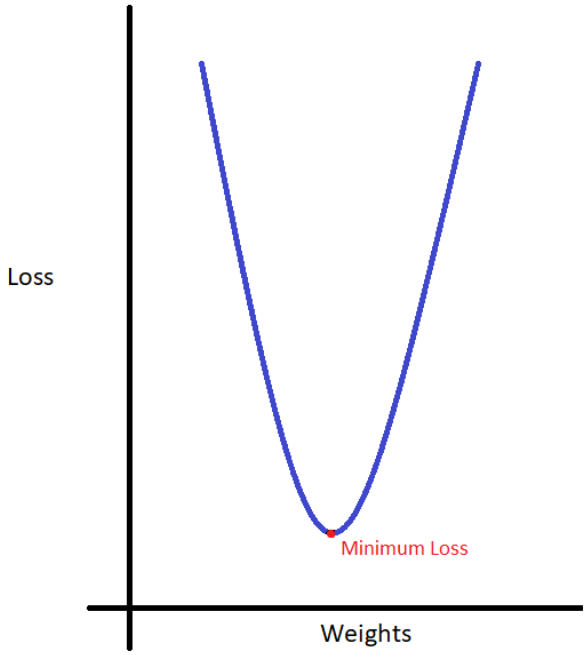


Figure 5.4 *Schematic plot of the performance function with the minimum loss point easily identified as the lowest value of the loss plot.*

In reality, the loss function is more complex than this and it is possible for localised minimum in the loss function to be identified, as is shown in Figure 5.5. The scaled conjugate gradient approach combines a Levenberg-Marquardt algorithm and the conjugate gradient method to create a training function that changes the learning rate with each iteration to avoid a local minimum being selected.

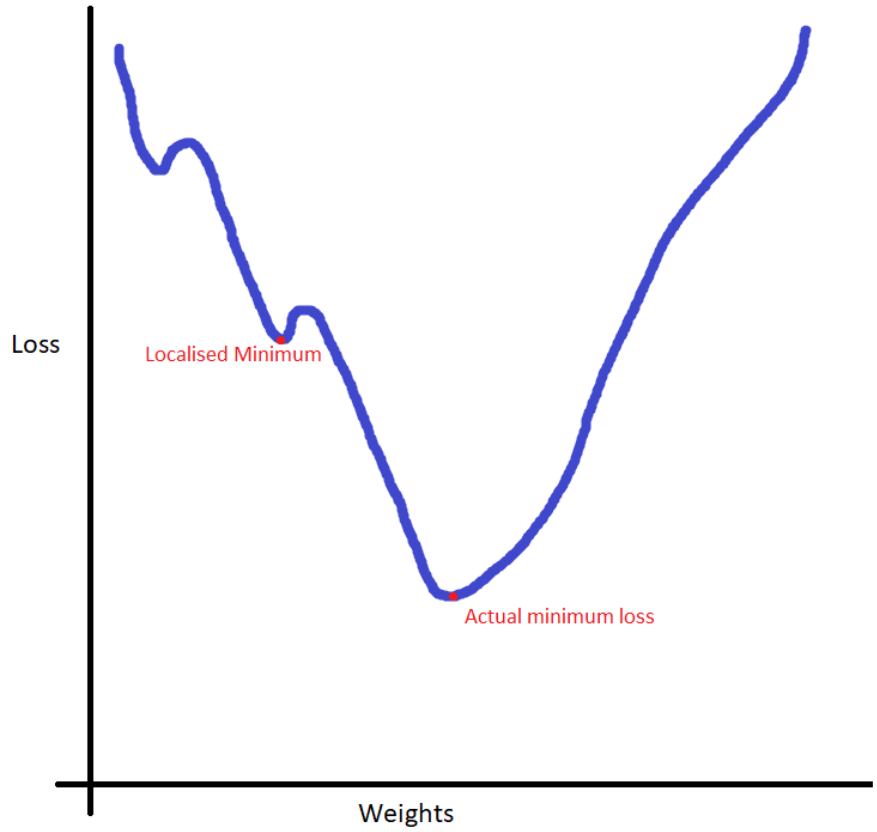


Figure 5.5 *Schematic plot of a loss function where a localised loss minimum has been identified. The identified local minimum has a greater loss than the actual minimum identified as the lowest point of the loss plot.*

5.3.1.2 Activation function

The activation function takes the weighted sum of the inputs and converts them into an output to the next layer. There are different activation functions depending on the data being analysed. If the data is not linearly separable a non-linear approach is required. In the present study, a sigmoidal activation is used. This function is given in Equation 5.1. This scales the weighted inputs to between 0 and 1 using the sigmoidal function plotted in Figure 5.6.

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

Equation 5.1 *Sigmoidal activation function this function is plotted in Figure 5.6. Where x is the data input.*

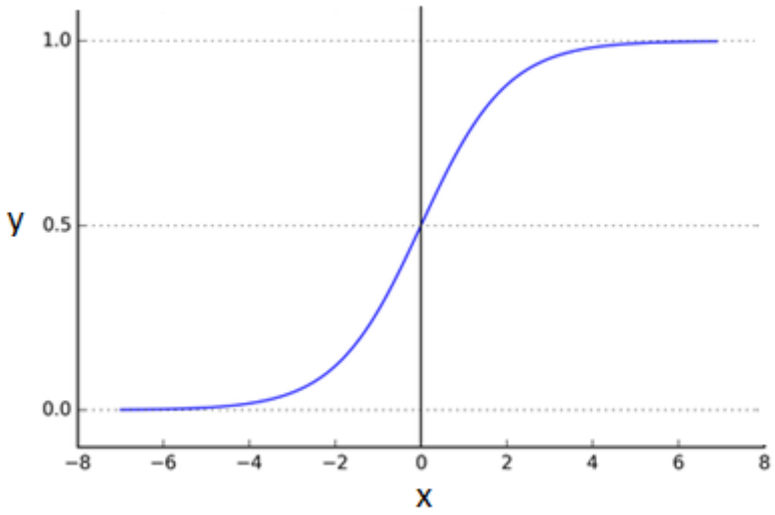


Figure 5.6 *Illustration of the Sigmoidal function. This function scales all inputs (x) to values (y) between 0 and 1 using the Sigmoidal function plotted here.*

5.3.1.3 Output function

The output function takes the output from the final hidden layer and assigns that dataset to a class. This study made use of a SoftMax function. This function finds the probability that the outputted data belongs to a particular class. It does this by calculating the probabilities of the exponentials of the input data. This function works on the assumption that a particular dataset can only correspond to one class therefore the data is assigned to the class to which it has the highest probability of being part of. As this is the case for the problem in this study this is an appropriate output function. The SoftMax layer has the same number of neurons as the output layer.

5.3.1.4 Performance function

In section 5.3.1.1 it was discussed that the training function seeks to minimise the loss and therefore produce the nearest approximation to the correct result. Once this has been applied and the network trained, the effectiveness of the network as a whole must be checked. This is done using a performance function. These functions compare the classifications produced by a network with the correct classification. The training and performance functions are very similar

in that they are assessing loss, however, differ in the loss they are assessing. The training function looks to minimise how far from the correct answer the network is and then updates the hidden layer weights to get a closer result. The performance function instead assesses how often the result is correct rather than how far from the truth it is.

In the machine learning technique developed and described herewith, cross-entropy loss is used to as the performance function. This method is used in a situation, where the output is a probability between 0 and 1. The cross-entropy loss function calculates how far from the ground truth this probability is, the further it is the worse the result. The perfect output from this would be 0. By using this performance function the overall performance of the network can be monitored.

The performance of the network can also be represented visually as a confusion matrix which compares the predicted class with the actual class showing the overall performance. These will be used in the present study to present the results of the effectiveness of the developed machine learning algorithm.

5.3.1.5 Training stops

At some point the training of the NN must stop. This is ideally when the lowest loss is achieved. However, there is a balance to be met between time, processing requirements and accuracy. In many cases, a close approximation to the answer will achieve the desired result. Therefore, there is little point in extending training beyond this point. It is not always possible to achieve 100% accuracy of a network, whether that is because the network architecture is not suitable or there is too much overlap between categories. In these cases, should training be left to achieve a required minimum loss, the training may continue uninterrupted with little to no change in accuracy as the loss ranges around the ideal.

To control the training stop points are defined. Once one or more of these stops are met, depending on the settings, the training of the NN will stop with the last iteration of the network being outputted. The user can then analyse the result to ensure that it is of a desirable accuracy and that it has not overfitted to the data. If the NN has not achieved the desired accuracy running the training function again may achieve a different result as the resulting NN will be different from the original.

The training stops used in this study are shown in Table 5.2. The number of validation checks applied to a network and the maximum error can both be controlled by the user depending on the performance requirements. When the desired performance has been achieved, the network is deemed validated and training is concluded. As this may not be achieved in all cases, the maximum number of epochs is also set to ensure the training will time out if it is not converging to a satisfactory loss.

Table 5.2 *Table giving the user defined training stops used for the neural networks trained using the RMS and FFT data. The maximum Epochs is the maximum runs through the network during training. The target loss is the target for the loss function. The minimum performance is the minimum gradient of the performance function and the validation checks is the number of performance checks that must be passed.*

Maximum Epochs	1000
Target loss	0
Minimum Performance Gradient	$1e^{-6}$
Validation checks	6

5.3.2 Deep learning pattern recognition neural network

Adding more hidden layers to a neural network creates a deep learning neural network. The extra hidden layers add the potential for more dimensions to be analysed by the neural network and therefore more features to be extracted. It is however a trade-off with processing time and computational requirements. In the present study, the effect of adding further hidden layers to the pattern recognition network applied to the RMS and FFT data was tested. The network

developed had the same properties as the shallow network described previously, with the main difference being only in the number of hidden layers applied.

5.3.3 Image recognition neural network

The networks described so far in this chapter have dealt with the AE data as numerical data. The FFT and RMS data can be plotted as a 2D graph as shown in Figures 4.5 and 4.6 in chapter 4. The CWT data is three dimensional. Therefore, the machine learning method described above is not applicable as data cannot be extracted from defined peaks that represent the overall dataset. The data from the CWT analysis is instead outputted as a spectrogram.

Instead of using a patternet NN, an image recognition NN was developed. As the spectrogram is essentially an image with a mixture of colour intensity and location indicating key features this is an appropriate approach ensuring that the extra dimensionality of this data can be analysed.

The image recognition NN developed here can be described as a deep learning NN as more than one hidden layer is used. For this application there is a clear case for the use of deep learning NN as opposed to a shallow network due to the extra dimensionality of the data. In the image recognition NN the hidden layers are not all the same, instead using a number of different layers which simplify and classify the data in different ways.

In the NN developed in the present study, the NN is made up of 8 different types of layers with several repeating leading to a 15 layer NN. The 8 Layers include, an image input layer, convolution 2D Layers, max pooling 2D layers, ReLU layers, average pooling 2D layer, fully connected layers, SoftMax layers and a classification layer. The order of these is shown in Figure 5.7. The maximum number of epochs for this network was set to 10.

1. Image Input Layer
2. Convolution 2D Layer
3. Max Pooling Layer
4. ReLU Layer
5. Convolution 2D Layer
6. ReLU Layer
7. Average Pooling 2D Layer
8. Convolution 2D Layer
9. ReLU Layer
10. Average Pooling 2D Layer
11. Convolution 2D Layer
12. ReLU Layer
13. Fully Connected Layer
14. SoftMax Layer
15. Classification Layer

Figure 5.7 *Ordered list of the layers used to define the architecture of the image recognition neural network developed for analysing the CWT analysed data.*

The image input layer takes the input image, which is the spectrogram, and inputs it into the NN defining its size as the number of colour channels used in the image, in this particular case 3 as it is an RGB image.

The Convolution 2D Layers apply a sliding convolution filter to the input. These are filters of a defined size and move through the input horizontally and vertically calculating the dot product of the weights and the input as well as adding a bias factor. To avoid the interactions of these filters with the edge of the image thereby losing data, padding is added around the image.

The max pooling 2D layer down samples the input from the previous layer by dividing it into rectangle pooling regions and computing the maximum value in each region. This is not applied to each value but strides through the data by a user set amount.

The ReLU layer or a rectified linear unit layer. This layer removes all negative values from the inputted data by defaulting these values to 0. This is applied to every element of the data. This is similar to the application of the sigmoidal function used for the RMS and FFT neural networks. The equation used is given in Equation 5.2, and the plot of this function is shown in Figure 5.8.

$$f(x) = \max(0, x)$$

Equation 5.2 *ReLU equation used to remove negative values from the input dataset.*

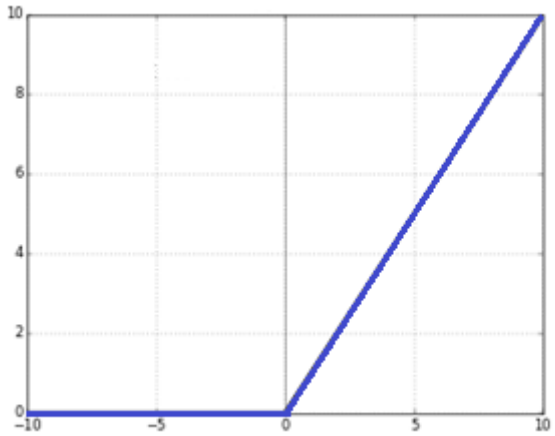


Figure 5.8 *Illustration of the ReLU function applied in the ReLU Layer. This function defaults all negative values to 0.*

The average pooling 2D layer is similar to the max pooling 2D layer described above. It differs in that it calculates the average value of each region rather than the maximum value that is calculated when using the max pooling 2D layer.

The layers described so far are in essence ways of simplifying the data inputted to the network, so only the most prominent or unique features are left within that data. The final layers are used to bring together this simplification to categorize the data. The first of these, the fully connected layer, multiplies the input by a weight matrix and then adds a bias vector in a similar way to the weights and biases used in the NN developed for the RMS and FFT data.

The SoftMax layer is used to normalise the output of the network to a probability distribution over all predicted classes. It takes all the values in the inputted vector and outputs them as a probability distribution, where all the values add up to one. This is the same approach used to classify the data in the RMS and FFT NN. As with these networks the final classification step uses cross-entropy loss.

5.4 Results

The data processed to develop the machine learning techniques below is split into a training and test dataset. The training dataset will act as the known/seen data and the test data as a labelled unseen dataset. The same dataset is used for all the machine learning techniques developed here. The split of the data into these datasets is shown in Table 5.3 and is approximately a 80/20 training to test data split.

Table 5.3 *Table showing the data split between the training a test dataset for each class of severity. The data is split to approximately 80/20 training to test data split.*

Descriptive label	**Class**	**Training**	**Test**
Moderate	3	175	43
Heavy	4	807	201
Severe	5	736	183
Very Severe	6	490	122
Total	**-**	**2208**	**549**

It is not expected that the networks developed here will reach 100% accuracy. This is due to the way in which the data is classified. As has been discussed the data has been grouped into categories. Whilst these are based off a classification system currently in use on the network, they are arbitrary in nature. Each length category is of a set length and there is no overlap. Crack growth does not follow this precise categorization instead growing gradually.

This presents a problem when applying classification into categories as is developed here. Those crack lengths which lie in the middle of the category are easily distinguishable from the

next category as they will be significantly different in length, and therefore if they can be classified by their AE signal this too should be significantly different. The problem lies at the boundary of these categories. At these boundaries the crack length and therefore corresponding AE signal will not be significantly different and therefore confusion may arise. This is the primary reason 100% accuracy is not expected. Other source of accuracy loss may be due to background noise of novel AE signals caused by untrained scenarios.

The performance of the developed NN are plotted using confusion matrixes for example figure 5.9. These present visually the difference between the predicted and actual class. Using this comparison, the overall accuracy of the network is shown. The confusion matrix also gives an idea of which classes the NN are confusing and struggling to classify.

Output class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
4	**0** 0.0%	**0** 0.0%	**145** 6.6%	**668** 30.3%	**522** 23.6%	**285** 12.9%	41.2% 58.8%
5	**0** 0.0%	**0** 0.0%	**24** 1.1%	**124** 5.6%	**183** 8.3%	**165** 7.5%	36.9% 63.1%
6	**0** 0.0%	**0** 0.0%	**6** 0.3%	**15** 0.7%	**31** 1.4%	**40** 1.8%	43.5% 56.5%
	NaN% NaN%	NaN% NaN%	0.0% 100%	82.8% 17.2%	24.9% 75.1%	8.2% 91.8%	40.4% 59.6%

Target class

Figure 5.9 *Example of a confusion matrix used to display the performance of the NNs.*

The confusion matrix is arranged with each row being an output class and each column being the target or ground truth class for that data. Each outputted data set is therefore placed into the cell corresponding to the output and target class for that data set. Where these are the same and the network has therefore classified correctly the cells are coloured green. When this is not the case the cells are coloured red. In each cell the number of data sets in that cell are represented

by the first number in bold and the percentage of the overall data set which that represents is the percentage below this in the cell. The cells containing coloured text represent the percentage of correct (green text) and incorrect (red text) is each output and target class. The overall accuracy is represented in the bottom right-hand cell.

For all confusion matrixes displayed here classes 1 and 2 are displayed however with no results. This is because there was no available data for these classes. They are still displayed as the ability to process them has been programmed into the machine learning algorithm to future proof them for further data, as well as avoiding confusion when comparing to the classes defined in Table 5.1.

In this thesis it has been stated that the NN accuracy can change when the same dataset and network architecture are used and therefore to achieve the highest performance/accuracy a NN architecture should be trained multiple times until the highest or desired accuracy is achieved. In the results presented here only the highest accuracy results from training the NN developed are given. In all cases multiple runs of the same architecture and data yielded little difference in accuracy as the networks had already been tailored for use on the datasets used in this study.

5.4.1 RMS

5.4.1.1 RMS Shallow Neural Network

The first NN developed in this study for the analysis for the RMS processed data was a shallow Neural Network. Figure 5.10 shows the confusion matrix of the accuracy of the RMS Shallow Neural Network developed in this study following training and testing of the network on known/seen data.

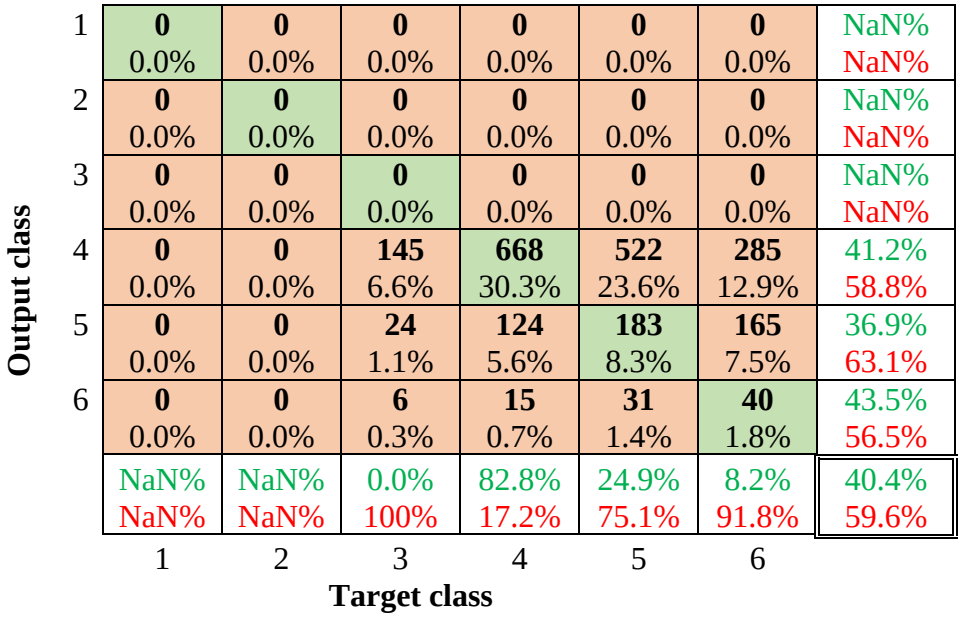


Figure 5.10 *Confusion matrix representing showing the performance of classifying known data (the training set) by the shallow NN trained using the RMS processed data. An overall accuracy of 40.4% is achieved, shown as the green percentage in the bottom right of the confusion matrix.*

The overall accuracy of the network, measured as a percentage of the time the network correctly classifies data in the correct class, is given as the percentage in green in the bottom right-hand cell of the confusion matrix in Figure 5.10. In this case the accuracy of the RMS shallow neural network is 40.4%.

This is a low result in terms of accuracy. For use under real world conditions this network work would not be suitable for working alone to monitor a network. This is particularly true as this is the accuracy on known data that the network has been specifically trained to. On unknown data this accuracy will drop further. Within the confusion matrix, there are some results that stand out as being particularly high in comparison to the other results. For example, where a dataset has a target class of 4 the network correctly classifies this data in 82.8% of cases. This is, however, contrasted with target class 3, where none of this class were correctly classified.

Within this confusion matrix, there is a clear indication of overfitting. This is particularly clear for class 4. In this case as already discussed 82.8% of datasets that had a target class of 4 were correctly classified as this. However, of those classed by the NN as class 4, only 41.2% should have been categorised in this category. In fact, 73.3% of the total dataset was classed as class 4, if the NN was correctly trained this value should have been 36.5%. The overfitting of the NN to this class may be caused by the larger number of samples avaible to train the network from in this class compared with the other possible classes.

Figure 5.10 showed the results of the trained network on known data. This should give high levels of accuracy should the network have been trained affectively. The trained data must then be tested on an unknown dataset to confirm that it has not been overtrained to the training dataset. From the results of the trained data (Figure 5.10) it is expected that the accuracy of the network on unknown data will be low and have a bias to classifying data sets as class 4.

The results from the test data are given in the confusion matrix in Figure 5.11.

Output class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
4	**0** 0.0%	**0** 0.0%	**39** 7.1%	**148** 27.0%	**127** 23.1%	**65** 11.8%	39.1% 60.9%
5	**0** 0.0%	**0** 0.0%	**4** 0.7%	**47** 8.6%	**43** 7.8%	**44** 8.0%	31.2% 68.8%
6	**0** 0.0%	**0** 0.0%	**0** 0.0%	**6** 1.1%	**13** 2.4%	**13** 2.4%	40.6% 59.4%
	NaN% NaN%	NaN% NaN%	0.0% 100%	73.6% 26.4%	23.5% 76.5%	10.7% 89.3%	37.2% 62.8%
	1	2	3	4	5	6	

Target class

Figure 5.11 *Confusion matrix representing the performance of the shallow NN trained using the RMS processed data when applied to an unseen data set (the test data set). The accuracy of the network is shown as 37.2% this is represented as the green percentage in the bottom right of the confusion matrix.*

The confusion Matrix in Figure 5.11 confirms the expected result that the accuracy of the NN on a test dataset on an unknown dataset (37.2%) will be lower than when test on a known dataset (40.2%). Figure 5.11 also confirms the overfitting of the NN to the class 4 category with 73.6% of those with a target class of 4 categorised correctly. However, of those datasets classed as class 4 only 39.1% were correctly classified. The trend for the other data classes is also repeated, with no class 3 datasets accurately classified. Classes 5 and 6 were classified correctly, in 23.5% and 10.7% respectively, in comparison with 24.9% and 8.2% correctly classified in the known dataset.

The higher number of samples avaible to train the NN on in class 4 in comparison with the other classes will have some effect on the overall accuracy of the NN. The overall low accuracy of the NN trained on the RMS data implies that there is another factor influencing the accuracy of this classifier. The structure and hyperparameters controlling the training of the NN have a large impact on the way in which the NN is taught and therefore its overall accuracy on a given dataset. Many of these parameters are set primarily though trial and error to achieve the highest possible accuracy. The author has conducted trials to optimise the accuracy of the given dataset. The results presented herewith show the highest accuracy level that was achieved. That is not to say that there is not a NN architecture that cannot achieve better results, but rather that the structures and parameters used in this study to create a shallow NN were unable to give a better result.

5.4.1.2 RMS Deep Neural Network

It may be that the shallow NN does not have enough dimensionality to classify the RMS data. A potential approach in combating this is to use a Deep NN. By adding further hidden layers more complex data can be analysed and indicative features extracted. In this section, the ability of a Deep NN to combat the shortfalls of the shallow network is discussed. As it has been

previously described, the Deep NN developed herewith, has the same structure as the shallow network. However, the number of hidden layers is increased.

The accuracy of the Deep NN developed for the RMS data when trained on known data is shown in Figure 5.12. As with the confusion matrix shown for the shallow NN, the accuracy of the trained network is shown as the green percentage in the bottom right of the matrix. For this NN on trained data the accuracy of the network is 44.1%. Whilst this is higher than the 40.4% accuracy of the shallow NN in Figure 5.10, considering this NN has been trained on the data that is classified to produce Figure 5.12 the accuracy is still very low for this data processing method.

Output class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
4	**0** 0.0%	**0** 0.0%	**120** 5.4%	**577** 26.1%	**351** 15.9%	**237** 10.7%	44.9% 55.1%
5	**0** 0.0%	**0** 0.0%	**42** 1.9%	**173** 7.8%	**298** 13.5%	**155** 7.0%	44.6% 55.4%
6	**0** 0.0%	**0** 0.0%	**13** 0.6%	**57** 2.6%	**87** 3.9%	**98** 4.4%	38.4% 61.6%
	NaN% NaN%	NaN% NaN%	0.0% 100%	71.5% 28.5%	40.5% 59.5%	20.0% 80.0%	44.1% 55.9%

Target class

Figure 5.12 *Confusion matrix representing showing the performance of classifying known data (the training set) by the deep NN trained using the RMS processed data. An overall accuracy of 44.1% is achieved, shown as the green percentage in the bottom right of the confusion matrix.*

Figure 5.12 also shows that the deep NN has not solved the issue of overfitting of the network to class 4 as it was seen with the shallow NN. As with the shallow NN, the deep NN showed the greatest accuracy in classifying this class of data, with those datasets with a target class of 4 being correctly classified in 71.5% of cases. For the shallow NN, it was noted that 73.3% of the total available data were classified as class 4, when in fact, this class only makes up 36.5% of the total dataset. From Figure 5.12 it can be calculated that for this network 58.19% of the data is classed as class 4 with the data used the same as that used for the shallow NN. From this, it can be seen that the Deep NN has begun to address the overfitting of the NN to the class 4 data.

As with the shallow NN, it is expected from the Deep NN, that the overall accuracy of the network will be lower when tested on an unknown dataset. The results of testing on the known dataset is shown in Figure 5.13.

Output class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
4	**0** 0.0%	**0** 0.0%	**36** 6.6%	**128** 23.3%	**102** 18.6%	**40** 7.3%	41.8% 58.2%
5	**0** 0.0%	**0** 0.0%	**3** 0.5%	**54** 9.8%	**55** 10.0%	**41** 7.5%	35.9% 64.1%
6	**0** 0.0%	**0** 0.0%	**4** 0.7%	**19** 3.5%	**26** 4.7%	**41** 7.5%	45.6% 54.4%
	NaN% NaN%	NaN% NaN%	0.0% 100%	63.7% 36.3%	30.1% 69.9%	33.6% 66.4%	40.8% 59.2%
	1	2	3	4	5	6	

Target class

Figure 5.13 *Confusion matrix representing the performance of the deep NN trained using the RMS processed data when applied to an unseen data set (the test data set). The accuracy of the network is shown as 40.8%, this is represented as the green percentage in the bottom right of the confusion matrix.*

From Figure 5.13 it is seen that, as expected, the accuracy of the NN, when tested on unseen RMS data is lower than that on known data as was shown in Figure 5.12. For the unseen test dataset from Figure 5.13 the overall accuracy of the trained deep NN is 40.8%.

Overall, the test dataset shows a similar response to the training dataset in terms of bias towards class 4 data, although, as expected this is lower than for the known trained dataset showing an accuracy for classification of this class of 63.7% compared with 71.5% of the known data. Interestingly, the accuracy of the classification of target class 6 data increases from 20% in the training dataset to 33.6% in the test dataset.

As with the shallow dataset the class 3 target class data is never classified correctly. As in training this class is never classified correctly it will not be possible for the NN to identify this class from an unknown dataset as it will not have learnt any features associated with it.

The inaccuracy of the class 3 dataset may be due to having a smaller dataset to work from in comparison with the other classes of data, or no unique features are extractable from the data using the RMS processing route. It may be that the signals associated with early crack growth are not distinct from the background noise, therefore the classifier is unable to isolate these signals.

5.4.2 FFT

5.4.2.1 FFT Shallow Neural Network

As with the RMS results the first results obtained from the FFT shallow NN are the effectiveness of that network on the known training data. This result is displayed as a confusion matrix in Figure 5.14. It can be seen that the overall effectiveness of the NN on known data, the green percentage in the bottom right of the confusion matrix, is 96.1%. This is a very high level of accuracy particularly considering the problems already discussed with respect the broad classification bands used.

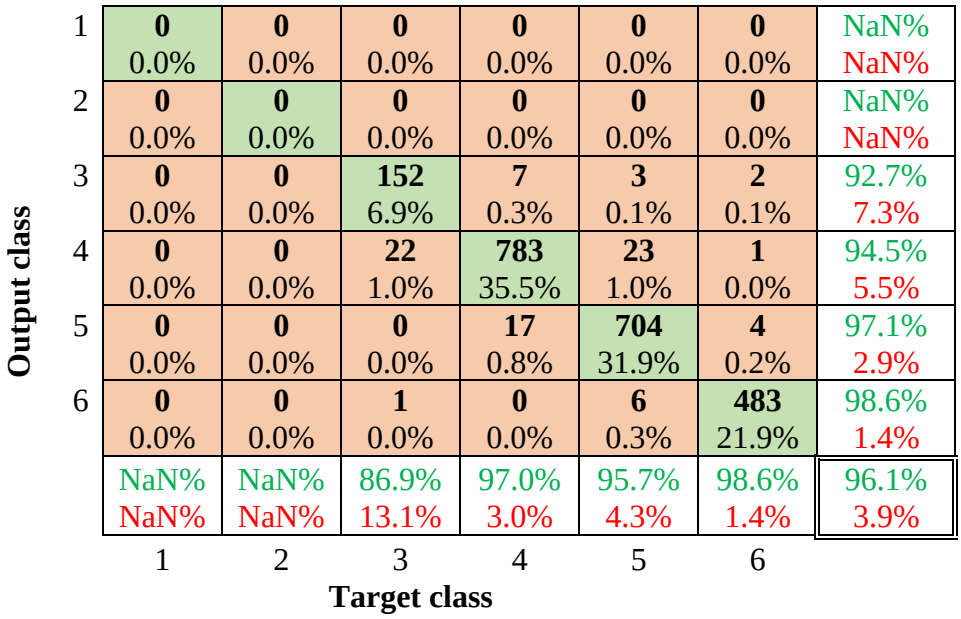


Figure 5.14 *Confusion matrix representing showing the performance of classifying known data (the training set) by the shallow NN trained using the FFT processed data. An overall accuracy of 96.1% is achieved, shown as the green percentage in the bottom right of the confusion matrix.*

High accuracy is observed across all the classes with NN most accurate is classifying class 6, 98.6% and least for class 3, 86.9%. The lower accuracy seen for class 3 may be partially affected but the number of samples used to train this class. At 175 samples this is significantly lower than the 490-807 used for samples 4-6. Even with this lower number of classes, at 86.9% accuracy this still indicates an effective classification system.

It can be seen from Figure 5.14 that 86 data point (3.9% of the training dataset) were inaccurately classified in the final trained network. 53.49% of these inaccurate results were classed inaccurately as class 4. With the majority of these classed as either of the adjacent severity classifications. It was expected that there would be a level of confusion between the severity classes adjacent to each other due to the arbitrary nature of the classes. The fact that the primary confusion is between the adjacent categories would suggest that this is a leading

factor in the inaccuracy of the network. As class 4 is the largest class of data it may be that the network is overfitting to this class leading to over classification to this class.

As already discussed, class 3 achieved the lowest accuracy at 86.9%, as with misclassification of class 4 this was primarily to an adjacent class of severity. The same trend is seen to a lesser effect, in that fewer data points were misclassed, in all the other categories.

With such a high level of accuracy achieved on the known training dataset, even with the expected overlap of results between adjacent categories of data, the network must be tested using a labelled but unseen dataset. This will ensure that the training of the network on the training dataset has not led to overfitting to the training dataset. To evaluate this, the trained network was trialled using the test dataset. The confusion matrix showing the result of this test dataset is shown in Figure 5.15.

Output class \ Target class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**22** 4.0%	**3** 0.5%	**1** 0.2%	**1** 0.2%	81.5% 18.5%
4	**0** 0.0%	**0** 0.0%	**20** 3.6%	**179** 32.6%	**12** 2.2%	**1** 0.2%	84.4% 15.6%
5	**0** 0.0%	**0** 0.0%	**0** 0.0%	**19** 3.5%	**164** 29.9%	**3** 0.5%	88.2% 11.8%
6	**0** 0.0%	**0** 0.0%	**1** 0.2%	**0** 0.0%	**6** 1.1%	**117** 21.3%	94.4% 5.6%
	NaN% NaN%	NaN% NaN%	51.2% 48.8%	89.1% 10.9%	95.7% 4.3%	95.9% 4.1%	87.8% 12.2%

Figure 5.15 *Confusion matrix representing the performance of the shallow NN trained using the FFT processed data when applied to an unseen data set (the test data set). The accuracy of the network is shown as 87.8%, this is represented as the green percentage in the bottom right of the confusion matrix.*

From Figure 5.15, it can be observed that whilst the overall accuracy has dropped from the 96.1% seen on the training data set to 87.8% for the test dataset. This drop in data is not unexpected as the NN is trained and tailored to the training dataset. Therefore, some loss in performance is expected when applied to an unknown dataset. This is partially due to the rigidity of machine learning techniques which are unable to adapt to novelty. The high performance on the test data does indicate that the NN has not overfitted to the training data set and is therefore capable of identifying the crack length from unknown dataset.

The same trend is observed in Figure 5.15 as in Figure 5.13, in that where the NN misclassifies the data it is primarily to an adjacent category or severity. In the case of the test dataset, shown in Figure 5.15, 67 datasets were misclassified making up 12.2% of the dataset. It would follow that the data classes that were most often misclassed in the training dataset would also be misclassed in testing. This is the case with target class 3 achieving the lowest accuracy at 51.2% and class 6 achieving the highest at 95.9% accuracy. Whilst the overall accuracy for the test data is high, at 51.2% accuracy the classification of class 3 is very low and is unlikely to be deemed acceptable for network use. As with the training data set the number of samples available for training in this category is low. A lower number of representative samples to train from will lead to lower accuracy for the network. It may therefore, be possible to increase this accuracy by increasing the number of samples in this class.

In contrast to this the class with the highest number of data samples, class 4 has the highest number of datasets misclassed as this class. It was noted when discussing the training data that as this is the largest class, it may be skewing the training to overfit to this class. As this pattern of misclassification is repeated in the test data, this may partially be the cause.

5.4.2.2 FFT Deep Neural Network

Through the application of a deep NN to the FFT dataset, it may be possible to further increase the accuracy of the network due to the extra dimensionality of these networks. It is worth noting that a high level of accuracy has already been achieved for the shallow NN as discussed in section 5.4.2.1. Therefore, it is noted that 100% accuracy is not achievable due to the constraints of the class boundaries, and it is expected that any increase in accuracy will be marginal.

The accuracy of the deep NN when applied to the known training dataset for the FFT processed data is shown in Figure 5.16. The overall accuracy of the deep NN is displayed in Figure 5.16 is 91.1%.

Output class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**141** 6.4%	**5** 0.2%	**1** 0.0%	**1** 0.0%	95.3% 4.7%
4	**0** 0.0%	**0** 0.0%	**30** 1.4%	**735** 33.3%	**59** 2.7%	**1** 0.0%	89.1% 10.9%
5	**0** 0.0%	**0** 0.0%	**1** 0.0%	**66** 3.0%	**661** 29.9%	**14** 0.6%	89.1% 10.9%
6	**0** 0.0%	**0** 0.0%	**3** 0.1%	**1** 0.0%	**15** 0.7%	**474** 21.5	96.1% 3.9%
	NaN% NaN%	NaN% NaN%	80.6% 19.4%	91.1% 8.9%	89.8% 10.2%	96.7% 3.3%	91.1% 8.9%

Target class

Figure 5.16 *Confusion matrix representing showing the performance of classifying known data (the training set) by the deep NN trained using the FFT processed data. An overall accuracy of 91.1% is achieved, shown as the green percentage in the bottom right of the confusion matrix.*

Whilst the accuracy is lower than that for the shallow network, which achieved 96.1% accuracy of the training dataset, it is still a high. As with the shallow NN the accuracy of classification was highest with class 6 at 96.7% accuracy. The lowest accuracy of classification was the

classification of data to target class 3 at 80.6%. As it has been discussed already, this is the smallest data class. Therefore, the network may be negatively biased in respect to this class. Whilst the classification of the class 3 data to the correct class is low, where data is classed by the deep NN as class 3 the accuracy of this result is high showing an accuracy of 95.3%. This is second only to the when data is classified as class 6, which has an accuracy of 96.1%. Therefore, from these results a user could have high confidence that should a sample be classified as class 3 that it is likely to be correct.

With the high accuracy shown for the training dataset there is a risk that the NN is overfitting to the training data. Therefore, it is not suitable for use on unknown data. The result of testing the FFT deep NN is shown in Figure 5.17.

Output class \ Target class	1	2	3	4	5	6	
1	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
2	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
3	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**0** 0.0%	NaN% NaN%
4	**0** 0.0%	**0** 0.0%	**33** 6.0%	**134** 24.4%	**106** 19.3%	**47** 8.6%	41.9% 58.1%
5	**0** 0.0%	**0** 0.0%	**9** 1.6%	**51** 9.3%	**61** 11.1%	**41** 7.5%	37.7% 62.3%
6	**0** 0.0%	**0** 0.0%	**1** 0.2%	**16** 2.9%	**16** 2.9%	**34** 6.2%	50.7% 49.3%
	NaN% NaN%	NaN% NaN%	0.0% 100%	66.7% 33.3%	33.3% 66.7%	27.9% 72.1%	41.7% 58.3%

Figure 5.17 *Confusion matrix representing the performance of the deep NN trained using the FFT processed data when applied to an unseen data set (the test data set). The accuracy of the network is shown as 41.7% this is represented as the green percentage in the bottom right of the confusion matrix.*

From Figure 5.17 the accuracy of the deep NN tested on the unseen test data set is 41.7%. This is significantly lower than the 91.1% accuracy of the network on the training dataset. This

strongly implies that the NN has overfitted to the training dataset. Therefore, it is not able to accurately analyse unseen data. As such, a dramatic shift in accuracy was not seen between the training and test data for the FFT shallow NN shown in Figures 5.14 and 5.15 respectively. This difference between the shallow and deep NN, where the only difference in the architecture is the number of hidden layers, implies that the extra dimensionality of the NN provided by the deep neural network is not leading to greater understanding of the data class feature but rather overfitting to a particular set of data.

What is particularly interesting about the test data results in Figure 5.17 are the classes that are the least accurate in classifying. For the training dataset in Figure 5.16 and the FFT shallow network class 6 showed the highest level of accuracy in terms of this target class being correctly classified. For the test data displayed in Figure 5.17 this is one of the least accurate target classes at 27.9% accuracy second only to class 3 which shows 0% accuracy for this target class whereas on the training dataset it showed a level of accuracy of 80.6%. The complete loss of accuracy for this target class and that no data was classified as this class even inaccurately, strongly suggests that the deep NN as overfitted to the training dataset as when using the training dataset, it can very accurately classify this class but when given unseen data of the same class is unable to classify it.

5.4.3 CWT

In the development of the CWT NN a major problem was encountered. The processing of the data and subsequent training of both the shallow and deep NN for the RMS and FFT approaches required the use of high-performance computing capability and parallel processing within this. With this capability these methods were able to process and train data relatively rapidly.

The CWT processing of the data is relatively computationally expensive in comparison with the FFT and RMS methods. The use of images as the input to the NN is also more computationally expensive than that of the pattern recognition networks used in the FFT and

RMS approaches as is the image recognition network required to process them. Even with the high-performance computational capability available to the author, it was not possible to effectively analyse the data using CWT at scale and subsequently train a NN. To give an idea of the time difference between the time required for the RMS and FFT processing routes in comparison to CWT, in testing the time required to process the entire training dataset for RMS or FFT only a single CWT data sample was processed.

Given enough time and higher processing levels it would be possible to process the data and train a network using this data. Whilst this would produce a trained network capable of analysing data in this way the high levels of computational processing power required would render this method unusable on the large datasets that would be generated under real world conditions. Image recognition NNs are widely used for applications of low processing systems such as face recognition of mobile phones. It is therefore, the mass processing and handling of the data using CWT methods that are a challenge to the use of this method. If a way can be found to reduce the computational load capability required to process the data using CWT this processing method is expected to lead to high levels of accuracy and the creation of a more robust classification system than that of FFT and RMS.

5.5 Discussion

The same training and test dataset were used for all the machine learning approaches presented herewith. Therefore, the accuracy of these networks can be directly compared and the optimal approach from those studied can be determined. Overall, there was a large range of accuracies displayed from 37.2% for the RMS shallow NN (Figure 5.11), to 87.8% accuracy for the FFT shallow NN (Figure 5.15) whilst the CWT approach was found to be to computationally very expensive.

Beyond the overall accuracy of each network the individual accuracies of each class are also of interest as they give an idea of how the NN are learning. That is whether it is overfitting and whether certain classes are harder for the network to identify. Overfitting is mostly clearly seen where the network shows a high level of accuracy on the training/known dataset but shows a very low accuracy in the test dataset. If across all networks and training and test datasets the networks are performing poorly for a particular class of data, it may suggest that that dataset is particularly difficult to differentiate either from the other datasets or in the case of the data used in this study from the background noise.

The clearest indication of overfitting in these results was the FFT deep NN (Figures 5.16 and 5.17). For this network the training data produced an overall accuracy of 91.1% in comparison with this the test data showed an accuracy of 41.7%. This is a large drop in accuracy of the data from a very high level to a potentially unusable level of accuracy on unseen data. The shallow FFT NN developed shows a similarly high accuracy on known data at 96.1% (Figure 5.14). However, unlike the deep network the overall accuracy does not show such a dramatic drop in accuracy, only decreasing to 87.8% (Figure 5.15). This indicates that the decrease in accuracy from the training to the test dataset is not inherent to the data and therefore the data processing method, but rather the network architecture. By increasing the complexity of the network to a deep neural network the features that can be identified are also increased. The risk of this is

that features that are indicative of only a small subset of examples not to the global examples of data are learnt. The overfitting in only the deep NN in this case implies that this is what has led to the overfitting to the training dataset.

Within the FFT deep NN class 3 data was particularly affected by this drop in accuracy falling from 80.6% target class accuracy for the training data (Figure 5.16) to 0% target class accuracy for the test data (Figure 5.17). Both the RMS dataset (Figures 5.10-5.13) struggled to classify accurately this target class with both NNs unable to classify this data either for the training and test dataset. This would imply that there was something inherent in that data class that meant that the NNs were unable to identify unique indicative feature for this class. This may be due to the smaller dataset in comparison with the other classes of data or potentially difficulties in extracting background noise for the quieter earlier crack growth. From the results in chapter 3, it is known that as crack growth accelerates, the energy of the AE signal increases. It is therefore easier to differentiate useful AE signals from the background noise. From this it is not unexpected that the NN may have difficulty in classifying earlier crack growth stages in comparison to latter stages.

In contrast to the other NN presented herewith, the FFT shallow network (Figures 5.14 and 5.15) did successfully classify the target class 3 data. Nonetheless, there is a significant drop in accuracy from the training set at 86.9% target class accuracy to the test set at 51.2% target class accuracy. The fact that this NN is able to classify the data more often than not even on unseen data implies that it is possible to separate this class using indicative features from the background noise.

It is worth noting that whilst the percentage accuracy implies a success in differentiating this class, when the number of datasets classified is looked at it can be seen that there is a low number of samples available for the NN to be tested on. Therefore, whilst 51.2% accuracy

seems potentially promising the difference between correctly and incorrectly classed data for this target class is 1, with 22 correctly classed and 21 incorrectly. This points towards one of the potential challenges already discussed for this dataset the small number of samples for this class particularly in comparison with the other classes of data. It was discussed in chapter 4 that for a NN to be accurately trained it must be trained on a large quantity of data covering all possible scenarios. The limited data avaible for training the NN in this study are therefore likely to have an impact on the accuracy of these networks and their ability to classify data, in particular those data classes that are potentially challenging to classify.

The impacts of the number of data sets in each class and the potentially higher signal to noise ratio of the higher severity classes will have two different effects of the hierarchy of accuracy of each class. For the increasing severity the accuracy would simply increase in order of increasing severity. The number of datasets for each data class are shown in Table 5.3. From the number of data sets in each class, assuming that a higher number of samples would lead to a greater accuracy, the class should be ranked from least to most accurate as; class 3, class 6, class 5, class 4.

Both challenges indicate that class 3 should be the most difficult class to classify and hence, the least accurate. As discussed already, this is the case across all the NN tested. It is thus, difficult to ascertain, which of these challenges is having the greater effect on the NN for this class. It does lend an explanation of why all the NNs struggled in particular to classify this data.

For the other data classes the difference in the number of samples and the increase in signal to noise ratio run in opposite direction. It would therefore be expected that there may be some levelling out giving overall good accuracy across all these classes. For the FFT shallow NN (Figures 5.14 and 5.15) this is true to an extent with a variation of 6.8% between the target class accuracy of class 4-6. In all other NNs developed here Class 4 showed a significantly

higher accuracy in comparison to target classes 5 and 6. As this class also has the highest number of data samples it implies that these networks are heavily influenced by the number of samples available.

As a NN implemented into a real-world application should have been trained with a very large dataset the effect of this challenge should be reduced. It may therefore be that those NNs that have a bias towards the accuracy of class 4 may become more accurate with an increase in the number of data sets available. That said if these networks are being heavily influenced by the number amount of data available, great care would need to be applied to ensure that this effect is mitigated. Should this not be achieved, the future use of this NN would be limited and the results treated with extreme caution.

One of the aspects investigated in this chapter was whether the use of a deep NN would increase the accuracy of classifying the data by adding extra dimensionality. For the RMS processed data an increase in accuracy was seen from 37.2% for the shallow network (Figure 5.11) to 40.8% for the deep network (Figure 5.13). In this case, the extra hidden layers have allowed the NN to learn the differences more accurately between the different categories. However, even for the deep NN the accuracy remains low being wrong more than it is correct.

In contrast to the RMS processed data, the FFT processed data shows a dramatic decrease in the accuracy of the NN when using a deep NN in comparison with the shallow NN. The shallow NN had an overall accuracy of 87.8% (Figure 5.15) in comparison with 41.7% accuracy for the deep network (Figure 5.17). For the FFT processed data, the shallow NN has a very high accuracy, which is likely to be deemed acceptable especially when considering the shortfalls of the classification system. The accuracy of the deep NN is however, much lower. At below 50% accuracy, it is unlikely to be acceptable for use on a live rail network.

When considering the use of a deep NN over that of a shallow NN it is important to ascertain if there is a benefit provided by using this deeper network, bearing in mind that there will be extra processing requirements as a result. Real world applications of the monitoring system being developed in this study will not have access to HPC technology and therefore will be limited in the NN they can support. In most cases a shallow network is capable of separating a data set into categories, therefore only very complex problems are likely to need deep NNs and their use are not guaranteed to increase accuracy in comparison to a shallow network.

The FFT processed data shows the potential downside of using a deep NN where despite the increased processing requirements a loss in accuracy is observed. As has been discussed this drop in accuracy is likely due to the NN overfitting to the training data. The RMS data on the other hand does show an increase in accuracy with the extra dimensionality offered by the deep NN. As has already been noted this increase in accuracy is low. Therefore, a decision must be made as to whether the increase in accuracy is sufficient to justify the increased processing requirement.

The development of an approach using CWT was proposed to overcome some of the issues faced by the other processing method. For example, the greater dimensional information provided may overcome the issues of signal to noise ratio faced by the other approaches. A big challenge facing the implementation of machine learning techniques into field operations is the computational processing requirement of these approaches. As already discussed, a balance must be met with these techniques between the accuracy and the processing power required. In the RMS data for example a greater accuracy was achieved using a deep NN in comparison to a shallow NN. However, greater processing power is required to achieve this. Thus, a decision must be made as to whether this increase in accuracy is worth the extra processing capability requirements.

For the CWT processing method, the processing requirements available to the author were found to be too high with the available hardware. It therefore poses a serious question, whether this method is suitable for use in this application. Whilst it does have advantages over the other two processing methods as discussed in chapter 4, if the processing requirements are higher than those avaible to the author through a HPC facility, it is unlikely that sufficient system could be implemented across a rail network with a considerable number of sensors and therefore datasets to process.

In this thesis the data from all of the samples was combined and then split randomly into and training and testing dataset. Whilst this does create a dataset of known data and unseen data there is a risk with this approach that the test dataset is not entirely novel. As the training dataset contains data from every sample there are no samples in the test dataset that the algorithm has not seen. Therefore, ideally the data should be split by sample into a samples used for training and those used for testing. In this case the data was randomised and combined to try and combat the limitations in the data set size and the none even spread of data for each class for each sample. That is each sample did not have the same number of data samples for each classification of damage.

Separating out samples to give independent datasets would give more robust results in terms of accuracy of performance and ensuring the modal has not overfitted. It is however not expected in this case that it would lead to a large change in the accuracy of the results given here. As other limitations on the data set such as its overall size and the idealised samples of conditions are lightly to have a greater effect on the results.

5.6 Conclusion

For a machine learning approach to AE monitoring of assets to be implemented in the field, the accuracy of this technique must be high to ensure safety and to overcome the challenges faced by current monitoring methods. It was anticipated from the start of the present study, that the investigation into the machine learning approach presented herewith, will not achieve 100% accuracy, due to the limitations of the arbitrary classification classes. That having been mentioned, it would be expected that even with this limitation a viable machine learning approach would be able to classify the data to a high degree of accuracy.

Three different processing methods were investigated including, RMS, FFT and CWT. Of the three processing methods considered, the CWT method was found to be to computationally heavy whilst both deep and shallow NN were developed and tested on the RMS and FFT dataset.

For the RMS processing route the highest accuracy was achieved when using the deep NN with an overall accuracy of 40.8% compared to 37.2% for the shallow NN. Despite the increased accuracy achieved for the RMS data when using a deep NN it still achieves a lower accuracy NN than that trained for use on FFT processed data. The FFT data achieved 87.8% accuracy when using the shallow NN but did fall short when using the deep NN falling to 41.7% accuracy.

Whilst the RMS data is more accurately classified using a deep NN, the accuracy of this result is still low and at less than 50% accuracy this is unlikely to be acceptable for real world use. The FFT processing route does achieve a high level of accuracy. This is also achieved through the less computationally heavy route of the shallow NN.

The processing requirements for both these techniques were found to be comparable under testing. This would suggest that for this data type and application using the NN developed

herewith, that the FFT processing route would be most appropriate. Within this approach the shallow NN showed the highest accuracy. Therefore, from the NN tested in this study the FFT shallow NN would be most appropriate for the development of a real time structural health monitoring system using AE data.

5.7 Summary

This chapter has presented the routes to the development and the subsequent testing of machine learning and signal processing techniques for the analysis of AE signals generated during the growth of a fatigue crack growth. Three signal processing routes were explored, including, RMS, FFT and CWT. The severity of the fatigue crack for each AE signal was categorised into 4 categories with 2 further categorise defined for future use.

For the RMS and FFT processed data both shallow and deep NN were developed and tested using pattern recognition NN. An image recognition NN was also developed and tested for the analysis of the CWT processed data. However, the computational requirements of this route proved to be too high for the available equipment.

The FFT shallow NN achieved the highest accuracy of all the networks tested achieving an accuracy of 87.8% when tested against an unseen test dataset. A dramatic drop in accuracy was observed when a deep NN was applied. The RMS NN achieved low accuracy results for both the RMS shallow and deep NN achieving 37.2% and 40.8% accuracy respectively. Unlike the FFT networks the RMS deep network achieved a greater accuracy than that of the RMS shallow network although the increase is marginal, and the overall accuracy remains low.

It is concluded from these results that of the approaches tested the FFT shallow NN is the most suitable for further testing as it surpasses all the other networks in accuracy.

Chapter 6

Conclusions and future work

6.1 Conclusions

This work aimed to further the development of a real-time structural health monitoring system for use on rail infrastructure, using Acoustic Emission (AE) techniques. Two main approaches were used: a conventional AE commercial system and a custom-built monitoring system.

The commercial AE system used a Hit Driven Data (HDD) approach focusing on the key AE parameters, Counts, Energy and Duration. The effectiveness of these parameters in monitoring fatigue crack growth in three steels, R220, R260 and cast manganese, was investigated with the aim of rapidly assessing the capability of AE monitoring techniques.

The quantitative accuracy of this system was found to vary with the material being monitored. It was not possible to find a close corelation between any of the AE parameters and the crack growth behaviour of the cast manganese steel. This was not unexpected due to the work hardening behaviour of cast manganese steels that has been reported previously. It would however, present a challenge in the development of an AE monitoring system based on this comparison. It is suggested that a system could be developed that is able to provide a warning of imminent failure based on the trough shaped drop and then rise in AE activity associated with the final stages of fatigue crack growth for this steel.

For both the R220 and R260 steels a correlation was found for all the AE parameters investigated and the fatigue crack growth behaviour of these steels with all parameters showing an upward trend broadly in line with the ΔK cs da/dn plot. This trend was clearer for the R220 steels. From this trend it may be possible to develop a system that is able predict the current

structural health of these steel under fatigue loading using these AE parameters. There is however a large amount of variation in the AE data for both of these steels making any model based on these plots limited in terms of its accuracy.

Whilst the commercial AE system does show that there is a connection between the key AE parameters investigated here, the development of an AE monitoring system based on this approach would be of limited accuracy due to the variability in the AE data in comparison with the crack growth behaviour. Also, the requirement for continuous monitoring to ensure an accurate comparison can be made leads to a loss of redundance in the system. Despite its limitations the commercial system investigated here has shown that there is connection between the AE signals recorded and the corresponding fatigue crack growth behaviour. It also highlights that each steel grade will result in a different AE response to fatigue crack growth and therefore any system developed must ensure that the data acquisition parameters are tailored to the material AE response generated during damage initiation and evolution. This is likely to involve better understanding of the behaviour of dislocations and the events occurring at the crack tip from a microstructural perspective.

The second approach to be investigated in this work was the use of a custom AE monitoring system which records the complete AE waveform and uses signal processing to analyse this. Due to the large quantity of data generated using waveform-based acquisition appropriate machine learning approaches where investigated. This system was developed for the R260 rail steel only due to time limitations.

Three different signal processing methods were investigated, including, Route mean square (RMS), Fast Fourier transformation (FFT) and Continuous wavelet transform (CWT). Deep and shallow Pattern recognition Neural networks (NN) were developed and tested for the RMS and FFT processing routes. An image recognition network was developed for the analysis of

the CWT processed data. However, this proved to be to computationally expensive for implementation with the hardware available to the author.

It was not expected that 100% accuracy would be achieved for any of the networks developed in this work due to limitations in the classification system. Nonetheless high levels of accuracy were achieved still. The highest accuracy was achieved by analysing the FFT processed data using a shallow NN, achieving an accuracy of 87.8% on unseen test data. No benefit was found in using a deep learning method for the analysis of the FFT data. Overall, the RMS processing route led to a lower level of accuracy compared with the FFT processing route with a maximum accuracy of 40.8% achieved. In contrast to the FFT data the higher level of accuracy was achieved using a deep NN, although the increase was marginal.

Despite the limitations of the arbitrary classifications used to train the NN in this work a system has been developed using a combination of a custom-built AE recording system and machine learning techniques capable of accurately predicting the structural health of a material under fatigue loading.

6.2 Future Work

The NN developed in this study was trained on a relatively small data set generated under laboratory conditions for one type of steel. Further work is therefore, needed to build on this work to achieve the full potential of AE monitoring systems.

Machine learning techniques are reliant on the data they are trained on, being unable to adapt themselves for scenarios they are not trained on. The development of a novelty analysis algorithm for use alongside the NN developed here could be used to indicate when such events occur.

A larger data set would increase the reliability of the NN on unknown data, ideally going forward this data would be generated under real world conditions. This would not require a

complete re-working of the NN developed here but rather further training of the NN using operational data. The early stages of this work have begun by both the author and others at the University of Birmingham alongside Network Rail, however a large amount of data across a range of conditions will be required before full implementation of the system on the Rail network. The methods used for this work are described in section 2.3 of chapter 2.

The development of the machine learning approaches focussed on the R260 rail steel. However, as noted from the analysis and evaluation of the commercial system data each material will have a different AE response to fatigue crack growth. Further work is therefore required to train the NN for use on all materials to be monitored. As with moving to real world data this would not necessarily require a complete redevelopment of the network. Transfer learning approaches could instead be used.

It was noted when analysing the commercial system data that changes in the microstructure of the material, such as the presence of inclusions, led to changes in the AE signals recorded and this corresponded to changes in crack growth behaviour. Given a large enough dataset encompassing all the defects expected to be found in the materials to be analysed it may be possible to detect when these changes in crack growth behaviour have occurred. This would provide a greater understanding of the structural health of an asset being monitored. This in turn may lead to the potential development of a digital twin to an asset allowing accurate prediction and monitoring of that asset. This combination of real time monitoring and digital twin analysis is suggested as the ideal system to be developed for the greatest benefit of structural health monitoring of critical rail infrastructure.

References

Aglan, H., & Fateh, M. (2007). Fracture and fatigue crack growth analysis of rail steels. *Journal of Mechanics of Materials and Structures*, *2*(2), 335–346. https://doi.org/10.2140/jomms.2007.2.335

Ahadi, M., & Bakhtiar, M. S. (2010). Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to Acoustic Emission signals. *Applied Acoustics*, *71*(7), 634–639. https://doi.org/10.1016/j.apacoust.2010.02.006

Aharoni, R., Glikman, E., & Krug, G. (2002). A novel high-speed rail inspection system. *Journal of Nondestructive Testing*, *7*(10), 9.

Alyaz, S. (2003). *Effects of heat treatment and chemical composition on microstructure and mechanical properties of hadfield steels*. Middle East Technical University.

Amini, A. (2016). *Online condition monitoring of railway wheelsets / by Arash Amini.* Thesis (PhD)--University of Birmingham, College of Engineering and Physical Sciences, 2016.

Amini, A., Entezami, M., Huang, Z., Rowshandel, H., & Papaelias, M. (2016). Wayside detection of faults in railway axle bearings using time spectral kurtosis analysis on high-frequency acoustic emission signals. *Advances in Mechanical Engineering*, *8*(11), 1687814016676000. https://doi.org/10.1177/1687814016676000

Anderson, J. A. (James A., & Rosenfeld, E. (1988). *Neurocomputing : foundations of research / edited by James A. Anderson and Edward Rosenfeld.* Cambridge, Mass.

Antoni, J., & Randall, R. B. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, *20*(2), 308–331. https://doi.org/10.1016/j.ymssp.2004.09.002

ASTM international. (2017). *A128 Standard specification for steel castings, Austenitic Manganese*. ASTM.

Bacioiu, D. (2019). *Vision-based Monitoring System for High Quality TIG Welding*.

Bal, B. (2018). A Study of Different Microstructural Effects on the Strain Hardening Behaviour of Hadfield Steel. *International Journal of Steel Structures*, *18*(1), 13–23. https://doi.org/10.1007/s13296-018-0302-9

Barile, C., Casavola, C., Pappalettera, G., & Pappalettere, C. (2015). Analysis of crack propagation in stainless steel by comparing acoustic emissions and infrared thermography data. *Engineering Failure Analysis*, *69*, 35–42. https://doi.org/10.1016/j.engfailanal.2016.02.022

Barke, D., & Chiu, W. K. (2005). Structural Health Monitoring in the Railway Industry: A Review. In *Structural Health Monitoring* (Vol. 4, Issue 1, pp. 81–93). https://doi.org/10.1177/1475921705049764

Barré, S., & Benzeggagh, M. L. (1994). On the use of acoustic emission to investigate damage mechanisms in glass-fibre-reinforced polypropylene. *Composites Science and Technology*, *52*(3), 369–376. https://doi.org/10.1016/0266-3538(94)90171-6

Bassim, M. N., Lawrence, S. S., & Liu, C. D. (1994). Detection of the onset of fatigue crack growth in rail steels using acoustic emission. *Engineering Fracture Mechanics*, *47*(2), 207–214. https://doi.org/10.1016/0013-7944(94)90221-6

Bhuiyan, M. Y. (n.d.). Acoustic emission sensor effect and waveform evolution during fatigue crack growth in thin metallic plate. *Journal of Intelligent Material Systems & Structures*, *29*(7), 1275–1285. https://doi.org/10.1177/1045389X17730930

Bhuiyan, M. Y., Bao, J., Poddar, B., & Giurgiutiu, V. (2018). Toward identifying crack-length-related resonances in acoustic emission waveforms for structural health monitoring applications. *Structural Health Monitoring*, *17*(3), 577–585. https://doi.org/10.1177/1475921717707356

Bishop, C. M. (1995). *Neural networks for pattern recognition / Christopher M. Bishop.* Clarendon Press.

Bonnett, C. F. (2005). *Practical railway engineering / Clifford F. Bonnett.* (2nd ed.). London : Imperial College Press.

BritishSteel. (2018). *Steel Grade Dimensions and Properties*. British Steel. https://britishsteel.co.uk/media/40810/steel-grade-dimensions-and-properties.pdf

Bruzelius, K., & Mba, D. (2004). An initial investigation on the potential applicability of Acoustic Emission to rail track fault detection. *NDT and E International*, *37*(7), 507–516. https://doi.org/10.1016/j.ndteint.2004.02.001

Bryson, A. E., & Ho, Y.-C. (1969). *Applied optimal control : optimization, estimation and control* (Y.-C. Ho, Ed.; Revised (ed.).). Washington, D.C.

Cannon, D. F., & Pradier, H. (1996). Rail rolling contact fatigue Research by the European Rail Research Institute. *Wear*, *191*(1–2), 1–13. https://doi.org/10.1016/0043-1648(95)06650-0

Chang, C. C., Chang, T. Y. P., Xu, Y. G., & Wang, M. L. (2000). Structural damage detection using an iterative neural network. *Journal of Intelligent Material Systems and Structures*, *11*(1), 32–42. https://doi.org/10.1106/XU88-UW1T-A6AM-X7EA

Chen, B., Yan, Z., & Chen, W. (2014). Defect Detection for Wheel-Bearings with Time-Spectral Kurtosis and Entropy. *Entropy*, *16*(1), 607–626. https://doi.org/10.3390/e16010607

Chen, S.-X., Zhou, L., Ni, Y.-Q., & Liu, X.-Z. (2021). An acoustic-homologous transfer learning approach for acoustic emission–based rail condition evaluation. *Structural Health Monitoring*, *20*(4), 2161–2181. https://doi.org/10.1177/1475921720976941

Culwick, R. (2019). *Remote conditioning monitoring of railway assets* (p. 89).

de Groot, P. J., Wijnen, P. A. M., & Janssen, R. B. F. (1995). Real-time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites. *Composites Science and Technology*, *55*(4), 405–412. https://doi.org/10.1016/0266-3538(95)00121-2

de Oliveira, R., & Marques, A. T. (2008). Health monitoring of FRP using acoustic emission and artificial neural networks. *Computers & Structures*, *86*(3), 367–373. https://doi.org/https://doi.org/10.1016/j.compstruc.2007.02.015

Ding, H. H., He, C. G., Ma, L., Guo, J., Liu, Q. Y., & Wang, W. J. (2016). Wear mapping and transitions in wheel and rail materials under different contact pressure and sliding velocity conditions. *Wear*, *352–353*, 1–8. https://doi.org/10.1016/j.wear.2016.01.017

Dunegan, H. L., Harris, D. O., & Tatro, C. A. (1968). Fracture analysis by use of acoustic emission. *Engineering Fracture Mechanics*, *1*(1), 105, 110–111, 122, IN23, IN24. https://doi.org/10.1016/0013-7944(68)90018-0

Ebrahimkhanlou, A., & Salamone, S. (2018). Single-sensor acoustic emission source localization in plate-like structures using deep learning. *Aerospace*, *5*(2), 50. https://doi.org/10.3390/aerospace5020050

Eden, H. C., Garnham, J. E., & Davis, C. L. (2005). Influential microstructural changes on rolling contact fatigue crack initiation in pearlitic rail steels. *Materials Science and Technology*, *21*(6), 623–629. https://doi.org/10.1179/174328405X43207

Eftekharnejad, B., Carrasco, M. R., Charnley, B., & Mba, D. (2011). The application of spectral kurtosis on Acoustic Emission and vibrations from a defective bearing. *Mechanical Systems and Signal Processing*, *25*(1), 266–284. https://doi.org/10.1016/J.YMSSP.2010.06.010

el Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? *Machine Learning in Radiation Oncology*, 3–11. https://doi.org/10.1007/978-3-319-18305-3_1

Ennaceur, C., Laksimi, A., Hervé, C., & Cherfaoui, M. (2006). Monitoring crack growth in pressure vessel steels by the acoustic emission technique and the method of potential difference. *International Journal of Pressure Vessels and Piping*, *83*(3), 197–204. https://doi.org/10.1016/j.ijpvp.2005.12.004

Erhan, D., Bengio, Y., Courville, A., & Vincent, P. (2009). *Visualizing Higher-Layer Features of a Deep Network Visualizing Higher-Layer Features of a Deep Network*. https://www.researchgate.net/publication/265022827

F50a - 200-800 kHz High-Sensitivity Flat Frequency Response AE Sensor, by Physical Acoustics. (n.d.). Retrieved March 15, 2022, from https://www.physicalacoustics.com/by-product/sensors/F50a-200-800-kHz-High-Sensitivity-Flat-Frequency-Response-AE-Sensor

Fausett, L. v. (1994). *Fundamentals of neural networks : architectures, algorithms and applications / Laurene V. Fausett.* Prentice-Hall International.

Fegredo, D. M., Shehata, M. T., Palmer, A., & Kalousek, J. (1988). The effect of sulphide and oxide inclusions on the wear rates of a standard C-Mn and a Cr-Mo alloy rail steel. *Wear*, *126*(3), 285–306. https://doi.org/10.1016/0043-1648(88)90171-8

Gagar, D., Foote, P., & Irving, P. E. (2015). Effects of loading and sample geometry on acoustic emission generation during fatigue crack growth: Implications for structural health monitoring. *International Journal of Fatigue*, *81*, 117–127. https://doi.org/10.1016/j.ijfatigue.2015.07.024

Garnham, J. E., Ding, R.-G., & Davis, C. L. (2010). Ductile inclusions in rail, subject to compressive rolling–sliding contact. *Wear*, *269*(11), 733–746. https://doi.org/10.1016/j.wear.2010.07.010

Godin, N., Huguet, S., Gaertner, R., & Salmon, L. (2004). Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers. *NDT & E International : Independent Nondestructive Testing and Evaluation*, *37*(4), 253–264. https://doi.org/10.1016/j.ndteint.2003.09.010

Grassie, S. K., & Kalousek, J. (1997). Rolling contact fatigue of rails: characteristics, causes and treatments. *6th International Heavy Haul Conference*, 38–404.

Grosse, C. U., Finck, F., Kurz, J. H., & Reinhardt, H. W. (2004). Improvements of AE technique using wavelet algorithms, coherence functions and automatic data analysis. *Construction and Building Materials*, *18*(3), 203–213. https://doi.org/10.1016/j.conbuildmat.2003.10.010

Gumus, B., Bal, B., Gerstein, G., Canadinc, D., & Maier, H. J. (2016). Twinning activity in high-manganese austenitic steels under high velocity loading. *Materials Science and Technology*, *32*(5), 463–465. https://doi.org/10.1179/1743284715Y.0000000111

Gunel, E., Yilmazer, P., Bicer, U., Altintas, B., Amini, A., Vallely, P., Huang, Z., García Márquez, F. P., Kaewunruen, S., & Papaelias, M. (2017). *Increasing the Reliability, Availability, Maintainability and Safety of Railway Network Operations through effective Remote Condition Monitoring*.

Guo, S. L., Sun, D. Y., Zhang, F. C., Feng, X. Y., & Qian, L. H. (20130730). Damage of a Hadfield steel crossing due to wheel rolling impact passages. *Wear*, *305*(1–2), 267–273. https://doi.org/10.1016/j.wear.2013.01.019

Hadfield, R. A. (Robert A. (1914). *The magnetic and mechanical properties of manganese steel / by Sir R.A. Hadfield and B. Hopkinson.* (B. (Bertram) Hopkinson, I. and S. Institute, & M. Society, Eds.). London : Iron and Steel Institute.

Hagan, M. T. (1996). *Neural network design / Martin T. Hagan, Howard B. Demuth, Mark Beale.* PWS Pub.

Hamel, F., Bailon, J. P., & Bassim, M. N. (1981). Acoustic emission mechanisms during high-cycle fatigue. *Engineering Fracture Mechanics*, *14*(4), 853–860. https://doi.org/10.1016/0013-7944(81)90097-7

Hamstad, M. A., O'gallagher, A., & Gary, J. (2002). A WAVELET TRANSFORM APPLIED TO ACOUSTIC EMISSION SIGNALS: PART 1: SOURCE IDENTIFICATION #. In *J. Acoustic Emission* (Vol. 20).

Han, Z., Luo, H., Sun, C., Li, J., Papaelias, M., & Davis, C. (2014). Acoustic emission study of fatigue crack propagation in extruded AZ31 magnesium alloy. *Materials Science and Engineering A*, *597*, 270–278. https://doi.org/10.1016/j.msea.2013.12.083

Harzallah, R., Mouftiez, A., Felder, E., Hariri, S., & Maujean, J.-P. (2010). Rolling contact fatigue of Hadfield steel X120Mn12. *Wear*, *269*(9–10), 647–654. https://doi.org/10.1016/j.wear.2010.07.001

Hebb, D. O. (Donald O. (1949). *The organization of behavior : a neuropsychological theory.* New York : Wiley.

Huang, Z. (2016). *Integrated railway remote condition monitoring / by Zheng Huang.* (U. of Birmingham. C. of E. and P. Sciences, Ed.). Thesis (Ph.D.)--University of Birmingham, College of Engineering and Physical Sciences, 2017.

Huang, Z. (2017). *Integrated railway remote condition monitoring*.

ISUNDE-ed. (2022). *Nondestructive Evaluation Techniques : Acoustic Emission Testing*. https://www.nde-ed.org/NDETechniques/AcousticEmission/AE_Theory-Wave.xhtml

Jayaswal, P., Verma, S., & Wadhwani, A. (2011). Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis. *Journal of Vibration and Control*, *17*(8), 1131–1148. https://doi.org/10.1177/1077546310361858

Johnson, H. H. (1965). *Calibrating the Electric Potential Method for Studying Slow Crack Growth*. Defense Technical Information Center.

Kang, J., Zhang, F. C., Long, X. Y., & Lv, B. (2014). Cyclic deformation and fatigue behaviors of Hadfield manganese steel. *Materials Science & Engineering A*, *591*, 59–68. https://doi.org/10.1016/j.msea.2013.10.072

Karaman, I., Sehitoglu, H., Gall, K., Chumlyakov, Y. I., & Maier, H. J. (2000). Deformation of single crystal Hadfield steel by twinning and slip. *Acta Materialia*, *48*(6), 1345–1359. https://doi.org/10.1016/S1359-6454(99)00383-3

Kongpuang, M., & Culwick, R. (n.d.). *Quantitative analysis of the structural health of railway turnout using acoustic emission technique*.

Kongpuang, M., Culwick, R., Cheputeh, N.-A., Marsh, A., Jantara Junior, V. luiz, Vallely, P., Kaewunruen, S., & Papaelias, M. (2021, June 18). Quantitative Analysis of the Structural Health of Railway Turnout Using Acoustic Emission Technique. *The Seventeenth International Conference on Condition Monitoring and Asset Management*.

Kostryzhev, A. G., Davis, C. L., & Roberts, C. (2013). Detection of crack growth in rail steel using acoustic emission. *Ironmaking & Steelmaking*, *40*(2), 98–102. https://doi.org/10.1179/1743281212Y.0000000051

Lapedes, A., & Farber, R. (1988). How neural nets work. In D. Z. Anderson (Ed.), *Neural Information Processing Systems* (pp. 442–456). American Institute of Physics.

le Cun, Y. (1986). Learning Process in an Asymmetric Threshold Network. *Disordered Systems and Biological Organization*.

Lebold, M., Mcclintic, K., Campbell, R., Byington, C., & Maynard, K. (1985). *Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics*. 623–634.

Lee, J. J., Lee, J. W., Yi, J. H., Yun, C. B., & Jung, H. Y. (2005). Neural networks-based damage detection for bridges considering errors in baseline finite element models. *Journal of Sound and Vibration*, *280*(3–5), 555–578. https://doi.org/10.1016/j.jsv.2004.01.003

Lee, S. H., & Lee, D. (2008). International Journal of Production Research In-process monitoring of drilling burr formation using acoustic emission and a wavelet-based artificial neural network In-process monitoring of drilling burr formation using acoustic emission and a wavelet-based artificial neural network. *International Journal of Production Research*, *46*(17), 4871–4888. https://doi.org/10.1080/00207540601152040

Li, D., Kuang, K. S. C., & Koh, C. G. (2017). Fatigue crack sizing in rail steel using crack closure-induced acoustic emission waves. *Measurement Science and Technology*, *28*(6), 065601. https://doi.org/10.1088/1361-6501/aa670d

Li, L., Lomov, S. v, Yan, X., & Carvelli, V. (2014). Cluster analysis of acoustic emission signals for 2D and 3D woven glass/epoxy composites. *Composite Structures*, *116*(1), 286–299. https://doi.org/10.1016/j.compstruct.2014.05.023

Lin, Y., Nie, Z., & Ma, H. (2017). Structural Damage Detection with Automatic Feature-Extraction through Deep Learning. *Computer-Aided Civil and Infrastructure Engineering*, *32*(12), 1025–1046. https://doi.org/10.1111/mice.12313

Lindley, T. C., Palmer, I. G., & Richards, C. E. (1978). Acoustic emission monitoring of fatigue crack growth. *Materials Science and Engineering*, *32*(1), 1–15. https://doi.org/10.1016/0025-5416(78)90206-9

Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, *4*(2), 4–22.

Lv, B., Zhang, M., Zhang, F. C., Zheng, C. L., Feng, X. Y., Qian, L. H., & Qin, X. B. (2012). Micro-mechanism of rolling contact fatigue in Hadfield steel crossing. *International Journal of Fatigue*, *44*, 273–278. https://doi.org/10.1016/j.ijfatigue.2012.04.010

Magel, E. E. (2011). *Rolling contact fatigue: a comprehensive review*. https://doi.org/https://doi.org/10.4224/23000318

Malekzadeh, M., Atia, G., & Catbas, F. N. (2015). Performance-based structural health monitoring through an innovative hybrid data interpretation framework. *Journal of Civil Structural Health Monitoring*, *5*(3), 287–305. https://doi.org/10.1007/s13349-015-0118-7

Masters, T. (1994). *Signal and image processing with neural networks : a C [plus plus] sourcebook / Timothy Masters.* New York.

Mcculloch, W. S., & Pitts, W. (1990). A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY* n. *Bulletin of Mothemnticnl Biology*, *52*(2), 99–115.

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons: Expanded Edition*. MIT Press.

Mohammad, M., Abdullah, S., Jamaludin, N., & Innayatullah, O. (2014). Predicting the fatigue life of the SAE 1045 steel using an empirical Weibull-based model associated to acoustic emission parameters. *Materials & Design*, *54*, 1039–1048. https://doi.org/10.1016/j.matdes.2013.09.021

Moore, P. O., Miller, R. K., & Hill, E. v. K. (Eds.). (2005). *Nondestructive testing handbook: Acoustic Emission Testing* (3rd ed., Vol. 6). American Society of Nondestructive Testing.

Moorthy, V., Jayakumar, T., & Raj, B. (19941101). Acoustic emission behavior during stage II fatigue crack growth in an AISI type 316 austenitic stainless steel. *Bull. Mater. Sci. (India)*, *17*(6), 699–715. https://doi.org/10.1007/BF02757552

Morscher, G. N. (1999). Modal acoustic emission of damage accumulation in a woven SiC/SiC composite. *Composites Science and Technology*, *59*(5), 687–697. https://doi.org/10.1016/S0266-3538(98)00121-3

Morton, T. M., Harrington, R. M., & Bjeletich, J. G. (1973). Acoustic emissions of fatigue crack growth. *Engineering Fracture Mechanics*, *5*(3), 691, 692–693, 697, IN17, IN18. https://doi.org/10.1016/0013-7944(73)90047-7

Muir, C., Swaminathan, B., Almansour, A. S., Sevener, K., Smith, C., Presby, M., Kiser, J. D., Pollock, T. M., & Daly, S. (2021). Damage mechanism identification in composites via machine learning and

acoustic emission. *Npj Computational Materials*, *7*(1), 95. https://doi.org/10.1038/s41524-021-00565-x

NationalInstruments. (n.d.). *Understanding FFTs and Windowing*.

Network Rail. (2018). *Module 07 Management of rail defects*.

NetworkRail. (2021). *Annual expenditure 2019-20*. https://www.networkrail.co.uk/who-we-are/transparency-and-ethics/transparency/our-information-and-data

New Measurement Train (NMT) - Network Rail. (2022). https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/our-fleet-machines-and-vehicles/new-measurement-train-nmt/

Office of Rail. (2022). *Freight rail usage and performance January to March 2022*.

Ono, K. (2011). ACOUSTIC EMISSION IN MATERIALS RESEARCH-A REVIEW Ancient metals View project ACOUSTIC EMISSION IN MATERIALS RESEARCH-A REVIEW. In *J. Acoustic Emission* (Vol. 29). http://www.physik.uni-augsburg.de/

Ould Amer, A., Gloanec, A. L., Courtin, S., & Touze, C. (2013). Characterization of fatigue damage in 304L steel by an acoustic emission method. *Procedia Engineering*, *66*, 651–660. https://doi.org/10.1016/j.proeng.2013.12.117

PAC. (2005). *R50α General Purpose, 500 kHz Resonant Frequency Acoustic Emission Sensor* (Patent No. 135–04). PAC.

PAC. (2022). *Physical Acoustics AE Technology - Overview of Acoustic Emission NDT Technology*. https://www.physicalacoustics.com/ae-technology/

Papaelias, M., Amini, A., Culwick, R., Heesom, J., Huang, Z., Jantara Junior, V. L., Kaenwunruen, S., Kerkyras, S., Kongpuang, M., Garcia Marquez, F. P., Shi, S., Upton, A., & Vallely, P. (2018). Advanced remote condition monitoring of railway infrastrucutre and rolling stock. *1st International Conference on Welding & NDT, of HSNT and WGI*.

Papaelias, M., Huang, Z., Amini, A., Ayorkinos Papaelias, M., Vallely, P., Day, N., Sharma, R., Kerkyras, Y., & Kerkyras, S. (2014). *Advanced wayside condition monitoring of rolling stock wheelsets Real-time monitoring of welding quality View project EC FP7-NIMO View project Advanced wayside condition monitoring of rolling stock wheelsets*. https://www.researchgate.net/publication/275040232

Parker, D. (1985). *Learning Logic Technical Report TR-87*.

Perez-Unzueta, A. J., & Beynon, J. H. (1993). Microstructure and wear resistance of pearlitic rail steels. *Wear*, *162–164*(Part A), 173–182. https://doi.org/10.1016/0043-1648(93)90498-B

Peters, N. W. (2005). *THE PERFORMANCE OF HADFIELD'S MANGANESE STEEL AS IT RELATES TO MANUFACTURE*.

Ph Papaelias, M., Roberts, C., & Davis, C. L. (2008). A review on non-destructive evaluation of rails: State-of-the-art and future development. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *222*(4), 367–384. https://doi.org/10.1243/09544097JRRT209

Rabiei, M., & Modarres, M. (2013). Quantitative methods for structural health management using in situ acoustic emission monitoring. *International Journal of Fatigue*, *49*, 81–89. https://doi.org/10.1016/j.ijfatigue.2012.12.001

Ramirez-Jimenez, C. R., Papadakis, N., Reynolds, N., Gan, T. H., Purnell, P., & Pharaoh, M. (2004). Identification of failure modes in glass/polypropylene composites by means of the primary frequency content of the acoustic emission event. *Composites Science and Technology*, *64*(12), 1819–1827. https://doi.org/10.1016/j.compscitech.2004.01.008

Ringsberg, J. W., & Bergkvist, A. (2003). On propagation of short rolling contact fatigue cracks. *Fatigue & Fracture of Engineering Materials & Structures*, *26*(10), 969–983. https://doi.org/10.1046/j.1460-2695.2003.00657.x

Roberts, T. M., & Talebzadeh, M. (2003). Acoustic emission monitoring of fatigue crack propagation. *Journal of Constructional Steel Research*, *59*(6), 695–712. https://doi.org/10.1016/S0143-974X(02)00064-0

Rochester, N., Holland, J., Haibt, L., & Duda, W. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *I.R.E. Transactions on Information Theory*, *2*(3), 80–93. https://doi.org/10.1109/TIT.1956.1056810

Rommelaere, M., & Maujean, J.-P. (2013). Pre-hardened cossings cut life-cycle costs. *Railway Gazette International*, *169*(2), 48–50.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.

Sawalhi, N., & Randall, R. B. (2004). The application of spectral kurtosis to bearing diagnostics. *Proceedings of the 2004 AAS Meeting*, *1*.

Schilke, M. (2013). *Degradation of railway rails from a materials point of view* [Doctor of Philosophy]. Chalmers University of Technology.

Schilke, M., Ahlström, J., & Karlsson, B. (2010). Low cycle fatigue and deformation behaviour of austenitic manganese steel in rolled and in as-cast conditions. *Procedia Engineering*, *2*(1), 623–628. https://doi.org/https://doi.org/10.1016/j.proeng.2010.03.067

Shan, D., & Nayeb-Hashemi, H. (1999). Fatigue-life prediction of SiC aluminum composite using a Weibull model. *NDT and E International*, *32*(5), 265–274. https://doi.org/10.1016/S0963-8695(98)00053-X

Shi, S., Han, Z., Liu, Z., Vallely, P., Soua, S., Kaewunruen, S., & Papaelias, M. (201804). Quantitative monitoring of brittle fatigue crack growth in railway steel using acoustic emission. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *232*(4), 1211–1224. https://doi.org/10.1177/0954409717711292

Shi, S., Soua, S., & Papaelias, M. (2017). Remote Condition monitoring of rails and corssings using acoustic emission. *NSIRC Annual Conference*.

Siddique, A., Yadava, G. s, & Singh, B. (2003). *Applications of artificial intelligence techniques for induction machine stator fault diagnostics: review*. https://doi.org/10.1109/DEMPED.2003.1234543

Sikorski, R. U. E.-W. (2013). Hit Detection and Determination in AE Bursts. In *Acoustic Emission* (p. Ch. 1). IntechOpen. https://doi.org/10.5772/54754

Smith, R. W., & Mackay, W. B. F. (2003). AUSTENITIC MANGANESE STEELS - DEVELOPMENTS FOR HEAVY HAUL RAIL TRANSPORTATION. *Canadian Metallurgical Quarterly*, *42*(3), 333–342. https://doi.org/10.1179/cmq.2003.42.3.333

Specht, D. F. (1967). Vectorcardiographic Diagnosis Using the Polynomial Discriminant Method of Pattern Recognition. *IEEE Transactions on Biomedical Engineering*, *BME-14*(2), 90–95. https://doi.org/10.1109/TBME.1967.4502476

Strantza, M., van Hemelrijck, D., Guillaume, P., & Aggelis, D. G. (2017). Acoustic emission monitoring of crack propagation in additively manufactured and conventional titanium components. *Mechanics Research Communications*, *84*, 8–13. https://doi.org/10.1016/j.mechrescom.2017.05.009

Suwansin, W., & Phasukkit, P. (2021). Deep Learning-Based Acoustic Emission Scheme for Nondestructive Localization of Cracks in Train Rails under a Load. *Sensors*, *21*(1). https://doi.org/10.3390/s21010272

Swindlehurst, W. (1973). Acoustic emission - 1 Introduction. *Non-Destructive Testing*, *6*(3), 152–158. https://doi.org/10.1016/0029-1021(73)90017-0

NSWRailcop. (2019). TMC 226 Rail Defects Handbook. In *Engineering Manual Track* (Issue 1.2). NSWRailcorp.

Uddin A.K.M, F. (2002). Application of AE to Fracture Toughness and Crack Analysis by BEM in Concrete. *NDT*, *7*(9). https://www.ndt.net/article/v07n09/01/01.htm

Ultrasonic Rail Testing - Pandrol. (2022). https://www.pandrol.com/product/ultrasonic-rail-testing/

Vallely, P. (2015). *A Holistic Approach to Remote Condition Monitoring for the Accurate Evaluation of Railway Infrastructure and Rolling Stock*.

Vallen, H. (2002). AE Testing Fundamentals, Equipment, Applications. *Journal of Nondestructive Testing (Germany)*, *7*(9), 1–30.

Watanabe, S. (1985). *Pattern Recognition: Human and Mechanical*. John Wiley and Sons.

Wen, Y. H., Peng, H. B., Si, H. T., Xiong, R. L., & Raabe, D. (2014). A novel high manganese austenitic steel with higher work hardening capacity and much lower impact deformation than Hadfield manganese steel. *Materials in Engineering*, *55*, 798–804. https://doi.org/10.1016/j.matdes.2013.09.057

Werbos, P. (1974). *Beyond regression : new tools for prediction and analysis in the behavioral sciences /*.

Widrow, B., & Hoff, M. (1960). Adaptive Switching Circuits. *1960 IRE WESCON Convention Record*, 96–104.

Willberry, J., Papaelias, M., & Fernando, G. (2020). Structural Health Monitoring Using Fibre Optic Acoustic Emission Sensors. *Sensors*, *20*, 6369. https://doi.org/10.3390/s20216369

Yilmazer, P. (2013). *Structural health condition monitoring of rails using acoustic emission techniques*.

Yu, J., Ziehl, P., Zárate, B., & Caicedo, J. (2011). Prediction of fatigue crack growth in steel bridge components using acoustic emission. *Journal of Constructional Steel Research*, *67*(8), 1254–1260. https://doi.org/10.1016/j.jcsr.2011.03.005

Zhang, J., Ma, H., Yan, W., & Li, Z. (2016). Defect detection and location in switch rails by acoustic emission and Lamb wave analysis: A feasibility study. *Applied Acoustics*, *105*, 67–74. https://doi.org/10.1016/j.apacoust.2015.11.018

Zhang, X., Feng, N., Wang, Y., & Shen, Y. (2014). An analysis of the simulated acoustic emission sources with different propagation distances, types and depths for rail defect detection. *Applied Acoustics*, *86*, 80–88. https://doi.org/10.1016/j.apacoust.2014.06.004

Zhang, X., Feng, N., Wang, Y., & Shen, Y. (2015). Acoustic emission detection of rail defect based on wavelet transform and Shannon entropy. *Journal of Sound and Vibration*, *339*, 419–432. https://doi.org/10.1016/j.jsv.2014.11.021

Zhu, Y., & Olofsson, U. (2014). An adhesion model for wheel–rail contact at the micro level using measured 3d surfaces. *Wear*, *314*(1–2), 162–170. https://doi.org/10.1016/j.wear.2013.11.031