Incorporating automated rail fatigue damage detection algorithms with crack growth modelling

Bello Sambo

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Institute of Railway Research
School of Computing and Engineering
The University of Huddersfield

November 2018

ACKNOWLEDGMENTS

I would like to express my gratitude to my Supervisor(s), Dr Adam Bevan and Dr Paul Allen who provided me with professional expertise, technical support and encouragement all the way. The joy and enthusiasm they both share for their research was contagious and motivational for me during this PhD project.

Also very special thanks go to the director, Institute of Railway Research Professor Simon Iwnicki and entire members of the institute, who gave me the opportunity to work in an environment based on effective and friendly relationships built on trust, respect, commitment to continuing self-development and development of others. On a more personal level, I would like to thank my parents Sambo Ingawa and Bilkisu Sambo and my siblings whose assistance was important for the success of this project. They have been supportive in numerous ways giving me the much needed inspiration in my life. Thank you all for your kindness and dedication. I hope God blesses all of you.

Executive Summary

This thesis examines the feasibility of incorporating Non Destructive Testing (NDT) of rail surface damage by means of combining image processing with damage prediction models. As rail traffic and adherence to safety measures become increasingly strict on the network, the associated maintenance cost of rail infrastructure must be kept at a minimum. Proactive maintenance is crucial to maintaining the competitive advantage of rail transport. A considerable amount of research has been done on improving the practical tediousness associated with popular condition monitoring techniques in rail industry e.g. Ultrasonic, and Eddy current method. This thesis aims to fill the gap of yet to be explored benefit, of combining detection and prediction of RCF damage. This research project will contribute to the rail industry by simplifying maintenance operations and support decision making.

In this thesis, a summary of existing image-based NDT and crack growth models is presented as a foundation on which the novel application is built. It could be said that similar research mainly focuses on quantifying severity of damage without predicting crack behaviour. The simulated results of the proposed image processing algorithm confirm superiority of local illumination invariant enhancement, multi-window segmentation, and cascaded feature extraction. The influential parameters of these methods are consistent within each image data set but differ across all sets. This is observed to be as a result of difference in environmental and reflection properties of acquired images. A sensitivity analysis of the proposed algorithm on data set 2 suggests a non-linear relationship between severity of damage and pixel mean intensity including variance.

Taking to account fracture mechanics aspect of this thesis, the influence of crack geometry on growth rate and path has been established by case study of newly initiated and critically grown cracks. It was further established that larger cracks are observed to grow faster than smaller ones. In addition, the influence of track curve radius and supporting structures on wheel rail contact dynamics is well understood from the structural mechanic's tests related to

contact forces and bending moment. These translate to increase or decrease in contact stresses, strains, and the propagation rate of defects.

Unlike other predictive models, the method developed in this thesis focuses on replicating the actual surface condition of the rail prior to estimating the fracture parameters (using detailed 3D Finite Element model) that dictate residual life of the rail asset. The model makes it possible to combine two separate maintenance activities i.e. detection and prediction without inducing down time of the service.

A direct impact of this novel application is the utilisation of the actual crack boundary for prediction of fracture behaviour. It is insinuated that stress distribution of actual crack boundary differs from elliptical equivalent assumptions.

Further work would include improving detection aspect of the novel application to avoid intersecting boundary coordinates, which are not readily imported into the Linear Elastic Fracture Mechanics (LEFM) prediction model. It is also beneficial to expand the prediction aspect of the research work to include influence of neighbouring cracks and fluid entrapment for more flexible analysis of other environmental and contact conditions. To improve on current work, it will be useful to conduct laboratory investigations on the influence of Image Acquisition System (IAS) light source in relation to illumination inequality within the captured image. Also fracture mechanics experimental validation can be used to assert the accuracy of the method.

Table of contents

List of Figures

List of Tables

List of Symbols

Symbol	Description
Δ	Pixel mean intensity offset
μ	Mean intensity value
$I(x, y)$	Input image
R_m	Image required mean
R_v	Image required variance
$im(x, y)$	Windowed sample of the original
$\overline{im(x, y)}$	Mean intensity of the window.
$\sigma(x, y)$	Windowed standard deviation
$n(x, y)$	Total number of data points
$T_N(x, y)$	Niblack segmentation threshold
$T_S(x, y)$	Savuola segmentation threshold
$T_W(x, y)$	Wolf's segmentation threshold
T	OGT and AIT threshold
θ	Orientation of the gabor filter [degrees]
σ	Standard deviation of shape distance
d_i	Defect maximum length [mm]
$P(x, y)$	Wheel-rail contact pressure [N/m^2]
P_{max}	Maximum wheel-rail contact pressure [N/m^2]
σ_x, σ_y, and σ_z	Principal contact stresses [MPa]
J_I, J_{II}	Pure mode-I and II J integral [N/m]
J_{vmI}, J_{vmII}	Mixed mode-I and II J-integral [N/m]
K_I, K_{II}	Pure mode-I and II SIF [N/m]
η and ξ_o.	Elliptical coordinate representation
σ_θ	Crack border tangential stress [MPa]
σ_{nom}	Nominal stress [MPa]
σ_I, σ_{II}	Biaxial mode-I and II stress [MPa]
$\aleph$	Load biaxiality
β	Crack inclination angle [degrees]
$\dfrac{da}{dN}$	Growth rate [nm/cycle]
ΔK_{eq}	Equivalent range SIF [N/m]
ΔK_{th}	Range of SIF threshold
N_f	Number of loading cycles until failure

List of Abbreviations

ΔKI	Range of Mode-I SIF
ΔKII	Range of Mode-II SIF
$\Delta Keqv$	Range of equivalent SIF
AI	Automatic Iterative algorithm
AIT	Automatic Iterative thresholding
ANN	Artificial Neural Network
AVS	Automated Vision System
BPHEME	Brightness preserving histogram equalisation with maximum entropy
CDF	Cumulative distribution function
CLAHE	Contrast limited adaptive histogram equalisation
CTS	Compact tension shear
DB	Deutsche Bahn
DFT	Discrete Fourier Transform
DoT	Department of Transport
DSP	Digital signal processing
E1	Local normalisation enhancement method
E2	Visibility measure enhancement measure
E3	Linear moving average filter enhancement method
E4	Fast Fourier Transform enhancement method
EC	Eddy currents
EMSR	Extended minimum subscribed rectangle
EPFM	Elastic plastic fracture mechanics
EU	European Union
FA-H1	First sample of field acquired image for heavily damaged rail surface
FA-L1	First sample of field acquired image for lightly damaged rail surface
FA-M1	First sample of field acquired image for moderately damaged rail surface
FE	Feature extraction
FE1	Standard deviation histogram shape distance feature extraction method
FE2	Gabor filter feature extraction method
FE3	Morphological feature extraction method
FEA	Finite element analysis
FFT	Fast Fourier Transform
FRA	Federal Railroad Administration
FT	Fourier Transform
GCC	Gauge Corner Cracking
GHE	Global Histogram equalisation
GLG	Gray level grouping
GSM	Global system for mobile communication
GSP	Global positioning system
HPF	High pass filter
IAS	Image acquisition system
ICRI	International Collaboration on Rail Initiative
Keqv	Equivalent SIF
KI	Mode-I SIF
KII	Mode-II SIF
Lab-1	First sample of laboratory acquired image of damaged rail surface
Lab-2	Second sample of laboratory acquired image of damaged rail surface
Lab-3	Third sample of laboratory acquired image of damaged rail surface

Lab-4	Fourth sample of laboratory acquired image of damaged rail surface
LCCD	Linear charged coupled device
LDR	Light dependent resistor
LEFM	Linear elastic fracture mechanics
LHE	Local histogram equalisation
LMAF	Linear moving average filter
LMF	Linear mean filter
LN	Local normalisation
LPF	Low pass filter
MERR	Maximum energy release rate
MLC	Michelson contrast
MSR	Minimum subscribed rectangle
MVS	Machine vision system
NDE	Non-destructive evaluation
NDT	Non-destructive testing
NRC	National Research Council-Canada
OGT	Occurrence global thresholding
PPO	Point processing operation
PSNR	Peak signal to noise ratio
RAMS	Reliability Availability Maintainability and Safety
RCF	Rolling contact fatigue
RM	Required mean
ROC	Receiver operating characteristics
ROI	Region of interest
RSSB	Rail Safety and Standards Board
RV	Required variance
S1	Occurrence global thresholding method
S2	Wolfs thresholding method
S3	Automatic iterative thresholding method
SDHSD	Standard deviation of histogram shape distance
SED	Strain energy density
SEDM	Spatial domain enhancement method
SGLG	Selected gray level grouping
SIF	Stress intensity factor
UT	Ultrasonic testing
V/T SIC	Vehicle/Track Systems Interface Committee
VM	Visibility measure
WLRM	Whole Life Rail Model

Chapter 1 Introduction

A considerable amount of derailments and service disruptions in railway systems has been attributed to track-related issues, as discussed by Banerjee, U. (2013). The Office of Rail and Road in 2006 confirmed the derailment in Hatfield UK, was caused by multiple Rolling Contact Fatigue (RCF) cracks resulting in rail breaking as the train passed. As a result of increase in axle loads, vehicles primary yaw suspension stiffness, wheel-rail conicity and capacity, greater damage to the rail infrastructure is observed on most networks around the world. The statistics of rail safety in Great Britain released by the Office of Rail and Road in 2017 accounts for 687 cases of train accidents, most of which as a result of rail defects. RCF cracks initiate in the surface or subsurface of rails and can propagate to cause rail failure, impacting on safety, maintenance and costs. The vision for Great Britain (GB) railway, as set out in the Rail Technical Strategy, Network Rail Limited, (2016), emphasises the need for improvements in condition monitoring of rail assets from current inspection methods to more automated procedures ('intelligent' and 'self-monitoring' systems) capable of improving the efficiency of maintenance planning and decisions. Rowshandel, H. Papaelias, M. Roberts, C. & Davis, C. (2010) highlighted the benefits of this vision including targeted risk-based maintenance and repair of rail infrastructure to deliver improved reliability, fewer delays, increased capacity, resilience, safety and all round improved performance of the railway industry. The limitations of manual inspection of rail defects are well understood. In fact, much earlier works of Martin, A. & Tosunoglu, S. (2000) also concluded that conventional rail non-destructive evaluation (NDE) techniques should be replaced by more reliable, efficient, and flexible systems. Although associated with higher complexities, there exist hybrid systems that offer more reliable high-speed inspection by simultaneously using two or more rail inspection methods (e.g. ultrasonic probes, and eddy current probes). It is therefore sensible to investigate and model RCF detection systems that are more intelligent and self- learning with fewer complexities. This PhD thesis presents a technology readiness level 1-2 report on the development of an intelligent image processing algorithm capable of generating statistical data

and automatically estimating the propagation behaviour of each detected artefact (more intelligently). This was achieved by a combination of spatial filtering and illumination invariant enhancement for pre-processing of the input image, after which segmentation of the RCF (or other damage types) artefacts by binary thresholding is performed in addition to morphological cleaning. The resulting binary image is calibrated with known dimensions of rail geometry for extraction of defect geometry (statistical data) for automatic incorporation within a structural mechanic's model. Such an incorporated model furthers the scope of rail asset remote predictive maintenance in industry mainly by capitalizing on the yet to be explored benefits of linking NDE method(s) to prediction models for residual life estimation which supports optimization and planning of rail maintenance (e.g. grinding and renewal). This research also studies the feasibility of adopting the proposed technology in industry by identifying the enabler and blocker of the application after detailed sensitivity and parametric studies of all components of the model.

1.1 Overview and project brief

1.1.1 Novel aspect of the research

The novelty of this research work not only resides within the algorithms related to image processing; capable of detecting surface RCF defects, but also in the capacity of the application to generate geometrical information (statistical data) related to each artefact, and further automatically process the geometrical data in accordance with LEFM model(s) for prediction of defect fracture mechanics behaviour in support of rail maintenance activities.

1.1.2 Aim(s), and objectives of research

This research aims at developing an intelligent image processing algorithm capable of detecting fatigue defects from an on-board surface RCF image acquisition system (called the RAILSCOPE image acquisition system). The algorithm is intelligent as it generates statistical data such as total number of detected defects per image, the maximum length, area, orientation, and boundary coordinates of each defect. Such data can be used to estimate real time severity of damage and predict future damage including maintenance needs, thereby

improving predictive maintenance of rail assets in terms of flexibility, reliability, remoteness, safety, and overall cost.

To achieve the above aim(s) the following intermediary objectives were set:

- Develop image processing RCF detection algorithm (MATLAB).

- Develop a global contact (wheel-rail) model for generating wheel-rail contact stresses and strains (COMSOL Multi-physics).

- Develop LEFM surface RCF crack propagation model (MATLAB).

- Perform field application and validation of proposed image processing algorithm (applied on NRC and ICRI data).

- Perform tangent and curve track parametric study of global track model with validation using VAMPIRE data.

- Link the defect detection algorithm to fracture mechanics prediction models and analyze the applicability of the technology in rail industry.

1.1.3 Contributions of the research to rail industry

Linking defect detection and prediction models attempts to contribute to rail industry, by the flexibility of remote automatic evaluation of rail integrity and by incorporating defect data to local fracture mechanics for predictive maintenance applications. The proposed method is anticipated to contribute to rail industry as follows:

This application jointly advises the maintenance decisions, through defect detection, quantifying defect severity, and predicting fracture behaviour of critically damaged track sections within a single package/program that makes the maintenance process to be easier, faster and cheaper.

Contrary to similar research work (in the area of crack propagation) that assumes the surface geometry of cracks on rails as being elliptical, this project accounts for the actual boundary

coordinates of each detected defect as captured from rail image(s) thereby accounting for more realistic stress concentration at the crack tip(s).

The ease of incorporating other contact conditions, such as influence of other crack growth conditions; motion of crack faces due to passing wheel, hydraulic pressure, fluid entrapment and squeeze film effect, pneumatic pressure etc. will offer the flexibility of analysing different and/or specific sites/environmental conditions.

The research identifies initial values of influential parameters related enhancement, segmentation, and feature extraction of both field acquired and laboratory controlled image samples. These include local window size for spatial filtering, block size for mean intensity and variance for illumination invariant enhancement. The pixel intensity threshold value for binary segmentation of suspected defect region is identified, while the filter frequency and energy threshold values for Gabor/Texture based feature extraction, including Standard Deviation Histogram Shape Distance (SDHSD) for morphological based feature extraction is presented.

1.1.4 Methodology

An intelligent image processing algorithm capable of detecting and quantifying the geometrical features of fatigue damage from rail head images is developed. The image data set 1-3 (detailed in section 3.2.1-3.2.3) relating to surface RCF damage on rails was obtained from Manchester Metropolitan University (MMU), National Research Council (NRC) Canada, and Collaborative project on RCF damage respectively. These images were analysed in MATLAB (Image processing toolbox) to firstly undergo enhancement, segmentation, morphological cleaning followed by defect feature extraction. In this PhD research, enhancement algorithms related to statistical filtering (Linear Moving Average Filter-LMAF), illumination invariant Local Normalisation (LN), Visibility Measure (VM), and Fast Fourier Transform (FFT) are individually investigated on data set 1, data set 2 and data set 3. More specific to field acquired images of RCF damage in rails (i.e. data set 2 and data set 3), a specific concatenation of the

aforementioned enhancement methods is proposed (in chapter 3) after detailed analysis of the Peak Signal to Noise Ratio (PSNR) of the enhanced images of each of the method mentioned above.

Segmentation by binary thresholding in the form of Wolf's algorithm, Object/Background Occurrence Probability approach (OGT), and an Automatic Iterative (AI) algorithm was investigated for fragmenting suspect defect regions from the enhanced image of all data sets. Although similar behaviour was observed during a comprehensive comparative study in all the aforementioned segmentation methods, the AI method was preferred due to its higher versatility across all samples compared to the others.

As a good practice in analysing segmented images, considering the fact that it is practically impossible to obtain a segmented image with no false segment, a morphological cleaning step is included. These morphological functions ranging from fill, clean, thin, dilate, erode etc. are utilised before feature extraction is performed.

The cleaned image is finally post-processed for true detection of RCF damage using either morphological, geometrical, or even texture features typical to defects. To this end, the following methods have been combined for improved defect detection; Standard Deviation Histogram Shape Distance (SDHSD), Geometrically Acceptable Rail Damage Feature Extraction and Gabor/Texture-based Feature Extraction. A concatenation of all these feature extraction methods are used for all data sets examined in this research.

After successful detection of RCF defects, the geometrical properties of each detected defect (of specific interest is maximum length of defect, surface orientation, and boundary coordinates) are automatically measured using the region property function of the MATLAB image processing toolbox. By means of the Livelink interface, the defect geometry is automatically incorporated into a detailed 3-D revolved S1002 wheel and an extruded 60E1 rail profile assembly. The assembly (detailed in chapter 4) referred to as the global contact model, considers structural steel material properties and is initialised using VAMPIRE contact data for generating the stress fields typical to wheel-rail interface subject to 80kN axle load.

Analytical evaluation of the J-integral is used to estimate the Stress Intensity Factor (SIF) values of the dominant RCF propagation modes (i.e. Mode I and Mode II). Processing the SIF in accordance with Paris Law and Maximum Tangential Stress (MTS) theory, the crack growth rate and branching direction are derived. Prior knowledge of the final crack length in relation to growth rate is utilised for residual life estimation.

1.1.5 Thesis structure

Chapter one introduces the PhD research work and establishes the specific aim(s), intermediary objectives, and contributions of the PhD project. An overview of the methodology is highlighted, including the background of RCF defects and fracture behaviour in rails. The chapter is concluded with an articulation of the trends, needs and significance of more intelligent and self-learning condition monitoring systems.

Chapter two describes the current rail inspection methods in industry focusing on NDT's. These methods are described in such a manner that highlights the best choice of delivering the aim and objectives of the PhD project. The chapter further examines initiation, fracture mechanism, and propagation prediction of RCF damage in rails. The discussions highlight the best suited model choice for linking fracture behaviour with NDT. As a conclusion the yet to be explored benefits of incorporation of detection and prediction model is elaborated.

Chapter three introduces the details of the proposed intelligent image processing algorithm for detection and data generation of surface artefacts. It also demonstrates the application of the proposed algorithm on both field and laboratory image data. The chapter is concluded with the observation and findings related to performance of the proposed algorithm.

Chapter four discusses the development of local fracture mechanics model, which is a combination of the stress concentration at the crack tip acquired from global contact model and the mathematical models representing tension and shear fracture mechanism (both

uniaxial and bi-axial loading cases). The conclusions and findings are presented after extensive parametric studies related to defect size, inclination angle, and track curvature.

Chapter five explains the details of linking the NDT and fracture prediction models using LiveLink client server. The consequence of adopting the actual crack geometry for fracture predictions as opposed to elliptical equivalent representation is a major contribution to knowledge presented in this chapter. The chapter is concluded with the identification of the enablers and blockers of adopting this technology in industry.

Chapter six presents the conclusions of the entire thesis, which serve as the basis of the reconditions and future works that can be explored to further improve the readiness level of the proposed technology.

1.2 Rolling Contact Fatigue damage in rails

Rail Safety and Standards Board (RSSB) Vehicle/Track System Interface Committee (V/T SIC) designates RCF as a major challenge in rail industry, in the sense that hasty rail replacement requires complex project planning, preparation, logistics difficulty, and unnecessary costs including service disruptions. In the UK, over 30,000 km of track is maintained by Network Rail, whose latest annual return report concludes that RCF is the main cause of rail surface and sub-surface initiated defects. These defects, isolated or continuous, are characterised as any failure necessitating repair or replacement for improved performance and conformity with Network Rail standards, Network Rail Limited, (2016). On the rail head, the level/severity of RCF damage especially along the rolling band is described by lengths of defects. Rail maintenance engineers have generally identified four distinct levels of severity namely- light, medium, heavy, and severe. However, in strategic points such as switches, crossings, stock rails, closure rails etc. is based on the frequency occurrence within the length of the component(s). Maintenance teams are advised to treat light and medium RCF damage

via cyclic inspection and periodic grinding, with no additional action required. It is only in the case of heavy and severe RCF that further immediate maintenance action may be required.

In addition to severe contact conditions at the wheel-rail interface, the complex influence of contact geometry, vehicle suspension characteristics, traction forces and radial curving forces, especially after large loading cycles can be responsible for accumulated dislocations called Persistent Slip Bands (PSB) in the form of extrusions and intrusions as depicted in the Figure 1.1 below.
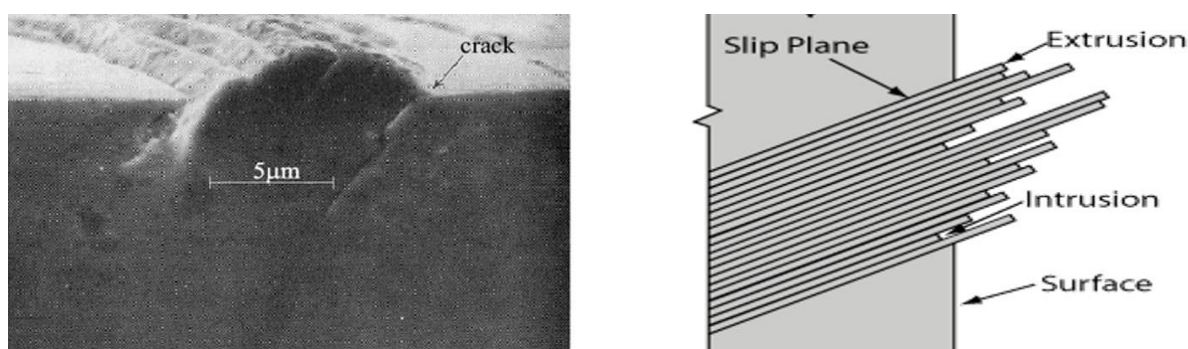
Figure 1. 1:Crack initiation by slip and extrusion planes. (Sangrid, M.D. 2013).

Burstow, M.C. (2006), described three modes (Mode 0, Mode 1, Mode 2) of generating RCF after extensive UK site specific studies related to vehicle dynamics and contact forces associated to these sites. Steady state forces related to curving (Mode 0) and discrete irregularities in track alignment especially when wheel set attempts to negotiate a piece of track (Mode 2) are most dominant cause of RCF on rails. More recently, Rovira, A. et al, (2011), explains that RCF damage is governed by the level of energy dissipated in the contact patch (T_γ), where a predominant cause of RCF is observed to be as a result of longitudinal creep force (especially if the force is in the direction of traction). For surface cracks and plastic deformation in rails, plastic ratcheting (which is relative to the tangential load) is said to be the main cause in Jones, R. Molent, L. Pitt, S. (2007).

The different types of RCF defects have been aptly reviewed by several authors over several years including the recent works of Magel, E. (2011) and are summarized briefly in the following sections:

- Squats

As described by Wilson, A. & Kerr, M. (2012), these are surface defects initiated by RCF when a pair or more of leading and trailing cracks propagate in opposite directions. The cause of such a phenomenon has been related to the influence of high traction forces and micro-slip.
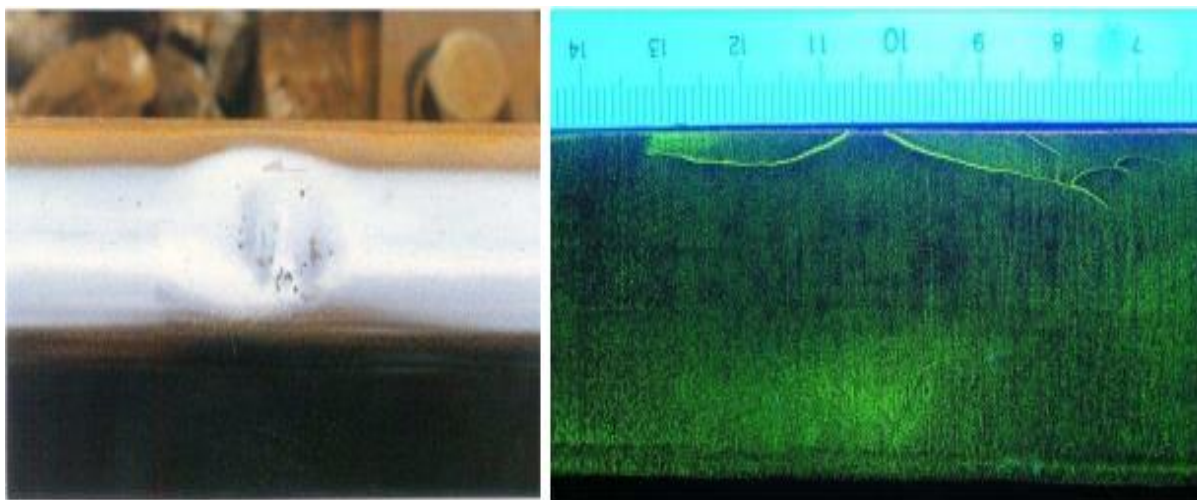
Figure 1. 2: The cross-section squat type crack with the long leading crack (in rolling direction) and the much longer trailing crack in the opposite direction (Magel, E. 2011).

- Gross plastic deformation

As depicted in Figure 1.3, these defects are as a result of a considerable deterioration and loss of structural integrity of rail head material especially under heavy loading cycles. These combined, may result into vertical irregularities on the rail head that is subject to more plastic deformation due to impact loads from passing wheels.

Figure 1. 3: Depicts a crushed head rail defect (Magel, E. 2011).

- Rail Gauge Corner Cracking (GCC)

Although GCC occasionally appear on tangent rails, and rarely seen on low rails of curved track, such defects are caused by high wheel-rail contact stresses in addition local shear stresses due to slip between the wheel and rail. GCCs have been characterized geometrically as being fairly equally spaced along lengths of track or clustered due to track geometry irregularities usually for long lengths of track section.
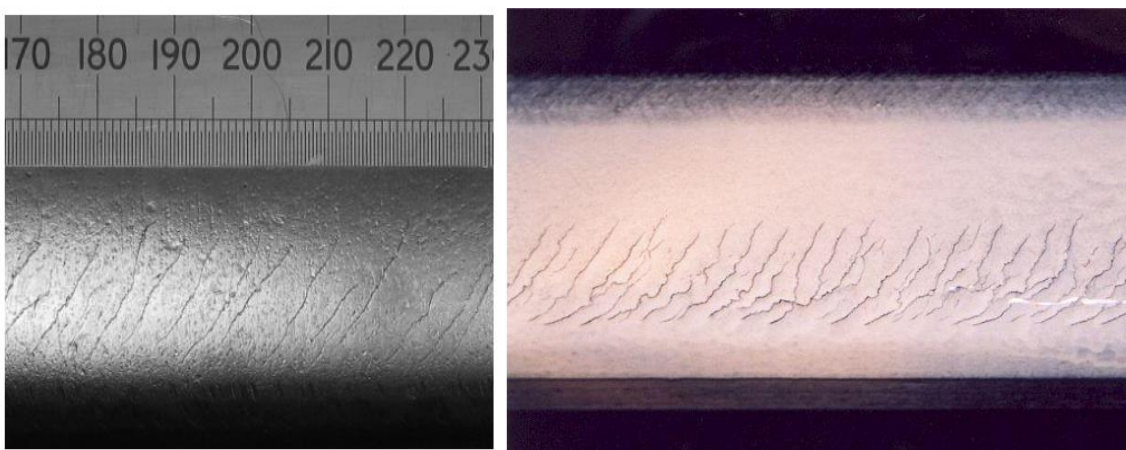
Figure 1. 4: Depicts a variety of GCC defects (Magel, E. 2011).

- Rail Transverse defects

Transverse cracks are initiated as a result of shells in the presence of metallurgical discontinuities, in the form of oxide stringers. These cracks propagate due to bending stress, residual stress, and thermal stress from rail contraction especially during the winter season. Field observations suggest such defects are the major cause of many broken rails and subsequent derailments if not curtailed.
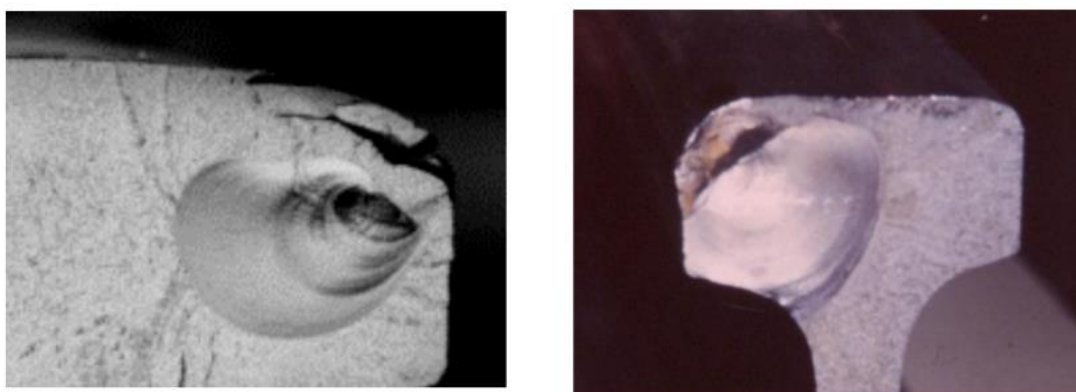
Figure 1. 5: Depicts transverse defect in rails (Magel, E. 2011).

1.3 Linear elastic fracture mechanics (LEFM) in rails

The existence of fatigue damage on rails is mainly responsible for significantly reducing the useful life of the rail infrastructure in addition to safety issues. The complexities related to fatigue and fracture can be observed on tangent track, and but more pronounced on curved track sections. RCF damage is also observed in sections of track with bi-directional operation, including rail butt welds, and switches. These complexities (related to fatigue and fracture) have been attributed to a combination of highly multifaceted dynamic loading, primary and secondary stresses including varying environmental conditions. To accurately model useful life estimations of rail assets in the presence of a crack requires perfect understanding of Stress Intensity Factor (SIF) around the vicinity of crack tip, the critical fracture toughness of the material, and the applicable fatigue crack growth rate expression, and finally the initial and final crack size.

Inglis in 1913 investigated the stress multiplier effect at the edge of a thin plate and established the relationship between the amount of stress magnification (at the crack tip) and radius of curvature for the crack (absolute crack size). A few decades later Griffith of the Royal Aircraft Establishment in England, concerned by the discrepancy between the conclusions of Inglis and the fact that larger cracks tend to grow faster than smaller ones furthered the investigation. He concluded that the amount of strain energy released must be greater than or equal to that required for the surface energy of the two new crack faces formed as a result of crack face extension. Experimental works of Irwin during the Second World War mathematically described the stress filed in the vicinity of the tip as a function of the applied remote stress and square root of crack length. Depending on the mode of fracture in consideration see Figure 1.6 below. the SIF is said to be a numerical estimate of the magnitude of the effect of stress singularity at a point around the crack tip vicinity.
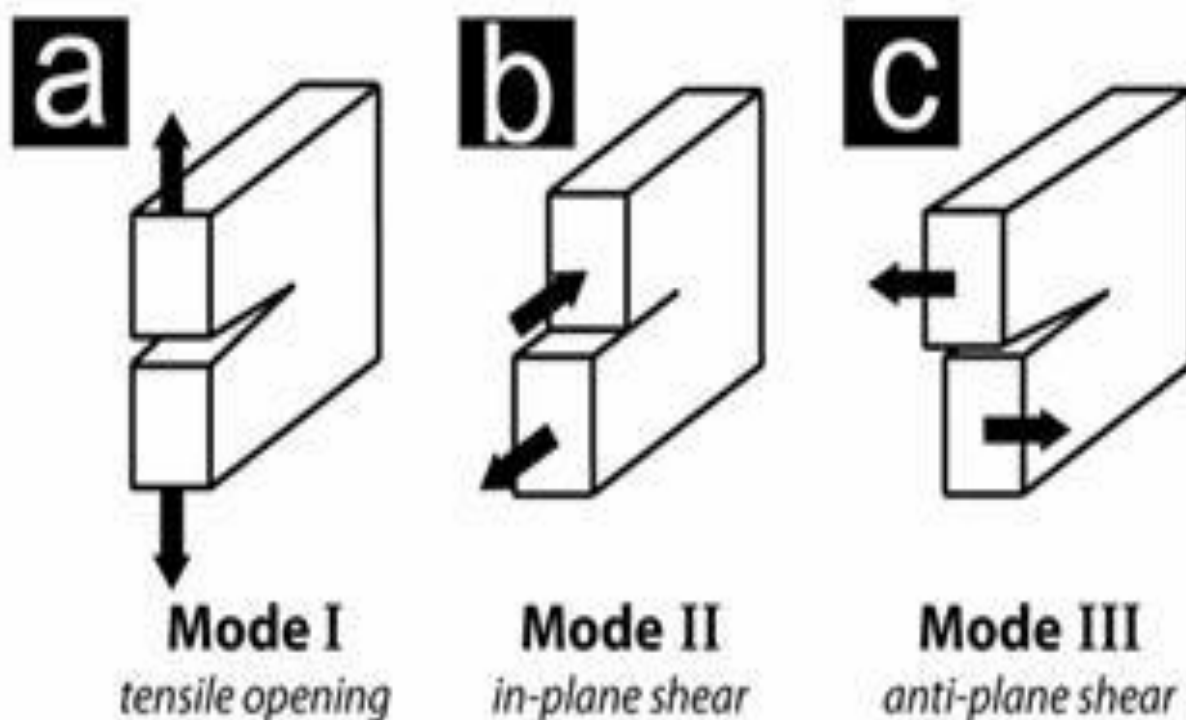
Figure 1. 6: Shows the tension (a), shear (b), and torsion (c) modes of fracture. (Kundu, T. 2008).

Consequently, the value of SIF at which the crack extension is observed defines the beginning of crack propagation. In practice Kundu, T. (2008) explains that the SIF is limited to the material yield stress thus a violation of the linear elasticity assumptions can be avoided by using Elastic-Plastic fracture mechanics (EPFM) approaches to describe fatigue and fracture process. However, in this write up the LEFM approach has been adopted to govern the behavior of the rail crack growth by assuming that the SIF is always less than material yield stress. Vasic, G. (2013) presented details of fatigue behaviour under cyclic loads under LEFM assumptions.
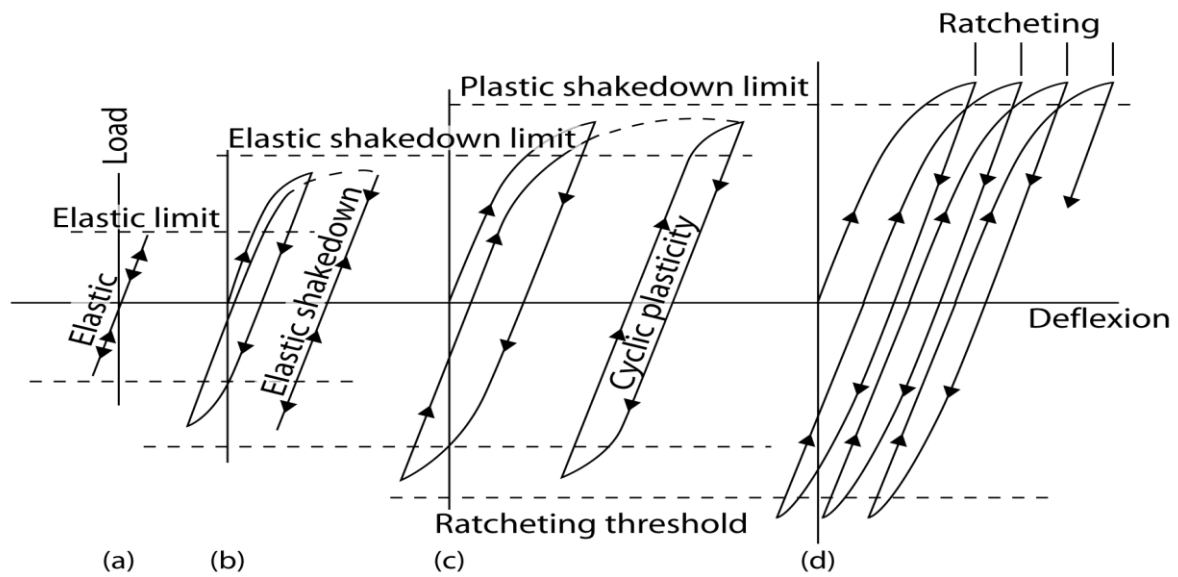
Figure 1. 7: Shows the material response to cyclic loading (Vasic, G. 2013).

As described by Figure 1.7 above, the cyclic loading of wheels on rails causes the rail material to respond either as perfectly elastic or plastic. This is only true given that yield stress of the material is not exceeded. It is also possible for the rail material to experience elastic shakedown, such that initial plastic deformation occurs, after which the material hardens and behaves elastically onwards. This behaviour occurs when the material is loaded above elastic limit but under the elastic shakedown limit. When material is loaded above elastic shakedown limit, a closed elastic-plastic loop with no net accumulation of plastic deformation results. In addition, cyclic loads above this plastic shakedown limit, causes rail material to accumulate unidirectional plastic strain (plastic ratcheting). From this behavior of rail response to cyclic loading, Plu, J. et al (2009) explains that three phases of crack propagation in rails depicted in Figure 1.8 below.
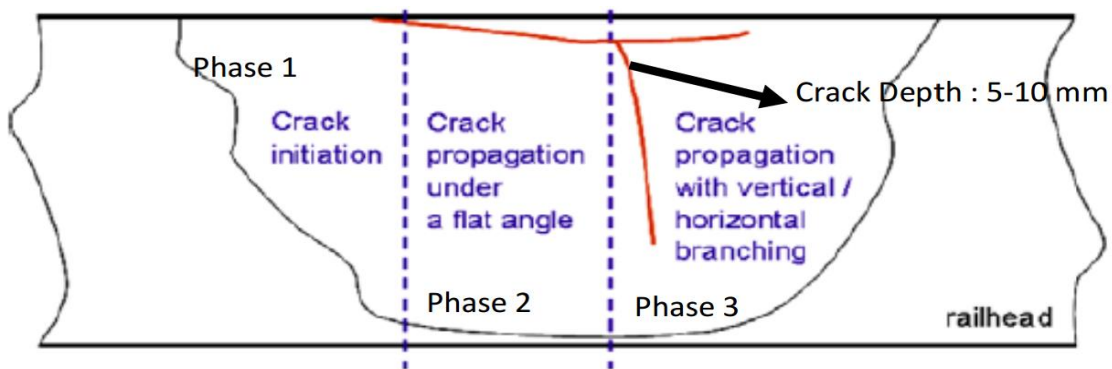
Figure 1. 8: Rail crack development (Plu, J. et al 2009).

Phase I is mainly as a result of strain accumulation beyond critical material ductility often indicated by shear in material (in the direction of motion). At this phase, cracks are characterized by early initiation, and decrease in growth rate as crack depth increases. Such cracks are caused by train axle loads, including the effect of longitudinal and lateral creep forces at the contact patch. These cracks usually appear following grain boundary of the material microstructure. In the second phase (Phase II) shallow angle crack growth occurs at this stage with propagation angle of $10^0 - 20^0$ relative to contact surface. The growth rate at this phase is observed to be proportional to contact stresses. There is sufficient evidence that the presence of fluid facilitating sliding of crack faces, increases in rate of propagation is observed. The final phase (Phase III) is also called the branch crack growth stage, where crack branching is caused by bending and residual stresses developed during rail manufacture. In this phase cracks branches up or down depending on the direction of the dominating stress. Propagation at this stage is very fast, and rapid fracture leading to rail breaks is possible.

1.4 Chapter conclusion

The specific aims and objectives of the PhD project have been discussed in detail. The chapter also introduced the background of the problem to be solved in this research work, which emphasises the positive impact of the yet to be explored benefit of incorporating detection and prediction models within a single process in railway condition monitoring. The fundamentals of RCF damage, including initiation, fracture mechanisms, and propagation behaviour are highlighted along with the fracture parameters that are essential in residual life estimation. An articulate understanding of the current industrial challenges, trends, and needs related to rail RCF damage detection and prediction will be discussed in the literature review presented in chapter two.

Chapter 2 Literature review

A detailed literature review was undertaken across a number of areas with regards to current rail inspection methods in industry, with specific interest in existing image processing algorithms adopted for rail inspection. The review further identifies the existing models adopted for predicting RCF damage in rails. This review synthesises key concepts, findings and arguments outlined in earlier studies in order to identify relevant issues and establish a theoretical framework for the proposed novel incorporation of detection and prediction models. Journals, books, industrial reports, and patent publications have been used for this review.

2.1 Review of rail surface inspection methods in industry

It is undeniable that the future of rail maintenance for guaranteed efficiency and safety is being challenged on a multifaceted platform including increase in loading, capacity, frequency of travels and climate changes. In the past decade, several research projects sought to explore and develop novel techniques capable of accurately detecting defects with improved flexibility, speed and accuracy of tests. Such objectives necessitate detailed survey of existing rail NDT to provide the required insight for assessing the suitability and effectiveness of new inspection technology for predictive maintenance in industry as presented in Sawadisavi, S. et al (2009) and Edwards, J.R. et al (2009). The significance of more intelligent or even hybrid systems has been recommended in several EU funded projects including but not limited to CRAFT Project, Rail-Inspect. This persistence in improving NDTs is mainly because no single method is capable of providing all the required information on the condition of rail infrastructure. In addition, Innotrack, (2010) in its concluding technical report laments on the practical tediousness of controlling sensor lift off variations in contactless probe applications while frequent cases of wear/breakage in the case of contact sensors is not favourable. These complexities have encouraged an outlook of development and deployment of more intelligent and self-learning rail inspection systems by most rail condition monitoring service providers including Deutsche Bahn (DB), SGS, Depotrail, amongst others. Such intelligent and self-learning RCM systems are capable of acquisition, processing, and generation of statistical

data engineered for better predictive infrastructure management and track engineer planning. The following conventional NDEs are elaborated with primary focus on operational principle, advances in the technology and possible novel applications.

2.1.1 Ultrasonic rail inspection

Since 1950s, ultrasonic tests (UT) have been performed on rails. The detection of defects are functions of amplitude and arrival time of reflected bulk waves due to the interaction of an electric charge with a piezo electric crystal. The method is capable of detecting both surface and subsurface defects by simply changing frequencies of excitation current. To-date this method remains one of the most popular in the U.K. Advances in technology have resulted into the use of portable instruments ranging from hand-held, and rollers. In Hesse, (2007) conventional ultrasonic contact probes are said to be most effective at detecting only subsurface defects at high speeds (above 75 km/hr). For contactless probes (with pre-determined frequencies optimal to defect types based on FEA simulation), it was suggested in Han, S. et al (2015) as a solution to practical challenges of electromagnetically generating and measuring waves in rails. Kenderian S et al, (2003) proved a concept that combines laser generation with air-coupled detection of ultrasound as the future of ultrasonic testing.

2.1.2 Eddy currents (EC) inspection of rails

Eddy current (EC) methods are based on interaction between eddy current sensor's magnetic fields and the homogeneity in rail steel material. Song, Z., et al (2011) concluded that surface roughness of rail corresponds to amplitude change in sensor field, while phase changes indicate surface crack. Improvements in EC NDE instruments have been significantly focused on the suitability of probe sensor vis-a-vis the test requirements as highlighted by most NDE manufacturers such as OLYMPUS. From measurements of either absolute probe (single coil), to that of differential probes (multi-active coil), the presence of a defect is established by either absolute or differential change in impedance. In a more recent publication by Lui, Z., et al (2013) a classical four-arm bridge with two known impedances utilised as an EC sensor such that the bridge differential is optimised for rail NDE using a digital lock-in amplifier algorithm

(for improved sensitivity of detection). The severity of sensor lift-off effect can be dynamically compensated by summing the signals of the detection coils.
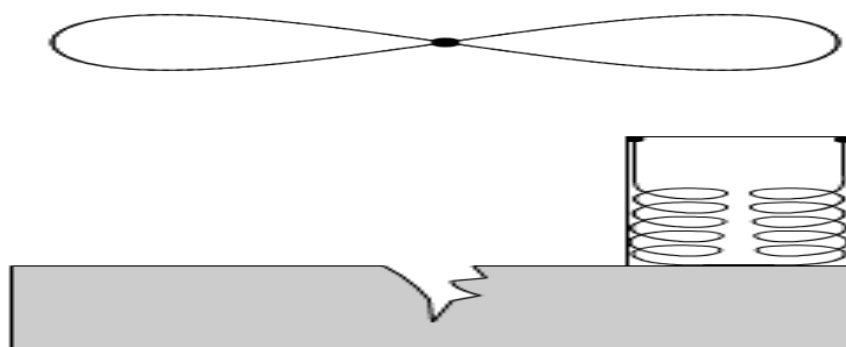
Figure 2. 1: Differential probe generates differential signal only when one end of the coil is over a defect free region and the other is at a defected position.

Garcia-Martin, J., (2011) discusses limitations related to: indirect estimation of defect depth, adverse effect of sensor lift-off variation (no more than 2mm), electrical conductivity and the magnetic permeability of the rail, and the signal to noise ratio as among the most influential variables that affect the performance of EC. In spite of this observation Network Rail in the U.K, confidently reports an increase in RCF damage on the network based on eddy current measurements more than previously reported by UT method.

2.1.3 Automated vision systems (AVS) for rail inspection

The availability of improved Image Acquisition Systems (IAS), even customised depending on the nature of inspection to be performed, has enabled automated vision systems to operate at very high velocities (speeds up to 100km/h). The type of test is observed to influence inspection speed, with the slowest being wheel burn tests at speeds less than 10km/hr. The flexibility and remoteness of AVS has attracted more curiosity according to recent review of applied AVS in rail industry as highlighted in Resendiz, E. (2013).

In this regard, Teng, Z., Lui, F., & Zhang, B. (2016) demonstrated the detection of railway track performed by Simple Linear Iterative Clustering (SLIC or super-pixels) based on colour similarity and proximity of pixels within an image. This detection algorithm is comprised of a weighting factor (TF-IDF) that indicates the significance (or frequency) of a particular super-pixel within the image and in the overall data set. True detection is said to be achieved (green

highlight depicted in Figure 2.2) when the super-pixel in consideration is repeatedly occurring in the image but low within the overall image data set.
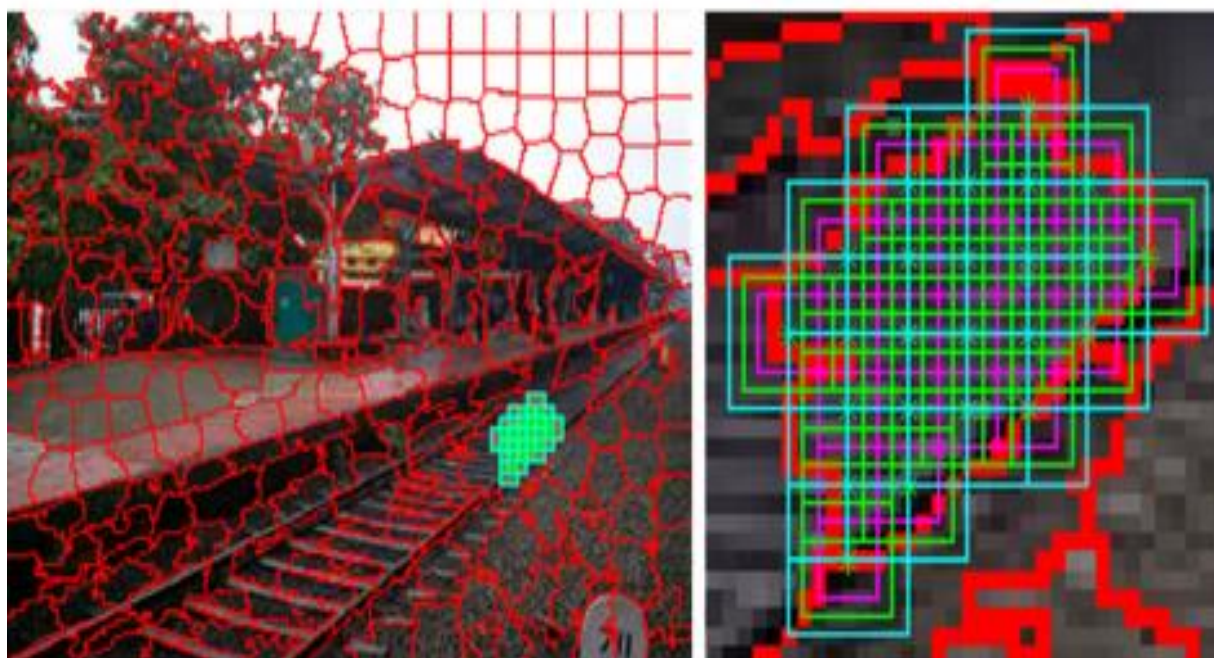
Figure 2. 2: Object detection using super pixels (Teng, Z., Liu, F., & Zhang, B., 2016)

Detection by forecasting governed by a combination of exhaustive search and jump search algorithms is discussed in Joy, G., Hyfa, N., Krishnan, R. (2014). In this method a window, exhaustively scans the image and true detection is said to coincide with the occurrence of left and right bolts at the same lateral axis (at the same y-coordinate). While a false detection occurs when the system fails to find an object in a forecasted position.

Resendiz, E., et al (2010) considered safety regulations in detection of irregularities and defects in rails, and proposed a Machine Vision System (MVS) using video data, adjustable lighting parameters, and virtual track model calibrated camera angles. The data is processed with a combination of edge and texture algorithms. Somalraju, S. et al (2012) presented a Robust Railway Crack Detection Scheme (RRCDS) using LED-LDR MVS. This MVS uses DC motors for each rail roller and an Audrino microcontroller ATMEGA 328 for deciding true crack detection based on photovoltaic properties of LDR sensor. The location (longitude and latitude) of the crack is determined using a combination GSM and GPS satellite systems. Jian-hua, Q., et al (2008) discussed the design of rail surface crack detection hardware using a TCD1208AP Linear Charge Coupled Device (LCCD) image sensor. The image processing

algorithms are executed on a DSP module based on noise elimination, edge detection, image segmentation and a warning system. AVS inspection of artefacts was discussed by Lui, Z., et al. (2010) in their publication focusing on spalling of rail head and general surface cracks. Algorithms comprised of algorithms related to de-noising, image feature extraction (region width and region position) applied in processing the acquired images of the rail. Further dynamic thresholding of ROI and feature matching is performed. The system is of in-service type with CCD line scan IAS coupled with embedded data processing system for the estimation of percentage of rail head wear and length of cracks on rail surface as an evaluation of flaw on rails. Li, Q., Ren, S. (2012) proposed an illumination invariant MLC contrast enhancement which offers normalised truncation of background and foreground pixels based on a combination of pixel intensity and local window mean values. Furthermore, Proportion Emphasised Maximum Entropy thresholding of discrete defects was performed by analysing the cumulative probability of a specific gray level, such that a normally distributed probability function is clearly divided into object (Po) and background (PB) probabilities. In addition, the threshold of segmentation is selected as the maximum total information provided by Po and PB respectively.
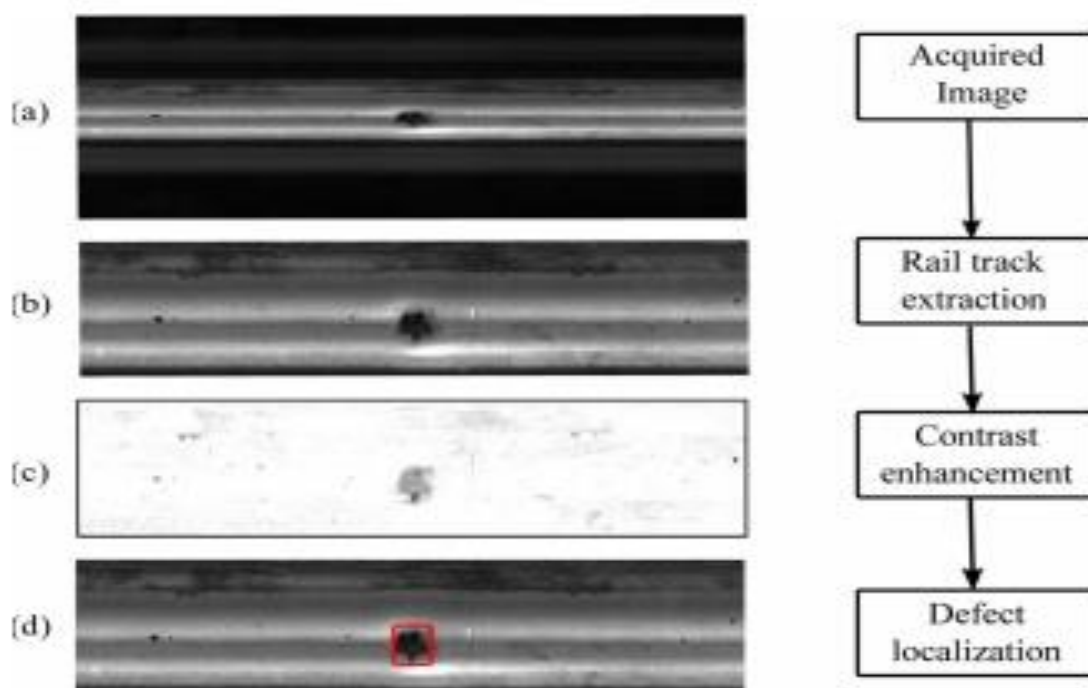
Figure 2. 3: Diagram of image analysis subsystem (Li, Q., & Ren, S. 2012)

Enhanced inspection speed and repeatable measurements have been ensured through simultaneous application of vision systems with vibration-based methods for fault detection and track data calibration as explored in Belkhade, A. & Kathale, A. (2014).

2.2 Existing image processing algorithms for rail condition monitoring

This section of the thesis presents the variety of the fundamental image processing algorithms that have/can been/be successfully applied to rail images for condition monitoring purposes. The algorithms are for enhancement, segmentation, and feature extraction.

2.2.1 Pre-processing

Image enhancement (pre-processing) are techniques with the principal objective of improving interpretability or perception of information within an image to suit a specific application. There exists a wide variety of enhancement techniques that are broadly categorised into spatial and frequency domain methods adopted for dynamic range modification, noise reduction, and edge enhancement. The appropriateness of a particular choice is influenced by the imaging modality, task at hand, image resolution and illumination/reflection properties of the rail surface.

2.2.1.1 Spatial domain enhancement method (SDEM)

This enhancement method involves the direct manipulation of neighbourhood pixel values (both linear and non-linear) in an image for enhanced overall perception of information. Bedi, (2013) has described conventional SDEM as follows:

2.2.1.1a Point processing operation (PPO)

Also called simple intensity transformation, a variety of which are depicted in Figure 2.4 below. These methods often described as manipulations on the intensity of individual pixels, the following PPO's have been applied in rail image processing.
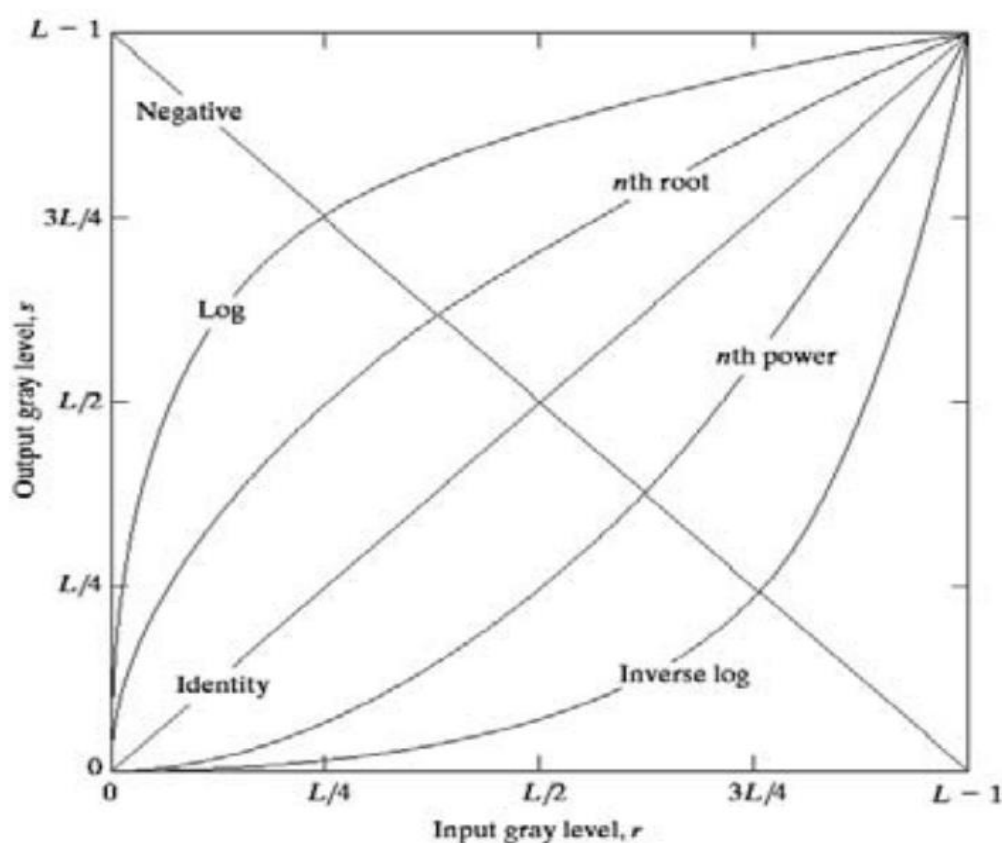
Figure 2. 4: Types of point processing operations on images. (Singh, G,M., 2013)

- Negative

The most basic and simple PPO is the negative g(x,y) transformation, expressed as the variation between each pixel intensity value f(x,y) to maximum intensity L as elaborated in equation 2.1 below.

$$g(x,y) = L - f(x,y) \qquad (2.1)$$

- Log transformation

Log transformation entails replacing each pixel 'l' with its logarithm as described in equation 2.2. For a constant value of c, with spatial coordinate 'r' positive, the model transforms a narrow range of low intensity values to wider ranges. Also, wide range of high intensities values are mapped to narrower ranges. The resulting image g, expands the values of dark pixels while compressing bright pixel intensity values.

$$g = c \times log(l + r) \tag{2.2}$$

- Piece wise linear transformation

The piece-wise linear transformation function enables modification of dynamic range of intensity values by changing the slope and intercept of user defined linear function as depicted in Figure 2.5. The complicated nature of optimising dynamic ranges, be it global or local is mainly dependent on the distribution of intensity values within the input image.

(a)

Mapping Function

(b)

(c)
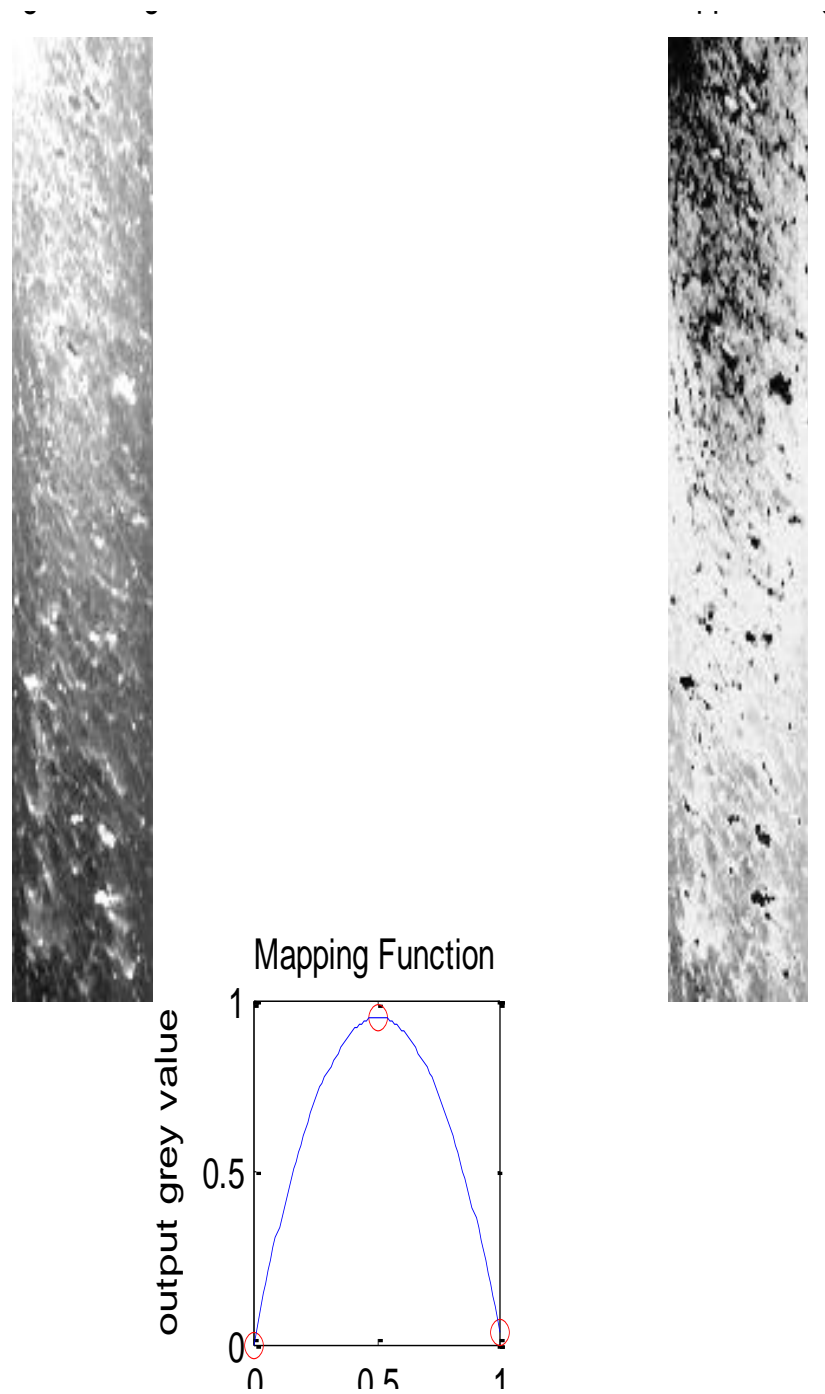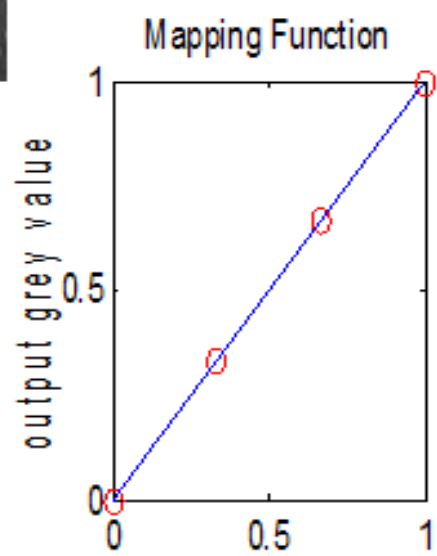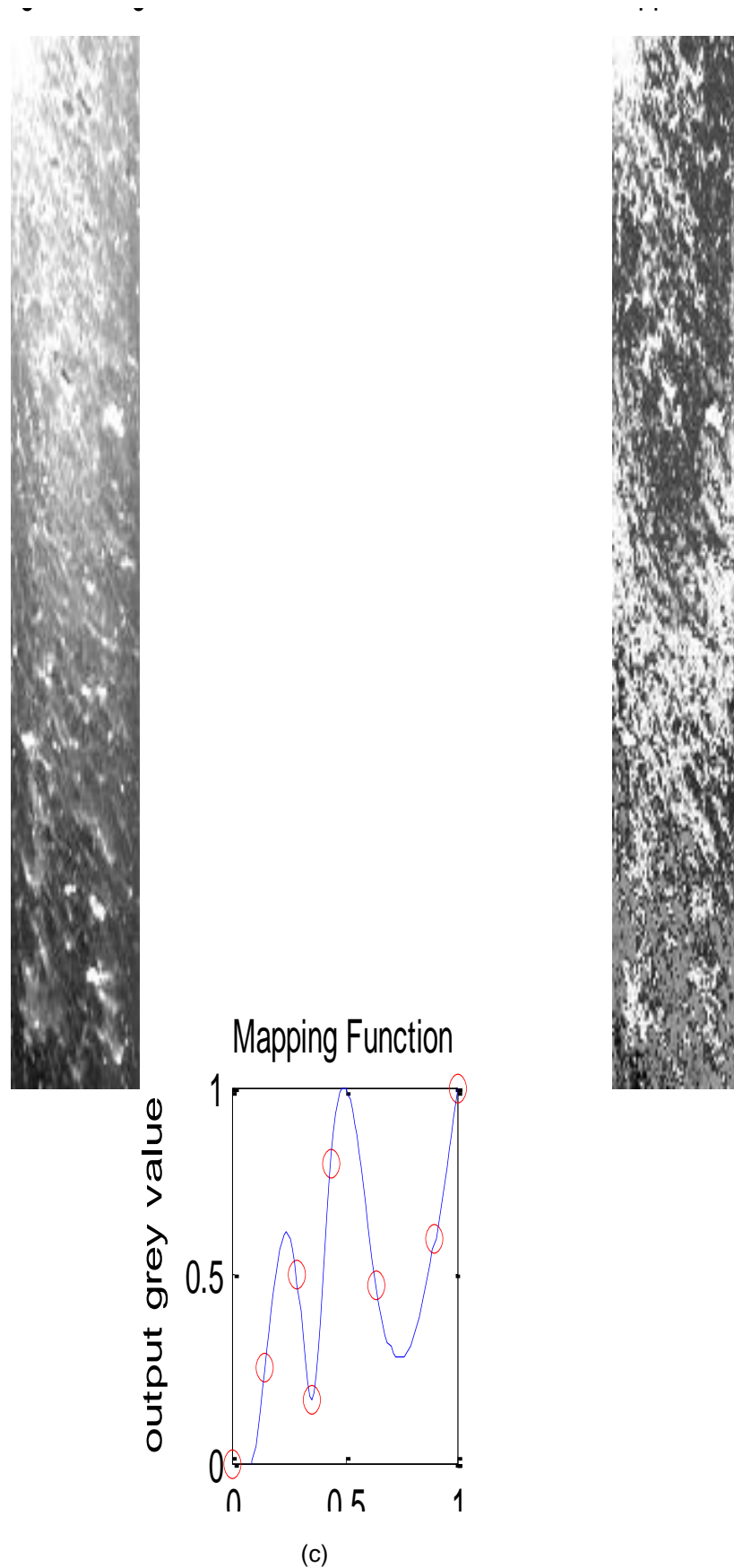
Figure 2. 5: MATLAB piece-wise transformation function applied on the original Heavily damaged rail surface image (on the left), such that on the left of(a) shows 3 control points, while left of (b) shows 4 control points, and to the left of (c) shows 8 control points; spline functions defining mapping of input to output intensity values.

2.2.1.1b Spatial (M x N neighbourhood) operation

Most image processing applications including document binarisation, medical image analysis, facial, and fingerprint recognition take advantage of the improved computational speed, and adaptive nature of spatial enhancement. The following types of spatial operations have been applied for rail image processing:

- Adaptive normalisation method

Both global and local enhancement using normalisation is achieved by scaling and stretching the range of grey level values to span a desired range of grey level intensities for revealing more clearly the local dark or bright regions of an image. Kim, B. & Park, D. (2002) discusses the implementation of equation 2.3, initially applied for normalisation of fingerprint images.

$$G(x,y) = \begin{cases} M_o + \sqrt{\dfrac{VAR_0(I(x,y)-\widehat{M})^2}{\widehat{VAR}}} & I(x,y) > \widehat{M} \\ M_o - \sqrt{\dfrac{VAR_0(I(x,y)-\widehat{M})^2}{\widehat{VAR}}} & otherwise \end{cases} \tag{2.3}$$

Where M_o and VAR_0 are the desired mean and variance respectively while $\widehat{M} \ and \ \widehat{VAR}$ are the computed mean and variance of input image, and can be evaluated locally (in a window of known size) or globally (for entire image).

In Xie, X., & Lam, K. (2006) a normalisation method which focuses on railhead illumination invariant enhancement according to the formulation in equation 2.4, is observed to decrease in the dynamic range of the grey scale between foreground and background pixel intensity values.

$$G(x,y) = \frac{F(x,y)-E(x',y')}{varF(x',y')} \tag{2.4}$$

Where $E(x',y')$ is the mean of $F(x',y')$ and (x',y') is the spatial coordinate of the local window W. F(x,y) is the original image intensity value at position specified by (x,y) related to the neighbourhood of W.

Figure 2. 6: Image enhancement by normalization.

Besides the mean and variance of the input image, Lui, Y., et al (2013) explains that illumination invariant normalisation can be achieved using gradients in the image as weights in averaging operation on intensity values.

- Spatial filtering enhancement

Spatial filtering of images is based on kernel or mask processing of aggregates pixels which has been demonstrated on crack images by Zhang, W., et al (2014) such that the centre pixel of each neighbourhood is enhanced in relation to the pre-determined origin of the mask. Depending on the operation to be performed, spatial filters are broadly classified into Linear and Non-linear filters, and in both cases either convolution or correlation operators are utilised for processing an MxN kernel coefficients on an M by N image. Singh Negi, S. & Gupta, B. (2014) explains that linear filters are mainly used to reduce noise or produce less pixelated images there by preserving details.

$$A = \frac{\sum f(x,y)}{n} \tag{2.5}$$

Where f(x,y) is the pixel intensity within a specified local window, and n is the number of pixels in the local window.

Weng, W. & Chen, H. (2015) explains smoothing of low gray values, to cushion the impact of environmental contaminants by using a cascaded combination of Linear Mean Filter (LMF), standard deviation based offset (for noise compensation), and image subtraction as Figure described in section 3.3.1.1.
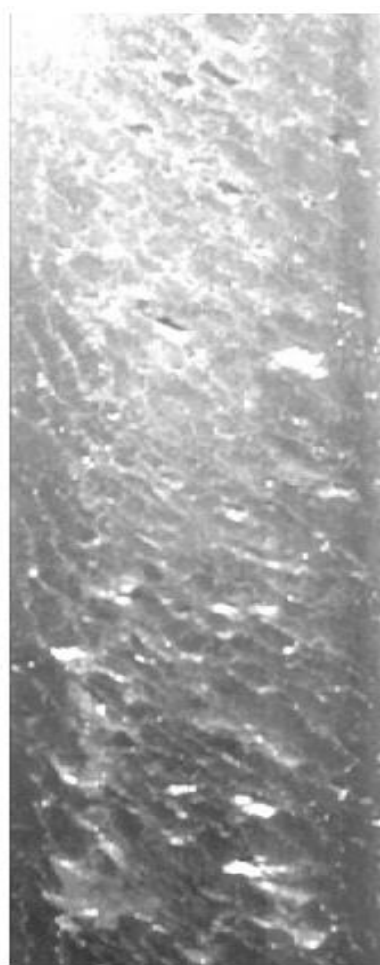
Figure 2. 7: Output of linear averaging filter on heavily damages rail surface image.

Non-linear order static filters are based on ordering statistical ranking of the pixels contained in a kernel. The value of the centre pixel (within the mask) is replaced by the output of a specified statistical operation. Median, Max, and Min filters response of a heavily damaged rail image is depicted in Figure 2.11 below. Simulated results confirm the conclusion of

Gonzalez, R,C., & Woods, R,E. (2005) that the median filter is the most robust at preserving sharpness of the original image.

<div align="center">(a)   (b)   (c)</div>
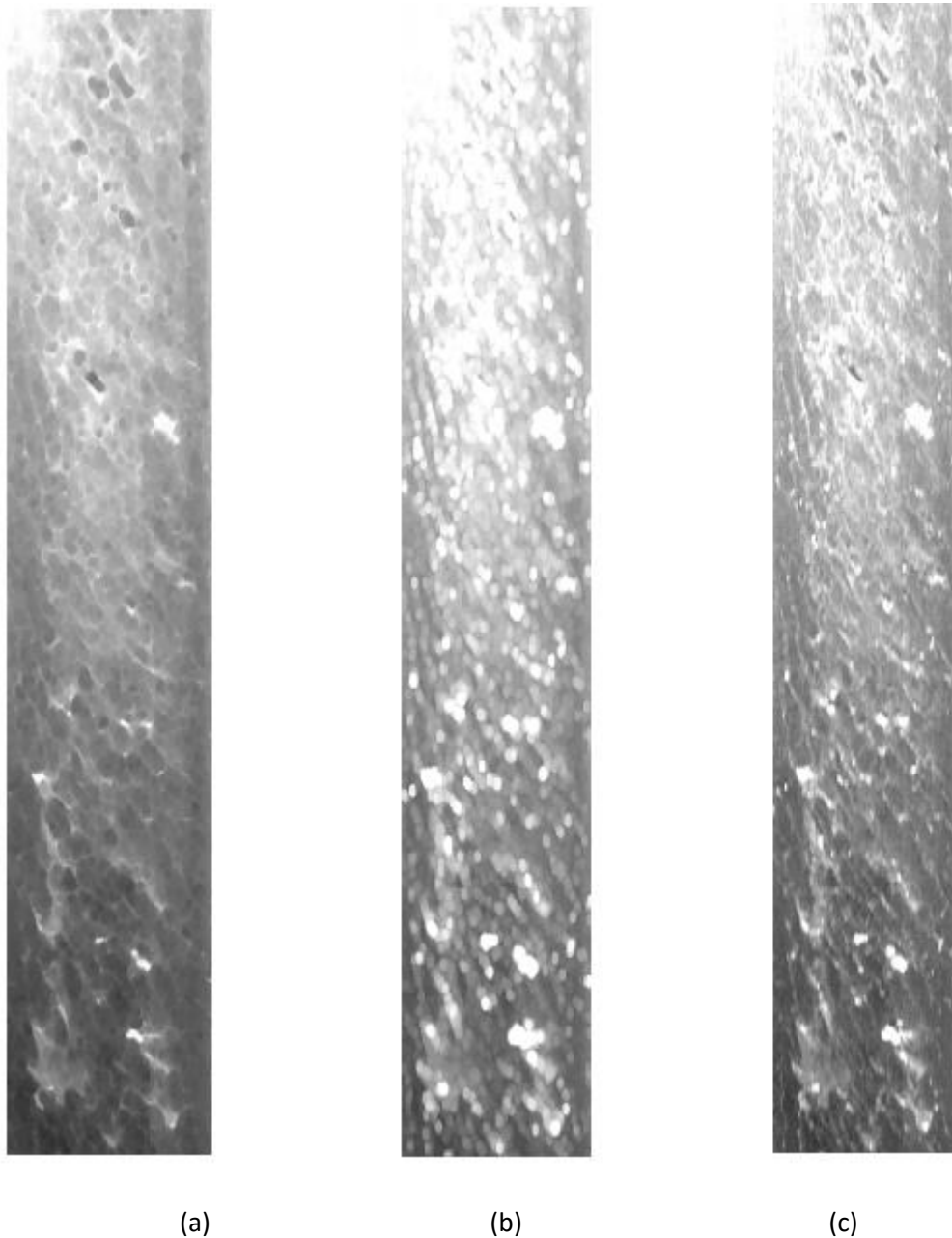
Figure 2. 8: Oder static filters applied on heavily damaged rail surface image; (a) Min filter, (b) Max filter, (c) Median filter.

- Histogram processing operation

Histogram equalisation is a well-known contrast enhancement technique achieved by altering the spatial histogram of an image to closely match a uniform distribution (p_n) of equation 2.6, based on Cumulative distribution function (CDF).

$$p_n = \frac{number\ of\ pixels\ with\ intensity\ n}{total\ number\ of\ pixels}, n = 0,1,2,\dots,L-1 \qquad (2.6)$$

The histogram equalised image g(x,y) for an input image f(x,y) is expressed in equation 2.7, where L is the number of possible intensity values mostly 256 for gray scale images.

$$g(x,y) = (L-1)\sum_{n=0}^{f(x,y)} p_n \qquad (2.7)$$

Global Histogram Equalization (GHE) in Figure 2.9, is an example of such conventional enhancement, and its performance is adversely hindered by high peaks or large spatial variation. As a solution, Local Histogram Equalization (LHE) is proposed, which is a neighbourhood based approach but is concluded to induce undesirable checkerboard effects. In Singh, G.M., Kohli, M.S., & Diwakar, M. (2013) a Contrast Limited Adaptive Histogram Equalization (CLAHE) that operates on contextual region of pixels and uses bilinear interpolation to eliminate artificially induced checkerboard effect is presented. In the investigations of Wang, C., & Ye, Z. (2005) extensions of HE include Brightness Preserving Histogram Equalization with Maximum Entropy (BPHEME), capable of identifying a specific target histogram that maximizes the entropy of the image. Kuber, M., Dixit, M. (2014) investigates the modifications related to enhancement applications in the past decade, taking cognisance of novel techniques such as the Grey-level Grouping (GLG) and its extended version, Selective GLG (SGLG) capable of grouping and ungrouping histogram components with specified intensity levels. A separate research investigating the so called Multi-HE was presented by Menotti, D., et al (2007), and it was concluded based on uniform decomposition of original image into various sub-images, followed by classical HE operations yielding improved results.
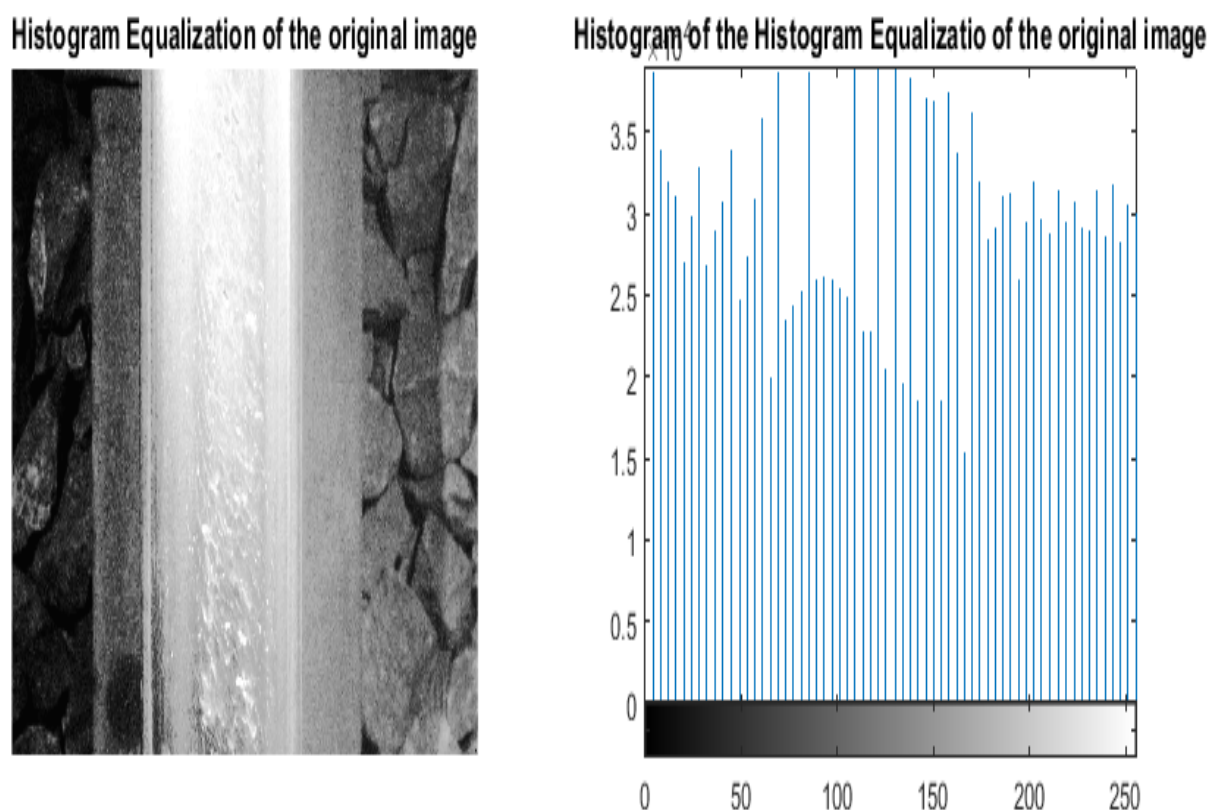
Figure 2. 9: Histogram equalisation applied on heavily damaged rail image.

2.2.1.2 Frequency domain enhancement
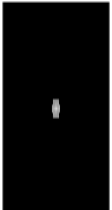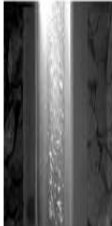
Variations from dark to bright and back to dark grey levels (or vice versa) correspond to change in frequency from High to Low values which are typical to existence of either edges, noise, or large background objects. The advantage of this approach is that, it is not limited to integral multiples of fundamental frequencies applicable to only periodic signals. Also the significance of estimated magnitude and phase response of images provide the much needed information on the relative amounts and positions of varied frequencies existing in the original image signals.

- Basic low and high pass filters

Low pass filters (LPF) are typical examples of smoothing filters that have been adopted for robust detection algorithms in railway as presented by Javed, A., et al (2012) and even for inspection of surface defects as demonstrated by Lui, Z., et al (2010). The authors suggested

that the kernel/mask size is proportional to computational time and extent of blur in the edges.

The transfer functions related to LPF are elaborated in Appendix A1.

IDEAL LOWPASS FILTER	BUTTERWORTH LOWPASS FILTER	GUASSIAN LOWPASS FILTER
Lowpass Image of the H1.jpg After Filterir Magnitude of Lowpass image Phase of Lowpass image	Lowpass Image of the H1.jpg After Filtering Magnitude of Lowpass image Phase of Lowpass image	Lowpass Image of the H1.jpg After Filtering Magnitude of Lowpass image Phase of Lowpass image

Table 2. 1: Shows the comparison of Ideal, Butterworth and Gaussian low pass filters response on a heavily damaged rail image.

High pass filtering (HPF) is another typical image sharpening technique and is concluded to suppress details of images especially in the case of kernels with positive values at origin and negative coefficients elsewhere Malik, Q. (2013). As depicted in Table 2.2 below, the filter response is observed to mostly enhance edges within an image. The transfer functions related to LPF are elaborated in Appendix A2.
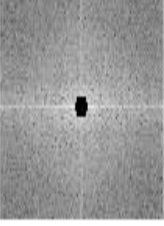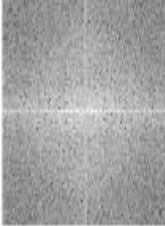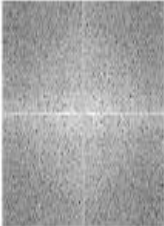
IDEAL HIGHPASS FILTER	BUTTERWORTH HIGHPASS FILTER	GUASSIAN HIGHPASS FILTER

Table 2. 2: Shows the comparison of Ideal, Butterworth and Gaussian high pass filters response on a heavily damaged rail image.

- Fourier Transform (FT) enhancement

In Kumari, A.V. (2015) the author explains the power of FT transforms in terms of its ability to synthesize non-periodic wave forms using sinusoids with continuous range of frequencies. Gonzalez, R,C., & Woods, R,E. (2005) described the method as principally based on the spatial convolution theorem which is equivalent to the more computationally efficient spatial multiplication in the frequency domain as depicted in Figure 2.10.
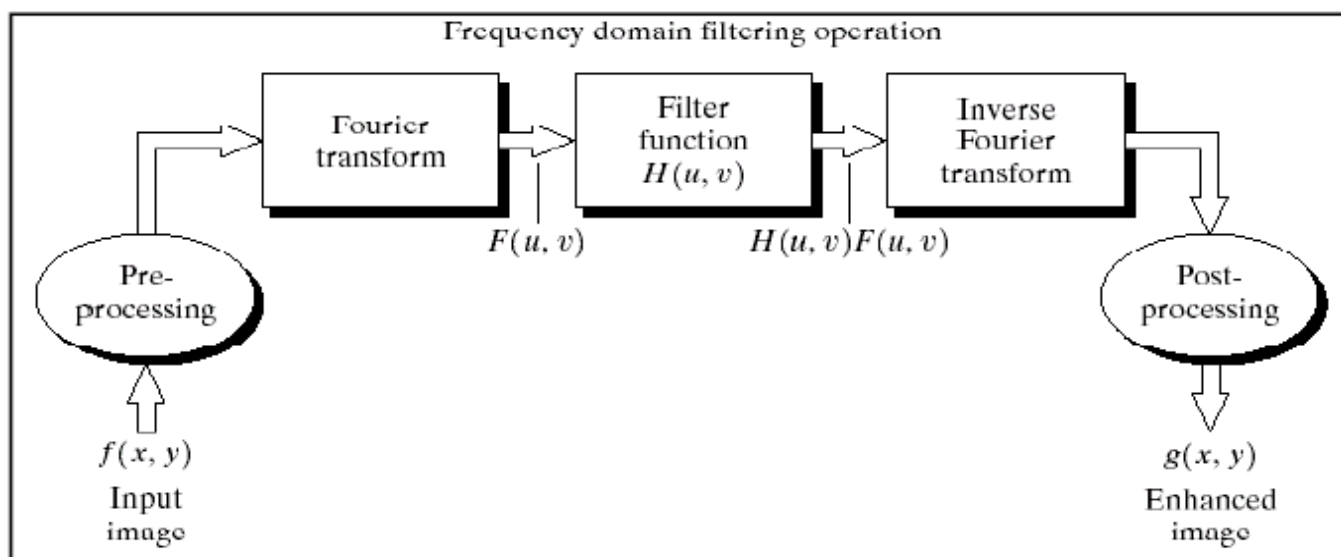
Figure 2. 10: Block diagram for FT based filtering of images (Gonzalez, R,C., & Woods, R,E. 2005).

Similar result has been obtained as summarised in Tables 2.1 and 2.2 with faster computational speed attributed to multiplication in frequency domain as opposed to time consuming convolution operation in the spatial domain.

2.2.2 Segmenting region of interest

Detection of objects of interest in image processing is mainly by partitioning the image into meaningful structures (number of partitions is subject to the resolution of object or region of interest) commonly termed as image segmentation. Zaitoun, M.N., Aqel, M.J. (2015) describes this process as not only classic but also hotspot in the field of image processing. The efficiency of segmenting object of interest is defined by the amount of connected pixels with distinct grey intensity values. In the absence of a unique solution to the problem of segmentation, there exist numerous approaches broadly categorised into region or edge based as depicted in Figure 2.11 below.
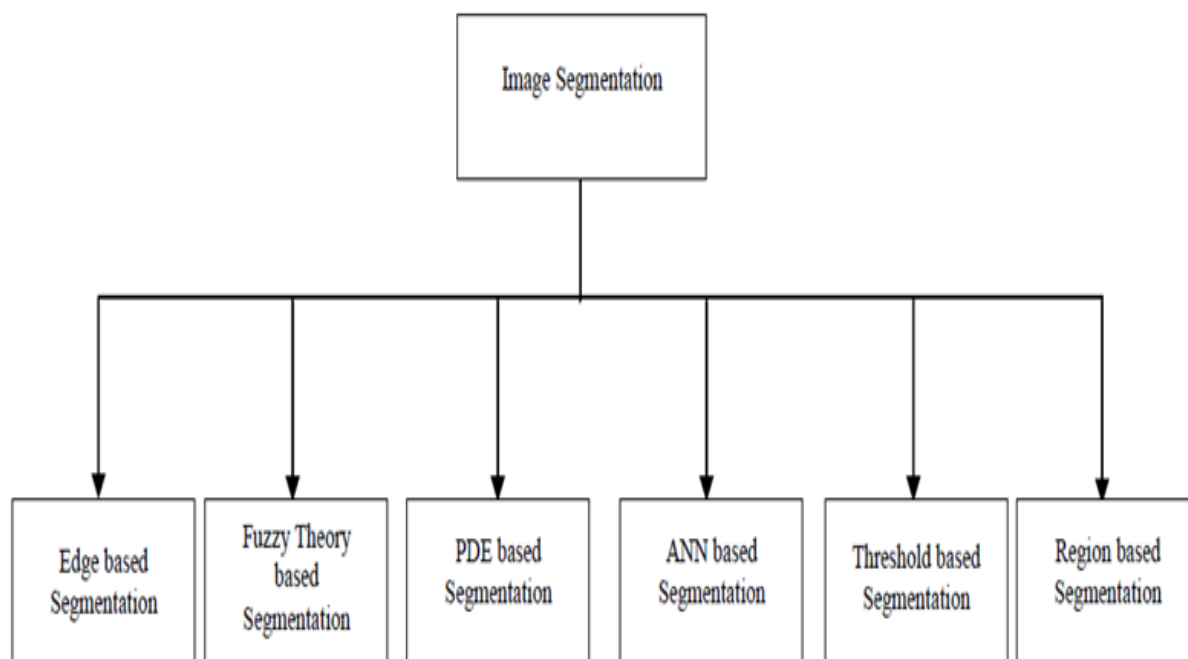
Figure 2. 11: Shows the types of segmentation methods in literature (Kahn, W. 2013).

2.2.2.1 Edge based image segmentation

In this technique, discontinuities or ending (edges) at the border of two partitions within an image (local variance in image intensity) is considered an edge. In Johnson, C.I. (2013) work of detecting and classifying a range of tramway rail surface features including other rail infrastructure, the author concluded that standard edge detection methods demonstrated inability of detecting the running band under varying circumstances. In Figure 2.12 a and b below, depicts Step and Ramp edges respectively, which describe an abrupt and gradual change in intensity from one value on one side of the discontinuity to a different value on the opposite side of the discontinuity. Figure 2.12 c and d, are line and roof edges, which describe abrupt changes only for short distance and gradual change in longer distance respectively.
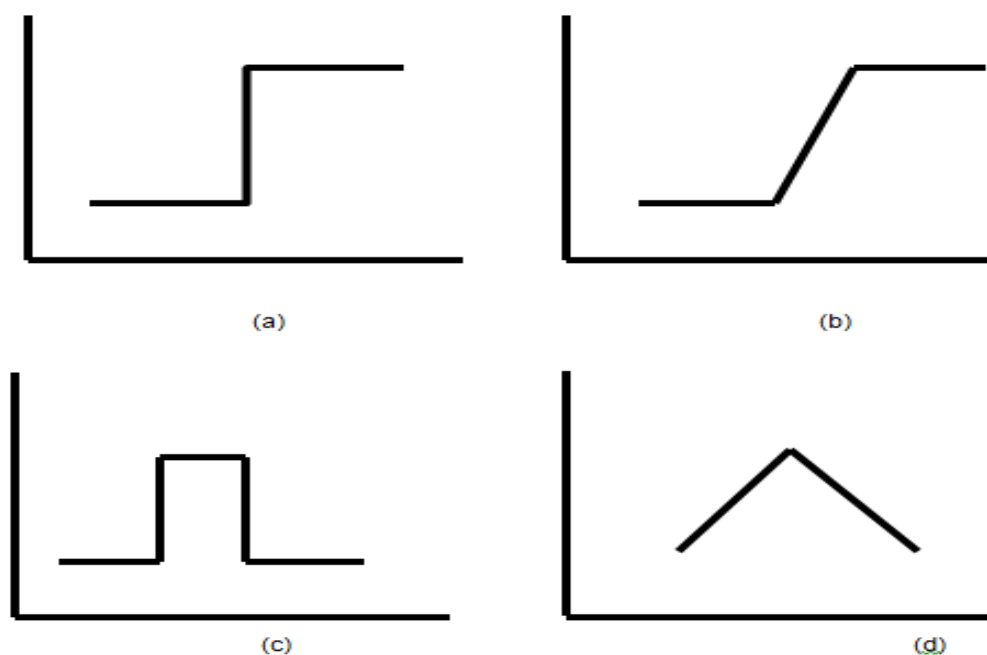
Figure 2. 12: Shows the types of edges that can be found in an image (Senthilkumaran, N., & Rajesh, R. 2009)

2.2.2.2 Theory based image segmentation

The intuition of this method is derived from diverse fields such as fuzzy logic based algorithms. Kandwal, R., Kumar, A., & Bhargava, S. (2014) explain that self-supervised and unsupervised learning capability of Fuzzification functions and clustering techniques are of great significance in the successful classification of image pixels based on some similarity criteria. Novel extensions of the fuzzy rule segmentation technique was introduced by Kahn, W. (2013) based on fuzzy IF-THEN rule structure and three membership functions namely; Region pixel distribution, measure of region closeness, and to find the spatial relationship among pixels. Furthermore, dynamic adjustment of linear weights in fuzzy connectedness for segmentation was demonstrated in a novel DyW algorithm, with successful application to images of varied modalities.

2.2.2.3 Region based segmentation

Instead of segmentation based on abrupt changes in pixel intensity, in this approach the image is partitioned into regions in accordance with some predefined similarity and termination criteria. In the case where the region is expanded with neighbouring pixel, satisfying the similarity rule will result to a special region based segmentation called region growing.

However, it is a case of Region Splitting and Merging if the partitions are comprised of arbitrary unconnected regions, which are then subsequently merged or split according to a pre-defined condition. Typically, the split-merge condition is such that every partition is initially split into four branches, and regions are only merged when no further splitting is achievable, and this process is iterated until no further merging is possible. Johnson, C.I. (2013) presented novel genetic region growing algorithm for segmentation based upon a supervised random examination of pixels which is not limited to the adjacent neighbouring pixels only as elaborated in equation 2.9 below.

$$\sum_{x=s(lb,rb),y=(H,W)}^{m,n} f(I(x,y)) \geq T \qquad (2.9)$$

Where (lb) (rb) are left and right running band edge cue of an WxH sized image respectively. The comparison function of regions (f), operates on a specific pixel to be examined I(x,y) generated from a random sample (s). A region is considered outside the running band data if f returns a value greater than or equal to threshold T.

2.2.2.4 Thresholding based segmentation

In thresholding based segmentation, pixels are allocated to background or foreground category according to intensity value characteristic. Kandwal, R., Kumar, A., & Bhargava, S. (2014) emphasises the computational ease and power embedded in this approach especially in the case of images with relatively darker background compared to foreground pixels or vice versa. Depending on the distribution of intensity values in the image, a global or local approach may be preferred.

2.2.2.4.a Global thresholding

Global thresholding is comprised of a single threshold value (T) for an entire image $I(x,y)$ as described in equation 2.10. Senthilkumaran, N. & Vaithegi, S. (2016) concludes that the popular Otsu threshold is most influenced by the class with larger variance, also best performance is realised for unimodal images. An improved Otsu method using weighted object variance for segmenting rail surface defects was presented by Yuan, X., Wu, L., & Peng, Q. (2015) and explains that results are better than most approaches.

$$q(x,y) = \begin{cases} 1, if\ I(x,y) > T \\ 0, if\ I(x,y) \le T \end{cases} \tag{2.10}$$

2.2.2.4.b Local thresholding

Carabias, D.M. (2012) elaborates that for multi-modal images, due to poor illumination, rust or other environmental factors, local thresholding (formulated in equation 2.11) is dependent on local neighbourhood statistics such as range, variance, or surface-fitting parameters which are of importance to good segmentation results.

$$b(x,y) = \begin{cases} 1, if\ W(x,y) > T_{W(x,y)} \\ 0, if\ W(x,y) \le T_{W(x,y)} \end{cases} \tag{2.11}$$

Where $W(x,y)$ is a specified neighourhood of pixels within original image $I(x,y)$. $T_{W(x,y)}$ is the estimated threshold of each neighbourhood in consideration.

2.2.3 Feature extraction (FE)

Feature Extraction is a pre-requisite in any image analysis algorithm that intends to perform some form of detection, classification, or matching. Low level feature detection (automatic feature extraction without shape information) can be achieved with satisfactory results using thresholding (as discussed in section 2.2.2). The fact remains that a threshold value large enough will remove unwanted blobs, but some discontinuities reduce efficient feature extraction, which is not acceptable for crack detection. Existing literature suggests that geometrical approach is best suited for extracting feature with known shape, while statistical and transform methods are generally considered more robust. Zhang, W., Zhang, Z., Qi, D., & Liu, Y. (2014) by means of several experimental verifications related to rail images, concluded that the most critical factor in differentiating background (railhead) and foreground (defect region) include: -

i)   The estimated standard deviation of the distance histogram expressed in equation 2.12, which describes the degree of irregularity of a spatial shape. Higher standard deviation values are realised for irrelevant objects with an irregular shape (i.e. heterogeneous distance distribution).

$$\sigma = \sqrt{\frac{1}{Nb}(p_i - \mu_p)^2} \qquad (2.12)$$

Where N_{di} is the number of pixels with the same di to centroid. di is the Euclidean distance between a pixel and centre pixel. $pi = \frac{N_{di}}{Nb}$ is the histogram distance of the image. Nb is the total number of pixels. μ_p: is the average distance of all pixels.

ii) The pixel number in most cases can be used to extract features especially in the case analysing false local dim regions (identified by morphological black top-hat transformation) such that these irrelevant objects have intensity values always less than those of actual defects. The statistical results indicate only about 10% of peak intensity values are actual defects.

iii) Average gray level is useful for feature extraction especially prior to binary segmentation, in the case of images having the actual defect region much darker than local dim regions (which correspond to high gray levels on the gray scale).

Feature extraction for general object detection along tracks is achievable with simple algorithms as demonstrated by Bhondwe, S.D., et al (2014). In the investigation of rail surface damage detection, Lui, Z., et al (2010) after analysing pre-processed (enhanced) images, proposed that the width and probability of occurrence of each blob (detected defect) can be used to extract spalling on rails. It was concluded that strips at the edge of the rail and grind marks (typical contaminants) of field acquired rail defect images are successfully removed using this approach. This method has been extended by Weng, W. & Chen, H. (2015) to include the estimated optimal segmentation threshold, thus enabling the determination of probable defect area through the comparison of each labelled defect area's relative gray value to that of the enhanced imaged (I_{enh}) according weighted function described by equation 2.13.

$$weight = \sum_{area} \frac{I_{enh}}{area} \qquad (2.13)$$

A threshold value is determined as a function of the threshold of defect area $Thr - area$ and the gray value of defect region with threshold $Thr - weight$. And for each suspect defect area true detection is achieved in the case the condition of equation 2.14 is satisfied.

$$Tar(x) = \begin{cases} True, & if \quad arear > Thr - area \quad weight > Thr - weight \\ false, & if \quad otherwise \end{cases} \tag{2.14}$$

Li, Q., & Ren, S. (2012) observed that the difference in contrast between a Minimum Subscribed Rectangle (MSR) and the Extended Minimum Subscribed Rectangle (EMSR) detailed in equation 2.15-2.17 can be considered a criterion for true defect detection, in addition to the minimum or maximum area of defects based on railway maintenance standards.

$$MSR = \{y1{:}y2 \,, x1{:}x2 \,\} \tag{2.15}$$

$$EMSR = \{y1 - \frac{h}{2}{:}y2 + \frac{h}{2} \,, x1{:}x2 \,\} \tag{2.16}$$

$$h = y2 - y1 \tag{2.17}$$

Where $y1, y2$ and $x1, x2$ are low, high border coordinates in longitudinal and in transversal directions respectively. h denotes the extended length in the longitudinal direction.

This relationship between defect area and contrast is optimised for each defect based on area threshold (Ta) contrast threshold (Tc) as described in equation 2.18 and 2.19.

$$Area\,(R) \geq Ta \; and \; C(R) \geq Tc \tag{2.18}$$

$$C(R) = \frac{\mu_{ER} - \mu_R}{\mu_{ER}} \tag{2.19}$$

Where the area threshold Ta is defined according to the specification of railway maintenance. μ_{ER} and μ_R are the mean intensity of gray levels of the EMSR and MSR respectively. The contrast threshold Tc is set as $Tc = 0.1$ in Li, Q., & Ren, S. (2012).

Mandriota, C., Nitti, M., Stella, E., & Distante, A. (2004) discussed the significance of Gabor functions as mother wavelet in texture analysis for feature extraction, highlighting the similarity of this method with human visual processing. It is further observed that the filter bank decomposition implementation of wavelet function offers improved computational speed of the method. In Xie, X. (2008) advances in texture based rail defect feature extraction was presented, and the conclusion drawn from this investigation suggests no superior method can be pin pointed depending on the application, even though hybrid texture analyses models are generally preferred in most applications.

2.3 Fatigue, initiation and propagation in rails

Post Second World War experimental work of George R. Irwin led to the establishment of purely elastic solutions for estimating energy required for fracture and this opened the gateway for precise modelling of fatigue life composed of initiation, stable propagation leading to final failure. In rail industry, rail steel response to cyclic loading (as in the case rolling contact), crack initiation and propagation and residual rail life estimations have been developed and even adopted in rail predictive maintenance through a variety of models including the Whole Life Rail Model. Optimised maintenance strategy through predictive modelling and simulation is highly prioritised by track engineers so as to curtail annual maintenance costs in European rail network running into the tune of 300 million Euros Brian, W. (2015).

In a review by Ringsberg, J.W. (2001) theoretical fatigue life prediction models have been discussed, (of which Equivalent strain approach i.e. fatigue initiation assumed for material plane with maximum range of shear strain) even though failure to account for the effect of mean stresses from the out-of-phase non-proportional loading on the fatigue life compromises accuracy of prediction. A critical plane approach that account for mean stress improve prediction of crack initiation plane and the fatigue life as demonstrated in the model of Brown, M.W., & Miller, K.J. (1973) and that of Fatemi, A., Socie, D.F. (1988). However, they do not account for material hardening satisfactorily, which is important in rolling contacts. In the case of Energy based models, the product of stress and strain is a measure of material energy

density per load cycle, and a crack is assumed to initiate and grow in the direction of the largest value of the energy-density. The limitation associated with these methods is that they do not address specific planes on which cracks will nucleate or directions along which cracks will propagate. More current situation report on RCF by Viduad, M. & Zwanenburg, W. (2009) discusses how cracks propagate slowly in the surface plain of the rail in the same direction as the plastic deformation at an angle of 15⁰ to 25⁰ angle to the running surface of the rail at depth of no more than 5mm.

2.3.1 Crack initiation models

2.3.1.1 Whole life rail model

In the U.K, Rail Safety and Standard Board (RSSB) have funded and managed RCF prediction research project leading to the development of Whole Life Rail Model (WLRM) as a tool to help track engineers in managing network infrastructure. Based on energy dissipated in the contact patch (Tγ), the wear number is estimated as a function of creep forces and corresponding creepages. Based on Figure 2.13, Iwnicki, S. (2009) elaborates on the fact that as Tγ function in equation 2.20 increases from 0 to 15 Nm/ms, no RCF damage is generated as there is insufficient energy to initiate RCF cracks. However, as Tγ increases from 15 to 65 Nm/ms, the probability of RCF initiation increases, to a maximum of 10 at a Tγ value of 65 Nm/ms. Above this value i.e. from 65 to 175 Nm/ms, the level of energy is such that the dominant form of surface damage is wear (rather than crack initiation) therefore the probability of RCF damage decreases as wear increases. Negative values of RCF damage index indicating the values of Tγ greater than 175 Nm/ms, results in wear and no RCF initiation.

$$F_\gamma = F_x\gamma_x + F_y\gamma_y + F_z\gamma_z \qquad (2.20)$$

Where F_γ is the Tγ, F_x is the longitudinal creep force, γ_x is the longitudinal creepage, F_y is the lateral creep force, and γ_y is the lateral creepage, F_z is the spin creep force, and γ_z is the spin creepage.
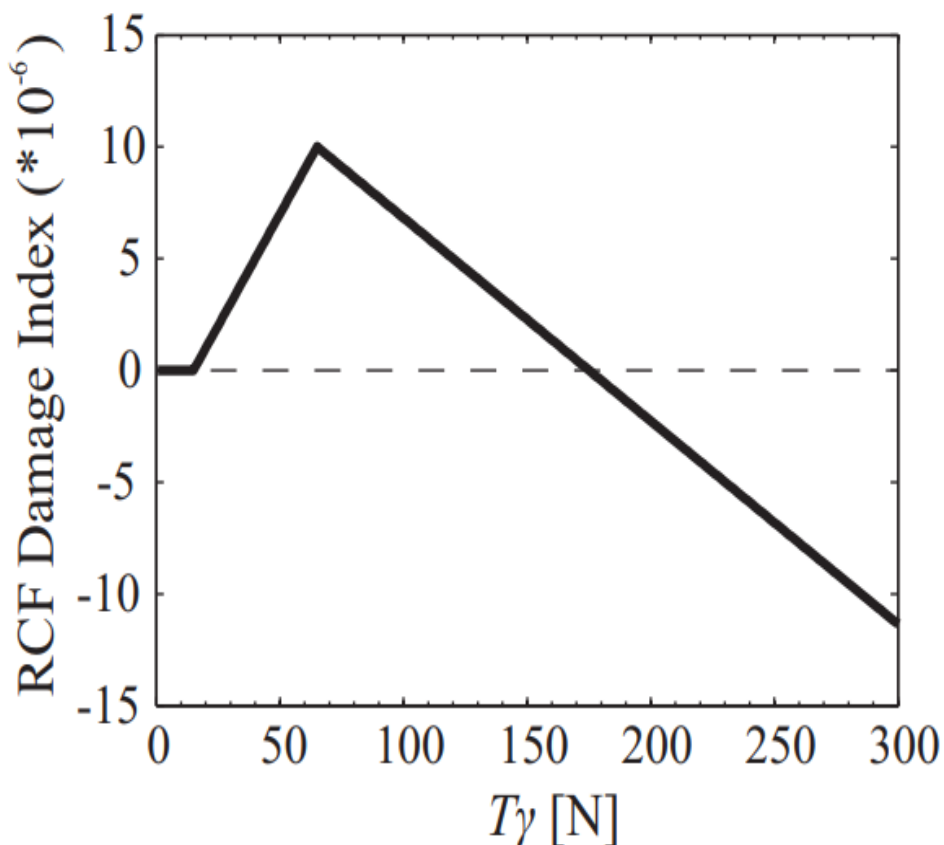
Figure 2. 13: Relationship between T_γ and RCF damage index. The units of the RCF damage index is such that for a damage index of 1, 100,000 axle passes would result in RCF initiation (Iwnicki, S. 2009).

In RSSB WLRM R&D report, RCF damage propensity was also defined as a function of damage parameter dependent on growth rate and wear rate according to equation 2.21.

$$RCF\ propensity = damage\ parameter - C \times wear\ rate \qquad (2.21)$$

where C is a constant term to achieve the correct balance between crack growth and wear.

In the absence of wear information, findings from the University of Sheffield Tyfour, W.R., Beynon, J.H., & Kapoor, A. (1996) suggests that the early growth of RCF crack is related to the amount by which the contact stress exceeds the appropriate shakedown limit. AEA Technology Rail studies, published in Beagles, A. E. (2002) showed that the contact conditions for the majority of vehicles, except on extremely tight curves, corresponds to damage parameter which suggests that RCF damage is as likely to occur on tangent track as

in most curved track, which does not agree with site observations. Failure to incorporate the position and area of the contact patch could give RCF propensities between 1.5 and 3 times greater than those calculated when such effects were included. Preliminary studies on this modification as presented in Beagles, M. (2002) yields equation 2.22 and 2.23 in terms of contact position and contact patch width.

$$Relative\ damage = \sqrt{1 - \left(\frac{Distance\ from\ centre}{Half\ width}\right)} \tag{2.22}$$

$$Half\ width = \sqrt{\frac{contact\ patch\ area}{\pi \times Ellipticity}} \tag{2.23}$$

2.3.1.2 Shakedown method

The shakedown method utilises a plot of contact stress against traction coefficient (defined as the ratio of the tangential to normal forces). The material properties set the 'shakedown limit' and exceedance of this means that RCF crack initiation is likely to occur. Field experience suggests that RCF behaviour in the case of low traction coefficient (in the range of 0 to 0.15) corresponds to contact conditions below the shakedown limit thus no RCF initiated and wear rate is significantly meaning cracks won't get worn out even if initiated. For moderate traction coefficient (ranging 0.15 to 0.3) significant contact conditions are above the shakedown limit and in spite of appreciation in wear rate, majority of RCF damage is within this range. High traction coefficient (above 0.3) contact conditions are well above the shakedown limit, but the wear rate is observed to dominate crack growth rate, resulting to minimum RCF damage in such regions. A modification of the conventional shake down fatigue damage index described in Dirks, B., Enblom, R., Ekberg, A., & Berg, M. (2015) is performed using equation 2.24, and damage is said to occur when the SI>0 occurs.

$$SI(x, y) = \sqrt{\tau_{zx}(x, y)^2 + \tau_{zy}(x, y)^2} - k \tag{2.24}$$

Where k represents the is yield limit in shear, longitudinal and lateral shear stresses for each cell $\tau_{zx}(x, y)$ and $\tau_{zy}(x, y)$ respectively.

2.3.2 Crack propagation models in rails

Brouzoulis, J., & Ekh, M. (2012) presented a crack driving force (G) model derived from the concept of homogenous hyper-elastic material forces. As depicted in Figure 2.20, the crack driving force is decomposable into parallel and perpendicular components. It is shown that only the component parallel to the crack tip of the crack driving force is a reliable quantity that can be used in RCF growth prediction.
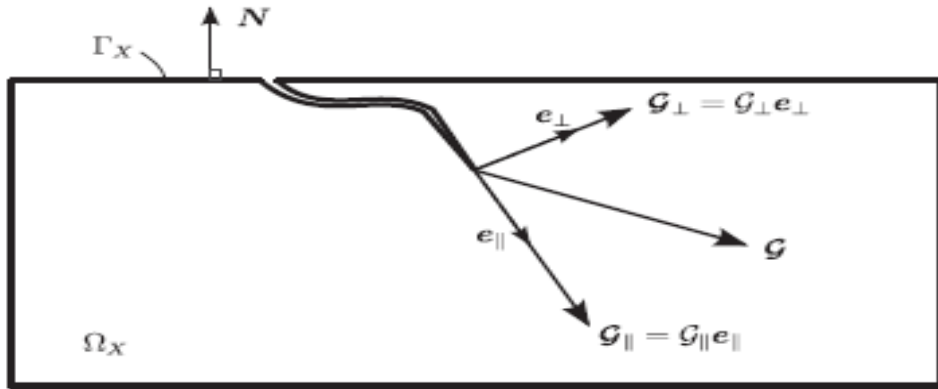
Figure 2. 14: Crack driving force concept short surface head check like cracks, often observed at the rail gauge corner. (Brouzoulis, J., & Ekh, M. 2012).

$$\mathcal{G} = \mathcal{G}_{int} + \mathcal{G}_{surf} \tag{2.33}$$

Where $\mathcal{G}_{int}$ represents contribution to the crack driving force from the internal Eshelby stresses and $\mathcal{G}_{surf}$ represents the contribution from surface tractions (e.g. contact between crack faces). The Elshelby stress tensor defined as Σ is the difference between the product of strain energy ψ and the product of deformation gradient (F) with First Piola-Kirchoff stress tensor (P).

Based on the crack driving force expressed in equation 2.33, the following rate independent propagation law in the time domain is proposed by Brouzoulis, J., & Ekh, M. (2012).

$$\frac{da}{dt} = \frac{1}{\gamma}\langle\dot{\Phi}\rangle e^* \tag{2.34}$$

Where a is crack tip position, e^* is the specified unit direction, γ material related energy dissipated in creating unit crack surface, Φ is crack driving potential,. The MacCauley brackets $\langle * \rangle = \frac{1}{2}(* + | * |)$ were used to ensure a propagation law that is proportional to the range of crack driving force.

Dirks, B., et al (2015) summarised in equation 2.30 to 2.32 above, such that the ratio of reference and actual values of fatigue damage with corresponding ratio of length or depth to be used in estimating crack growth rate for fatigue damage levels (as described in equation 2.35).

$$\frac{\rho}{\rho_r} = \frac{h}{h_r} = \frac{D}{D_r} \qquad (2.35)$$

Where the surface crack length (ρ), depth (h) corresponding to fatigue damage (D), and predefined reference is length (ρ_r), depth (h_r) and damage level (D_r) respectively.

From Figure 2.21 below, it is possible for an arbitrary surface crack length (ρ), to be derived from the depth given the angle of inclination of the crack into the railhead.
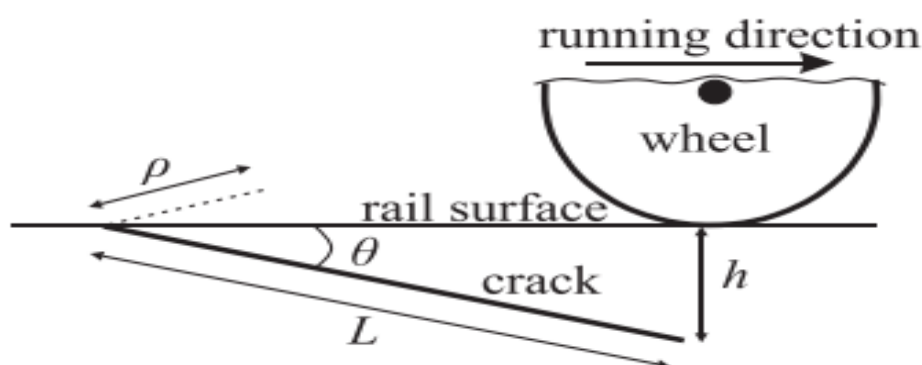
Figure 2. 15: Damage function based estimation of crack growth rate (Dirks, B., et al 2015)

In Failure Fatigue and Fracture of rail steel by Dalberg, T., & Ekberg, A. (2002), two popular parameters used to describe crack behaviour in structures are the SIF and Energy release

rate (G). In the case of LEFM assumptions, the two parameters are proportional load for constant material property. By considering the internal strain energy, including work done by loading forces and moments as criteria for formation of new crack surfaces in different materials the energy release rate/crack extension force (G) defined in equation 2.36, as the ratio of change in potential energy ($d\pi$) and change in crack area (dA).

$$G = -\frac{d\pi}{dA} \tag{2.36}$$

Accordingly, a crack is said to grow when the G equals a critical value related to fracture toughness of the material. However, in this literature review, the focus is on SIF description of crack propagation. Assuming that SIF is an indication of the amplitude of the crack-tip singularity, and can be determined from stresses near the crack tip along the normal to the crack front as shown in Figure 2.34 then equations 2.37-2.39 are applicable.
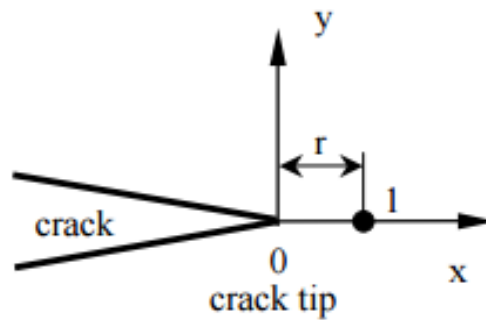
Figure 2. 16: SIF estimation from stress fields at vicinity of crack tip. (Dahlberg, T., & Ekberg, A. 2002)

$$K_I = \sigma_{yy}\sqrt{\pi a f} \tag{2.37}$$

$$K_{II} = \sigma_{yx}\sqrt{\pi a g} \tag{2.38}$$

$$K_{III} = \sigma_{yz}\sqrt{\pi a h} \tag{2.39}$$

Where K_I, K_{II}, K_{III} are the SIF's due to tension, shear, and torsion fracture mechanism respectively.

The SIF can alternatively be determined using displacement correlation near the crack tip (points 1 and 2 on in Figure 2.35) using the mode I, and II equations elaborated in Lui, L. (2008) shown in equation 2.45.
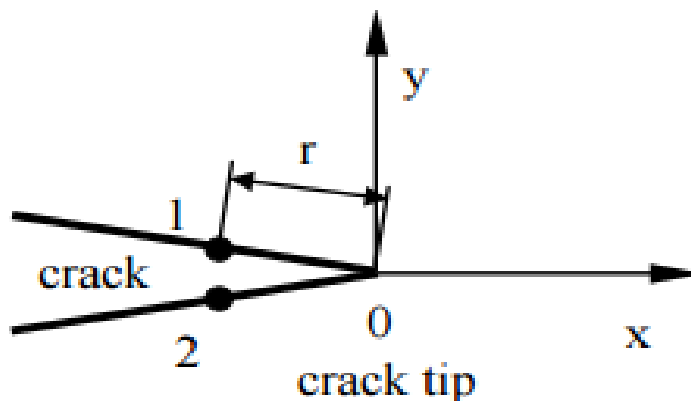
Figure 2. 17: SIF estimation from crack tip displacement fields (Lui, L. (2008)

$$K_I = (v_2 - v_1)\frac{E\sqrt{\frac{2\pi}{r}}}{8(1-V^2)} \tag{2.40}$$

$$K_{II} = (u_2 - u_1)\frac{E\sqrt{\frac{2\pi}{r}}}{8(1-V^2)} \tag{2.41}$$

Where K_I, K_{II} are the stress intensity factors, u_i, v_i $(i = 1, 2)$ are displacements in local x, y coordinate. r is a distance of the node on the crack surface from the crack tip, E is the Young's modulus and V is the Poisson's coefficient.

Several studies such as Beden, S.M., et al (2009), Kotsikos, G., & Grasso, M. (2011), and Bhalekar, B, D., & Patil, R,B. (2016) suggest that the propagation of a crack is driven by the stress field that develops at the vicinity of the crack tip. In rolling contact, mixed loading conditions are believed to be responsible for crack growth at the crack tip. Moreover, there exist various criteria proposed in literature for the calculation of effective mixed-mode SIF of which the following are highlighted in this thesis.

Tanaka, K. (1974) considered Mode I and II (see equation 2.42) deformation to be mutually exclusive with fatigue crack propagation only when sum of the absolute values of the displacements in a plastic strip reaches a critical value. The effective SIF with best fit with experimental data is defined accordingly.

$$\Delta K_{eff} = (\Delta K_I{}^4 + 8\Delta K_{II}{}^4)^{\frac{1}{4}} \tag{2.42}$$

Where ΔK_{eff} is effective SIF, ΔK_I is the rang mode-I SIF, ΔK_{II} is the rang mode-II SIF.

Yan, X., Zhang, Z., & Du, S. (1992) regardless of experimental validation suggested using maximum tangential stress criterion proposed by Erdogan, F., & Sih, G.C. (1963) taking into account crack growth direction as described in equation 2.43.

$$\Delta K_{eff} = \frac{1}{2}\cos(\theta) \times [\,\Delta K_I(1 + \cos(\theta)) - 3\Delta K_{II} \times sin(\theta)\,] \tag{2.43}$$

Where θ is the crack growth direction, ΔK_{eff} is effective SIF, ΔK_I is the range of mode-I SIF, ΔK_{II} is the range of mode-II SIF.

The Energy release rate models discussed in Forth, S.C. (2003) calculates SIF taking into account Mode III loading as elaborated in equation 2.44 below. The authors further simplified the model by showing passion ratio (v) is negligible.

$$\Delta K_{eff} = (\Delta K_I{}^2 + \Delta K_{II}{}^2 + \frac{1}{(1-v)}\Delta K_{III}{}^2)^{\frac{1}{2}} \tag{2.44}$$

Where v is Poisson ratio, ΔK_{eff} is effective SIF, ΔK_I is the range of mode-I SIF, ΔK_{II} is the range of mode-II SIF, ΔK_{III} is the range of mode-III SIF

Modifications made on existing fatigue growth laws utilising effective SIF to account for mixed mode loading (dominant in surface propagation of rail defects) has gained application in rail predictive maintenance modelling. Equation 2.45 shows a modified Paris law formulation which is convenient for cyclic loading case studies.

$$\frac{da}{dN} = C(\Delta K_{eff})^m \tag{2.45}$$

Where $\frac{da}{dN}$ is the crack growth rate, C is a constant, ΔK_{eff} is effective SIF, the parameters C is a constant and m is the slope on the log/log plot in region II crack growth stage.

Kim, J.K., & Kim, C.S. (2002) based on experimental investigations further accounted for the effect of stress ratio R, material constant γ, loading, and geometry as described in equation 2.46 below.

$$\frac{da}{dN} = C\left(\frac{\Delta K_{eff}}{(1-R)^{1-\gamma}}\right)^m \tag{2.46}$$

$$R = \frac{K_{II\,max}}{K_{I\,min}} \tag{2.47}$$

Where $\frac{da}{dN}$ is the crack growth rate, C is a constant, ΔK_{eff} is effective SIF, m is the slope on the log/log plot in region II crack growth stage, γ is the material parameter, R is the ratio of max mode-II SIF and mode-I SIF.

Forman, R.G., Kearney, V.E., & Engle, R.M. (1967) investigated the influence of crack growth characteristics at both low and high levels of ΔK. He concluded the relation in equation 2.48 suggesting that at high ΔK values, as max K approaches the critical level K_c, an increase in crack growth rate is observed.

$$\frac{da}{dN} = \frac{C(\Delta K)^m}{(1-R)K_c - \Delta K} \tag{2.48}$$

Where $\frac{da}{dN}$ is the crack growth rate, C is a constant, ΔK is effective SIF, m is the slope on the log/log plot in region II crack growth stage, γ is the material parameter, , R is the ratio of max mode-II SIF and mode-I SIF.

A summary of publications and their suggested R, C, and m parameters for different rail materials are shown in the Table 2.3 below.

Author	R-ratio	Paris law constant C	Paris law constant m
Kim and Kim, 2002	0.1	4.47E-9	3.13
Elshabasy & Lewandowski, 2004	0.1	-	3.50
Elshabasy & Lewandowski, 2004	0.7	-	6.20
Skyttebol, Josefson, Ringsberg, 2005	0.7-0.9	2.74EE-9	3.33
Zaijn, Sajuri, Yusof, & Hanafi, 2010	0.1	2.63E-12	3.29
Heshmat, 2011	0.1	1.13E-11	2.17
Kotsikos & Grasso, 2012	0.1	3.3E-12	2.63
George,2012	-	3.3E-13	2.63

Table 2. 3: Published values for Paris Law constant R, C, and m (Zafosnik, B., et al 2000).

2.3.3 Crack branching criteria

The Maximum Tangential Stress criterion (MTS) as detailed in the works of Erdogan, F., & Sih, G.C. (1963) used stress equations based on the assumption crack propagation occurs in the direction of maximum tangential stresses calculated on a circle of sufficiently small radius around the crack tip. The angle of crack propagation θ is in this case is defined as in equation 2.49.

$$\tan\left(\frac{\theta}{2}\right) = \frac{-2K_{II}}{K_I + \sqrt{(K_I)^2 + 8((K_{II})^2)}} \qquad (2.49)$$

In the case of Maximum Energy Release Rate criterion (MERR) initially proposed by Griffith, the intuition is that crack will grow in the direction of maximum energy release per unit extension (i.e. in the direction of the maximum stress intensity factor K). Helen, T.K., Blackburn, W.S. (1975) showed that the energy release rate may be expressed as a function of the J1 and J2 integrals. The corresponding angle of crack propagation is here obtained as in equation 2.50.

$$\theta = \arctan\left(\frac{J_2}{J_1}\right) = \left(\frac{2K_I K_{II}}{K_I{}^2 + K_{II}{}^2}\right) \tag{2.50}$$

The minimum Strain Energy Density criterion (SED) Zafosnik, B., et al (2000) is another postulation that considers critical value of the local strain energy as a criterion to determine the direction of propagation as the simultaneous solution of the following equations 2.51 and 2.52.

$$[2\cos(\theta) - (k-1)]\sin(\theta)K_I{}^2 + 2[2\cos(\theta) - (k-1)\cos(\theta)]K_I K_{II} + [(k-1-$$
$$6\cos(\theta))\sin(\theta)]K_{II}{}^2 \tag{2.51}$$

$$[2\cos(2\theta) - (k-1)\cos(2\theta)]K_I{}^2 + 2[(k-1)\sin(\theta) - 4\sin(2\theta)]K_I K_{II} + [(k-1)\cos(\theta) -$$
$$6\cos(2\theta)]K_{II}{}^2 > 0 \tag{2.52}$$

2.4 Conclusion

A detailed survey of existing NDTs highlighting differences between the methods in section 2.1, results to the conclusion that it is necessary to consider the type of data obtained from each NDT in order to assess the suitability of adopting a particular method for further incorporation to prediction models. With this understanding, not even the most widely adopted NDT in the U.K. network (Ultrasonic) or even the increasingly patronised Eddy currents technique is considered ideal for the aim of this PhD research. This survey has favoured the choice of automated visual inspection of rails (due to inspection speed, flexibility, adaptability to multiple tests, and remoteness) to demonstrate the readiness level of the proposed incorporated detection and prediction inspection technology. It was also deemed necessary to conduct a critical appraisal of the existing image processing algorithms particularly those

adopted in rail RCF detection, taking cognisance of the effect of rail head non-uniform reflection property, rust, grind marks and scratches; the enhancement, segmentation and feature extraction stages of the algorithms applied within the past decade suggest the advantage of local approach especially in the first two stages, while morphology and/or texture based analysis is popular in the latter stage. The mechanisms responsible for initiating and propagating (including branching criterion) RCF damage have been well understood in literature. In this review the conclusion affirms that in the U.K, majority of RCF is initiated as a result of Mode-0 and Mode-2 initiation mechanism, while the dominant loading condition for surface propagation of cracks are the tension and shear loading cases. Further clarity based on field observation of the behaviour of crack propagation by RSSB, establishes the influence of traction, wear and entrapped fluid etc. Incorporating the defect detection algorithm with detailed structural mechanics models will be explored as intelligent predictive maintenance tool in the next chapter.

Chapter 3: Application of defect detection algorithm

3.1 Introduction

In the UK, rail RCF cracking information (as part of the outcome of the post-Hatfield incident investigations), via the Arup/TTCI RCF database, was prepared for Railtrack maintenance support. This data with other European collaborative research sources is able to provide over 22000 track miles with data recorded for every 55 yards of track. Such information related to damage levels accompanied by infrastructure supporting information (such as rail age, profile, sleeper type etc.) are obtained through large numbers of different contractors, with inevitable differences in the thoroughness and method of surveying. Observations at the time of the investigation indicated that existing database considerably over-reported false defects (scratches, grind marks and rust). Increasing correlation between maintenance planning and such data is at the expense of accurately accounting for the location, geometry and severity of damage. AEA Technology Rail have made attempts to correct anomalies related to approximations of network annual tonnage, and incomplete data. This section of the PhD thesis, discusses the development of image processing algorithms presented in Appendix C, capable of detecting rail fatigue damage, extracting geometrical features, and supporting rail maintenance decision making. Three data sets, described in section 3.2 have been considered for this research work as detailed in Appendix C1 and Appendix C2. Modifications of recent rail image processing methodologies within the past decade are presented in section 3.3, the MATLAB algorithms related to enhancement, segmentation and feature extraction are detailed in Appendix C3, Appendix C4, and Appendix C5. In section 3.4, a comparative study of each stage of the developed algorithm(s) is presented with the simulated results depicted in Appendix C6. Based on the repeatability of simulated results after calibrating the image as shown in section 3.4.4, the suitability of a particular image processing technique is also assessed in terms of consistency in generated fatigue damage statistical data over multiple runs of the algorithm. The effectiveness of combining two or more methods at each respective stage of the algorithm is also considered on data set 3 as shown in the results presented in

Appendix C7, which supported the optimisation and automation of the final proposed algorithm in section 3.5. It is intended that the conclusions drawn in terms of performance and computational requirement discussed in section 4.5.1 and 3.6 will provide a foundation for remote and possibly real time predictive image based condition monitoring maintenance of rail infrastructure as the main contribution of this chapter of the PhD thesis.

3.2 Data set

Three image data sets (Data set 1, Data set 2, and Data set 3) are adopted from the different sources namely; Manchester Metropolitan University (MMU) U.K, National Research Council Canada (NRC), and International Collaborative Research Institute (ICRI) respectively, for the purpose of this research. These data sets are comprised of laboratory (controlled environment) acquired, and field acquired rail defect images as described below.

3.2.1 Data set 1

This is comprised of four laboratory coloured images as depicted in Figure 3.1 with samples Lab-1 and Lab-2 acquired with the potential of validating the generated defect data. Both samples are reasonably uniformly illuminated and contain random types of defects with spalling and fatigue damage considered in this investigation. However, it is observed that sample Lab-2 is heavily contaminated by rust and aging effect. While sample Lab-3 and Lab-4 are mostly dominated by cluster of fatigue damage, and the influence of poor reflective property of the rail is also present. It is worth mentioning that the details of the image acquisition system for this data set samples are not readily available.

Figure 3. 1: Shows laboratory acquired images of data set 1. (MMU).

3.2.2 Data set 2

The samples in Figure 3.2 were acquired via a 2 Mega pixel Rail scope image acquisition system (with up to 35mph inspection speed). NRC pre-classified and sampled five field acquired grey scale images each, for 3 distinct severities of RCF damage (identified as heavy, moderate, and light damaged sites). This enabled the establishment of key model parameters for varied severity of damage. All samples contain random types of defects, however it was noticeable that spalling is more common to heavily damaged samples while cluster of fatigue damage were persistent in moderately and lightly damaged samples. It is worth mentioning that no prior information with regards to geometrical data of RCF defects were provided by NRC.

3.2.3 Data set 3

Courtesy of the international collaborative research project on RCF damage, over 1000 samples of field acquired low and high rail grey scale images have been used in this investigation. All samples shown in Figure 3.3 contain random types of defects; however, no details of the image acquisition system or severity of damage information was provided along with the samples. It is also worth mentioning that the samples were obtained in snow weather condition which was observed to influence certain parameters of the model. A sub set of 30 images is considered in this section.

3.3 Methodology

The presence or absence of a defect in an image can be characterised by change in intensity value between neighbouring pixels. It is therefore most common to suspect local dim, or bright region within an image as suspect defect region. Such assumption can be modelled as function of mean intensity value (each pixel's contribution to the overall brightness of the image) and variance of adjacent pixels. It is therefore an intuitive solution in detecting RCF defects to prefer local methods of processing images to ensure accurate data for supporting predictive rail maintenance. After a concise review of recent image processing techniques presented in the previous chapter, the following methods are considered. Enhancement by

smoothing using a median filter, and sharpening by averaging filter are adopted for preparing the image for normalisation. The smoothened image is processed by a normalisation function, which focuses on local mean and variance. In the frequency domain, the performance of Local Fast Fourier Transform (FFT) is also considered in this section of the thesis.

Segmentation by thresholding functions of intensity values, defect occurrence probability, and Automatic Iterative algorithm are investigated. Also, the Feature Extraction (FE) stage is comprised of Gabor texture analysis, Standard deviation shape histogram distance and establishing a criterion for geometrically unacceptable defects. MATLAB

3.3.1 Enhancement of defect region

Image analysts emphasise the importance smoothing images for removal of unwanted peaks in intensity values prior to enhancement. However, the suitability of a particular enhancement technique is subject to the overall objectives of the application in consideration. To this end, in this thesis all images are smoothened by a median filter. Results from coins.png test image are presented in Figure 3.5. The Figure confirms Zhiyuan, Q., et al (2002) report that such statistical filters are efficient for noise reduction with improved preservation of useful detail in the original image.

Original image

Pre-smoothing spatial distribution of pixel intensity
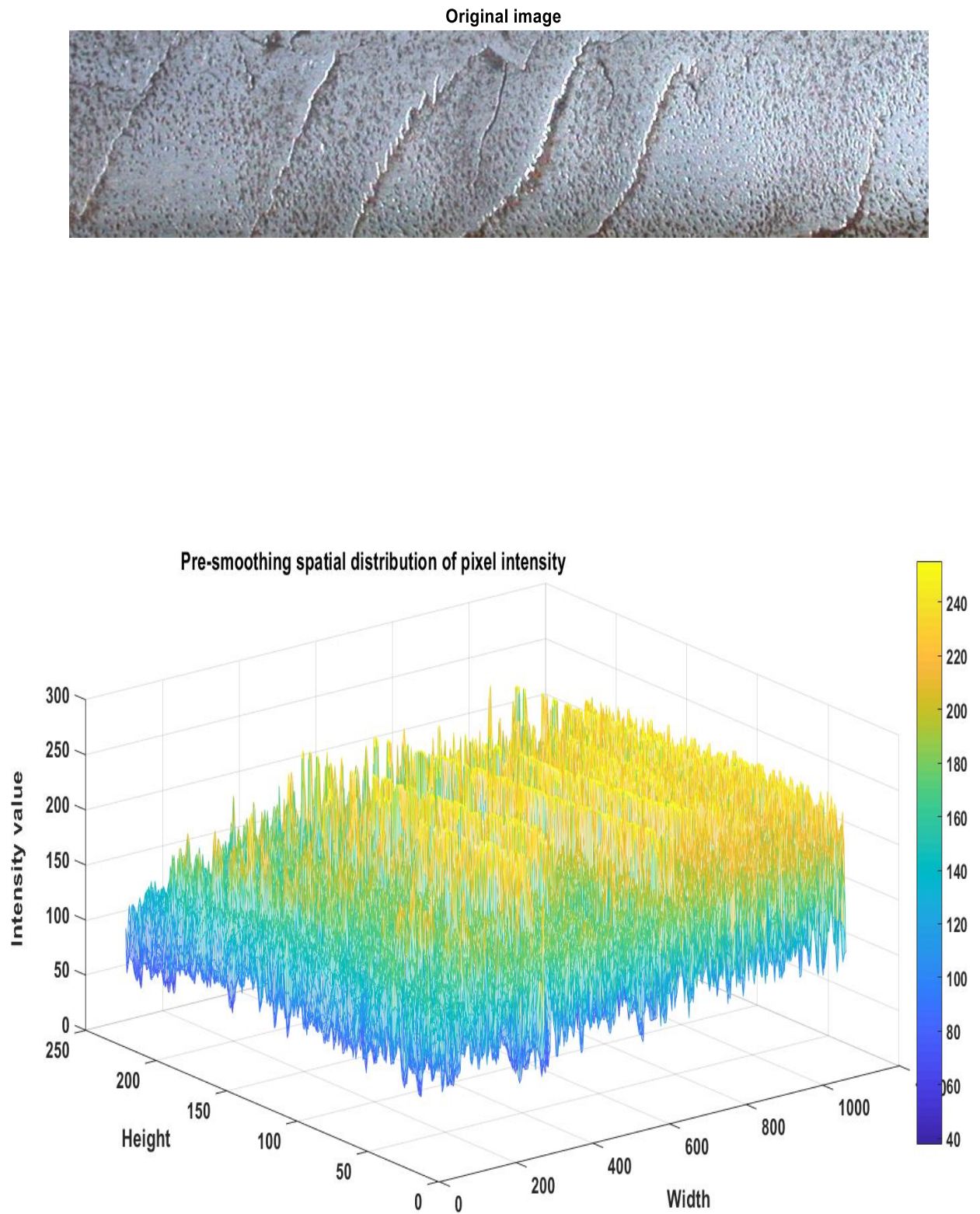
Figure 3. 2: Depicts the Lab-2 original image and pre-smoothing pixel intensity value distribution

Median filtered gray image

Post-smoothing spatial distribution of pixel intensity
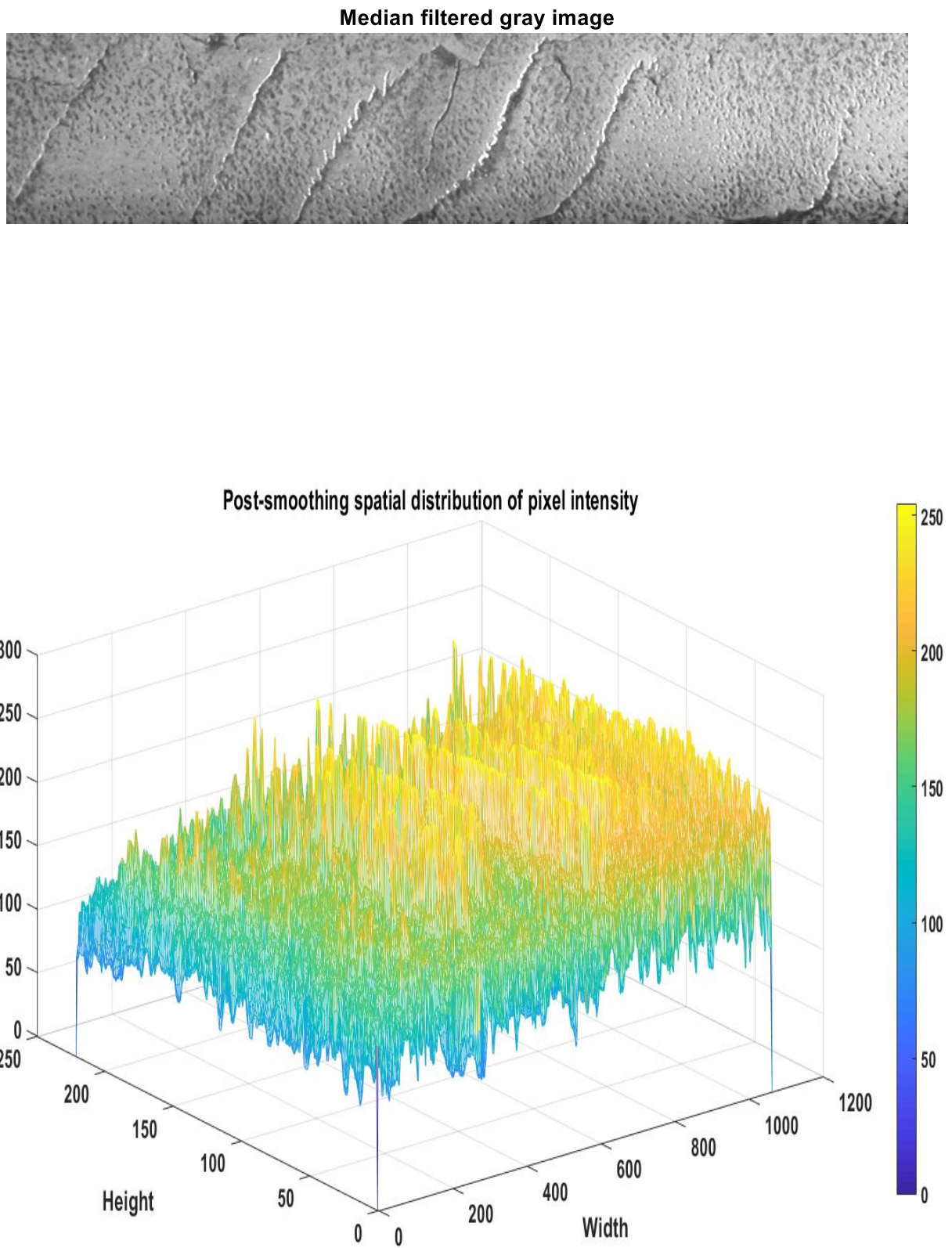
Figure 3. 3: Depicts the Lab-2 median filtered image and post smoothing pixel intensity value distribution

The median filter is preferred because each seed pixel is replaced by the median value of intensities around the seed as opposed to the averaging, or other statistical functions such as minimum or maximum value. After successful smoothing of the input image, four local enhancement techniques are comparatively studied and analysed using (data sets 1 and 2) and their respective performance is quantified by a measure of peak signal to noise ratio (PSNR). The mathematical models governing the different enhancement methods adopted in this thesis are described below.

3.3.1.1 Linear moving average filter (LMF)

In this method the averaging of a specified number of pixels (N) around a seed value (x, y) is performed along each strip (i.e. along rolling direction) of the image. And depending on the information regarding noise type and distribution of intensity values within the original image, a compensation mean offset (Δ) is introduced to exclude uneven grey values contamination. Weng, W. & Chen, H. (2015) presented the models in equations 3.1-3.4 and considered the influence of white noise thus assumed Δ=0.5, also suggesting the lower the value of Δ the more efficient noise reduction attained.

$$LMAF_{xy} = \sum_{y-N/2}^{y+N/2} \frac{I_{xy}}{N} \tag{3.1}$$

$$I_\Delta = LMAF_{xy} - \Delta \tag{3.2}$$

Where Δ value should be slightly greater than the standard deviation of N pixels around seed pixel. I_{xy} is original image. N is neighbourhood of pixels along the rolling direction.

Further consideration of environmental factors that challenge image processing of rail defects such as rust strips (usually distributed on both sides of the rail edges), are curtailed by means of image subtraction between I_Δ (compensated image) and I_{xy} (original image) as a remedy for such unwanted intensity levels.

$$I_{con} = I_\Delta - I_{xy} \tag{3.3}$$

I_{con} could be positive, zero, or negative values which corresponds to defect free regions, background regions, and possible defect region respectively. And according to equation 3.4 the final enhanced image (I_{enh}) is obtained as map of the LMAF grey value to a range of 0-L as opposed to $0 - I_{con}$.

$$I_{enh} = L - (\frac{LMAF_{xy} \times L}{I_{con}}) \tag{3.4}$$

Original image
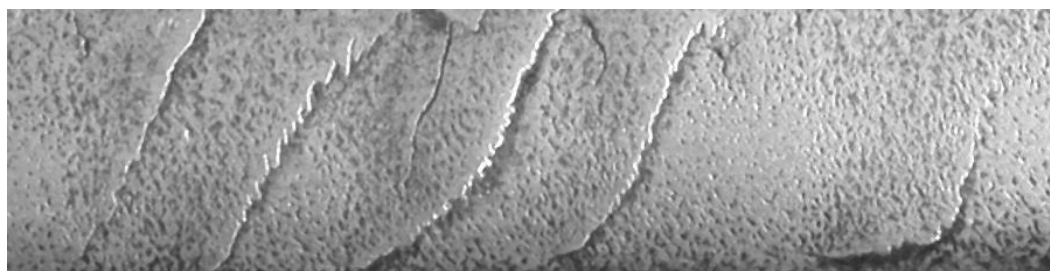
LMAF enhanced image
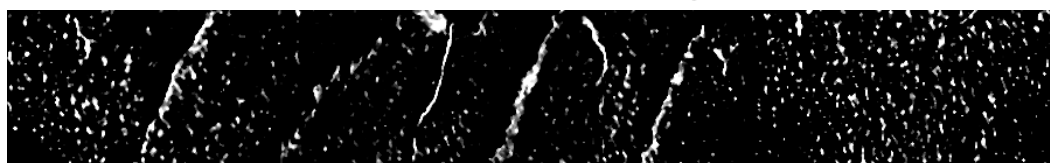
Figure 3. 4: Shows the result of Linear Moving Average Filtering (LMAF) of the original Lab-2 image.

3.3.1.2 Visibility Measure (VM) enhancement

Based on Michelson's definition of contrast of an image, Vijaykumar, V.R., & Sangamithirai, S. (2015) proposed VM contrast enhancement technique defined in a similar manner to the working principle of the human eye (in terms of visualising objects). Furthermore, the method

utilises the maximum and minimum luminance of an input image ($I_{(x,y)}$) within a local window. In addition to the mean intensity (μ) the local visibility of the neighbourhood is derived as presented in equation 3.5. To ensure uniform background in the output image from this method, the pixel location with grey intensity value higher than the mean intensity of its corresponding local window is truncated according to equation 3.6.

$$V(x,y) = \frac{I(x,y) - \mu}{I(x,y) + \mu} \tag{3.5}$$

$$V_{enh} = \begin{cases} \frac{I(x,y) - \mu}{I(x,y) + \mu} & if\ I(x,y) < \mu \\ 0 & otherwise \end{cases} \tag{3.6}$$

Where μ is the mean intensity of the local window in consideration

Original image
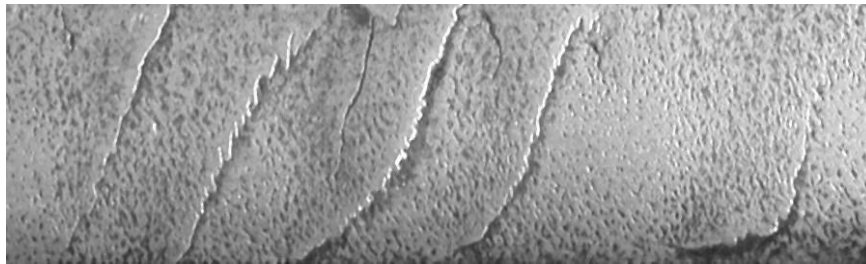
Visibility measure enhanced image
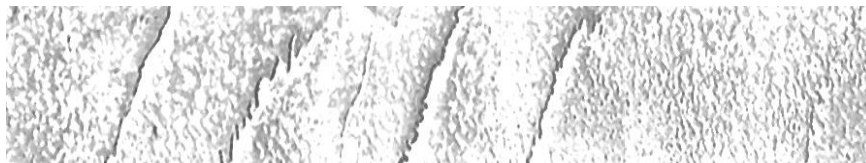
Figure 3. 5: Shows the simulated result for Visibility measure enhancement for Lab-2 image.

3.3.1.3 Local normalisation (LN) enhancement

Normalisation is used to force the input image to more familiar or easier to process form. This method is well received for poor contrast images and illumination sensitive applications. The concept of linear local normalisation method (see equation 3.7) focuses on minimum and maximum pixel intensity values.

$$I_{LN} = (I - I_{min}) \frac{newI_{max} - newI_{min}}{I_{max} - I_{min}} + newI_{min} \tag{3.7}$$

Where I_{LN} is the locally normalized image. I_{min} is the local minimum intensity value. $newI_{min}$ is the new minimum local intensity value. $newI_{max}$ is the new maximum local intensity value.

Xie, X., & Lam, K. (2006) took into consideration mean and standard deviation intensities as opposed to the range utilized in equation 3.7. The objective of this modification as per equation 3.8-3.9 is to locally establish the contribution of each pixel to a function of required mean (RM) and required variance (RV). This offers the solution of illumination and reflectance inequality across the rail head especially in field acquired samples.

$$LN(x, y) = R_m + \left[\sqrt{R_v} \times \sigma(x, y) \right] \tag{3.8}$$

$$\sigma(x, y) = \frac{im(x,y) - \overline{im(x,y)}}{\sqrt{\frac{\sum |im(x,y) - \overline{im(x,y)}|}{n(x,y)}}} \tag{3.9}$$

Where $im(x, y)$ is the windowed sample of the original image. $\overline{im(x, y)}$ is the mean intensity of the window, $\sigma(x, y)$ is the standard deviation, and $n(x, y)$ is the total number of data points.

Original image

Locally normalised image
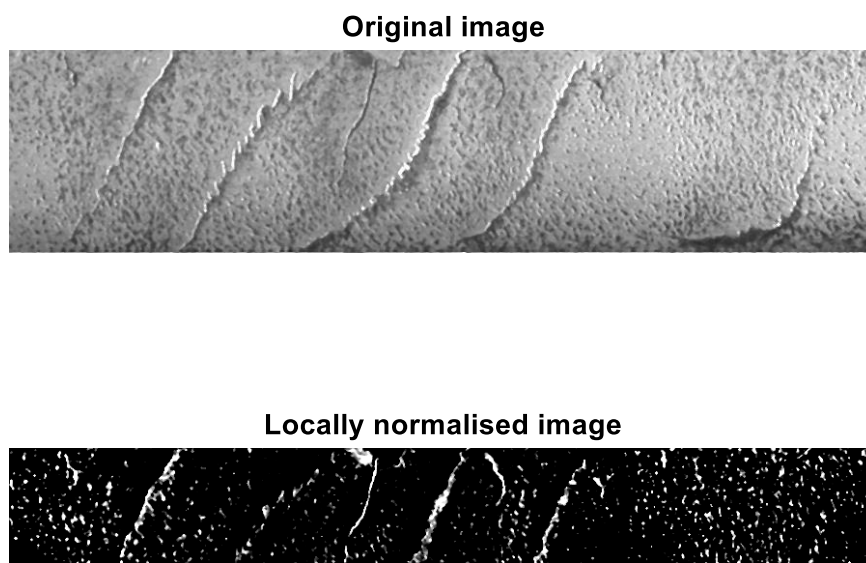
Figure 3. 6: Shows the result for local normalisation performed on Lab-2 image.

3.3.1.4   Fast Fourier Transform (FFT) enhancement

High frequency components of acquired images are most contaminated by noise even though containing most detail, while the low frequency component contains the most information within an image. Based on this understanding the detection of damage from rail images requires careful extraction of both high and low frequencies. Fingerprint detection and recognition algorithms have demonstrated the use of Fast Fourier Transform (FFT) for enhancement by exclusion of unwanted frequencies. In this method the original image is divided into overlapping local windows and the dominant frequency of each window is enhanced according to equation 3.10 below.

$$g(x,y) = F^{-1}[F(u,v) \times |F(u,v)|^{k}] \tag{3.10}$$

Where $g(x,y)$ is the FFT enhanced image.

For gray images the range of k values typically $0 < k < 1$ ensures adequate preservation of information by amplifying low frequencies while attenuating noise levels in high frequency component of the original image, where $F(u,v)$ is the Fourier transform of an image

$f(x, y)$ according to the Discrete Fourier Transform (DFT) relation in equation 3.11. Ishmael, S.M., (2011) further explains that better enhancement results are obtained if the magnitude of the FFT is squared or cubed before it is scaled by its magnitude raised to the power of k.

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)\exp(-2\pi(\frac{x}{M}u + \frac{y}{N}v)) \qquad (3.11)$$

For pixel locations u and x =1,2,3…, M-1 while v and y =1,2,3…,N-1.

And the inverse DFT function is also evaluated accordingly:

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v)\exp(-2\pi(\frac{u}{M}x + \frac{v}{N}y)) \qquad (3.12)$$

For pixel locations u and x =1,2,3…, M-1 while v and y =1,2,3…,N-1.

In Figure 3.7 below, the performance of FFT on Lab-2 image is presented, with a block size of 3×3, for the median filter and a maximum of 22×22 block size for the FFT function detailed in Appendix C3.

Original image
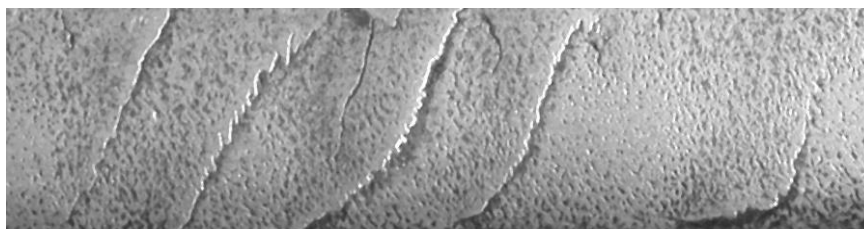
Fast Fourier Transform enhanced image
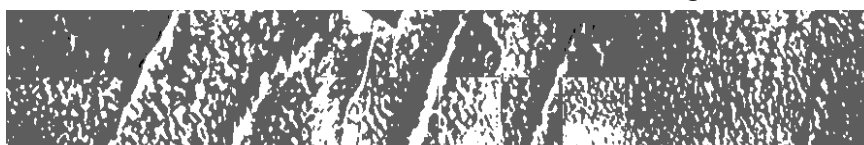
Figure 3. 7: Shows the result of Fast Fourier transform on Lab-2 image.

3.3.2 Segmentation of defect region

After successful enhancement of defect regions within an image, there may still exist some unwanted or peaks of intensity value that may result in false defect detection, which can be avoided using segmentation by thresholding functions. In this thesis local mean and standard deviation and occurrence probabilities are the thresholding based segmentation methods adopted for the purpose of sharpening the enhanced image and also for the binarisation of the image to enable easier and more accurate feature extraction.

3.3.2.1 Automatic iterative threshold (AIT)

The automatic iterative thresholding approach is based on computation of estimated mean intensity (T_{est}) of infinite partitions of the image until a variation of less than 0.1% is achieved as described by the algorithm below. The value of T_{est} is best initialized to the global mean intensity value of the image.

1. Initialize estimated threshold T_{est} approximately equal to average intensity of original image.

2. Partition original image to R1 and R2 based on T_{est}.

3. Calculate the new mean grey values μ_1 and μ_2 of R1,R2.

4. Select a new estimated threshold $T_{est} = \frac{1}{2}(\mu_1 + \mu_2)$.

5. Repeat steps 2-4 until the change in estimated mean value between successive iterations does not change by more than 0.1%.

Original image

Automatic iterative threshold
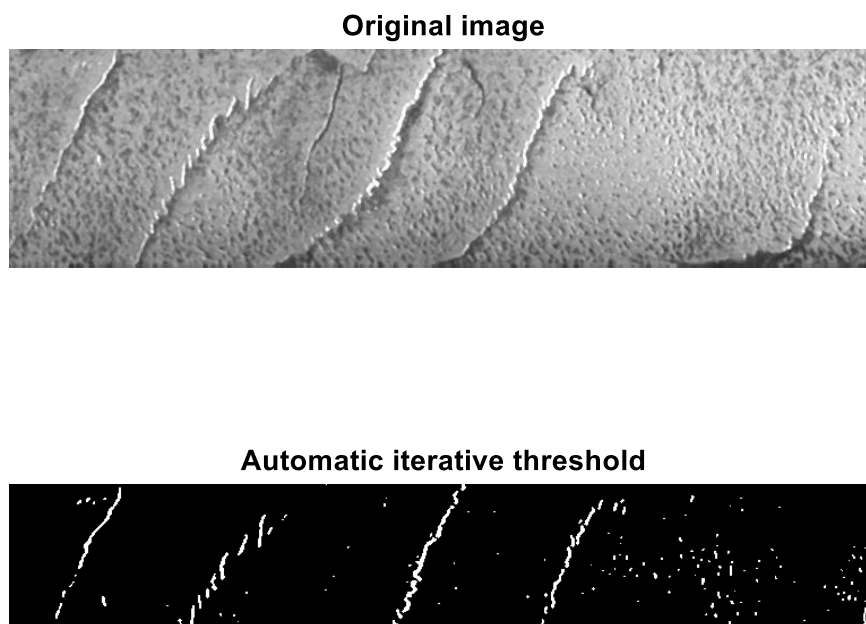
Figure 3. 8: Shows the result for Automatic iterative thresholding performed on Lab-2 image.

3.3.2.2 Combined mean and variance based segmentation

Document binarisation introduced by Niblack, W. (1986) was based on local mean and variance for the segmentation of regions of interest by the relation defined in equation 3.13.

$$T_N(x,y) = m(x,y) + K \times S(x,y)$$ (3.13)

Where $T_N(x,y)$ is the NIblack threshold for each local wxw window. While $m(x,y)$ and $S(x,y)$ are the corresponding local average intensity value and standard deviation respectively. k is bias constant, set as 0.5 in biomedical application of Senthilkumaran, N. & Vaithegi, S. (2016).

Efficient segmentation of poorly illuminated images with foreground pixel intensity value tending to zero while background tends towards maximum intensity value of 255, for 8-bit image representation required inclusion of the dynamic range of pixel intensity values as demonstrated in the works of Sauvola, J. & Pietikainen, M. (2000).

$$T_S(x,y) = m(x,y) \times (1 - k \times (1 - \frac{S}{R}))$$ (3.14)

Where $T_S(x, y)$ Savuola threshold for each local window wxw. $m(x, y)$, $S(x, y)$, and k are as defined in equation 3.13, while R is the maximum value of the standard deviation of the gray image.

For images with narrower ranges of foreground and background intensity values, Mostafa, A., et al (2016) explains the use of contrast normalised input image according to the mathematical model of Wolf in equation 3.15.

$$T_W(x, y) = (1 - k) \times m(x, y) + k \times \frac{S(x,y)}{R}(m - M) \tag{3.15}$$

where $T_W(x, y)$ Wolf's threshold for each local window wxw. $m(x, y)$, $S(x, y)$, k, and R are as defined in equation 3.14. And M is the minimum grey value of the image.

Feng, M-L., & Tan, Y-P. (2004) introduces two concentric windows; a primary window for the local estimation mean (m), minimum gray intensity value (M), and a secondary window for the estimation of standard deviation and dynamic range standard deviation (S_1) and (S_2). However, despite addressing the R problem in latter case, the introduction of two windows and three parameters (see equation 3.16) makes the approach tedious. Mahmoudi, A. & Regragui, F. (2009) in his research on improvement of thresholding speed, suggests that the intensity at a position is equal to the sum of the intensities of all the pixels located on the top-left of this position (integral image equivalent).

$$T_F(x, y) = \left(1 - k_1\left(\frac{S_1(x,y)}{S_2(x,y)}\right)^\gamma\right)m + k_2\left(\frac{S_1(x,y)^\gamma}{S_2(x,y)}\right)(m - M) + k_2\left(\frac{S_1(x,y)^\gamma}{S_2(x,y)}\right) \times M \tag{3.16}$$

Where K1, $k2$, and γ are predefined constant and could be varied between [0 1].

M signifies the minimum scalar value within the frame window, m represents the mean vector value within frame window. S_1 and S_2 are the standard deviations of primary and secondary windows respectively.

Original image
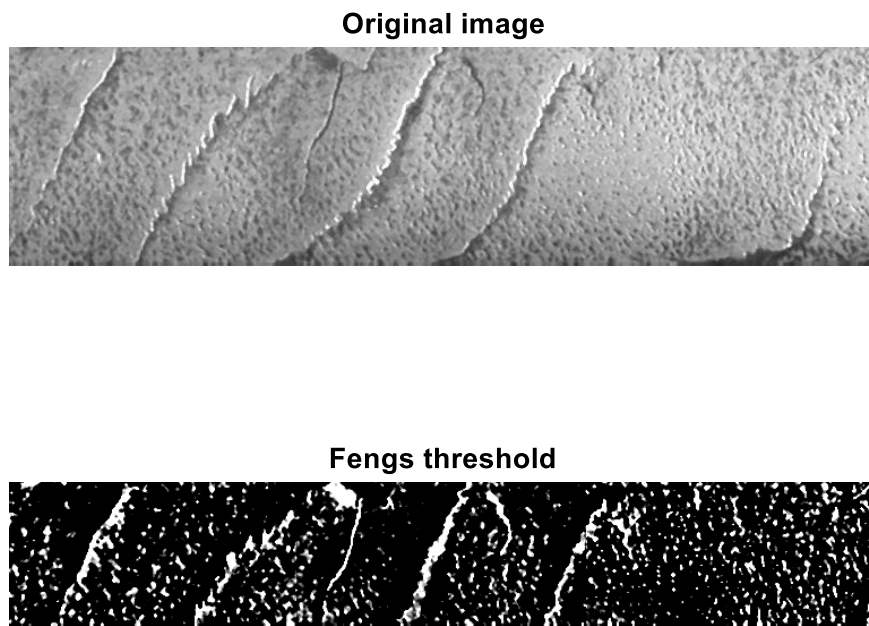
Fengs threshold

Figure 3. 9: Shows the performance of Feng's thresholding on Lab-2 image.

3.3.2.3 Object and background occurrence probability based segmentation

Zhang, W., et al (2014) exploits Occurrence Global Threshold (OGT) method which focuses on local dim regions identified using a black top hat transform. A threshold is selected as the level of intensity (ranging 0-255) which satisfies equation 3.17 defined as a function of background occurrence probability (ω_0), object occurrence probability(ω_1), background mean (μ_0) and object mean (μ_1)levels.

$$\omega_0 \omega_1 (\mu_0 - \mu_1)^2 |_{l=T_0} = \max[\omega_0 \omega_1 (\mu_0 - \mu_1)^2] \tag{3.17}$$

Final threshold $f_{image}(x, y)$ value is further offset by an acceleration parameter (t_a) of which a suitable choice will result in removal of false defects while too high or too low a value result in exclusion and inclusion of pixels respectively. In this thesis $t_a = 0.1\% \times T_0$ for all data set samples.

$$T = T_0 + t_a \tag{3.18}$$

$$f_{image}(x, y) = \begin{cases} 0 & if x \leq T \\ 1 & if \ x > T \end{cases} \tag{3.19}$$

Where T is the segmentation threshold value of occurrence probability method. Figure 3.10 10 shows the performance of OGT on Lab-2 image, and it can be seen that there exists a good agreement compared to AIT method although the latter is not favoured due to the iterative nature of the approach.

Original image
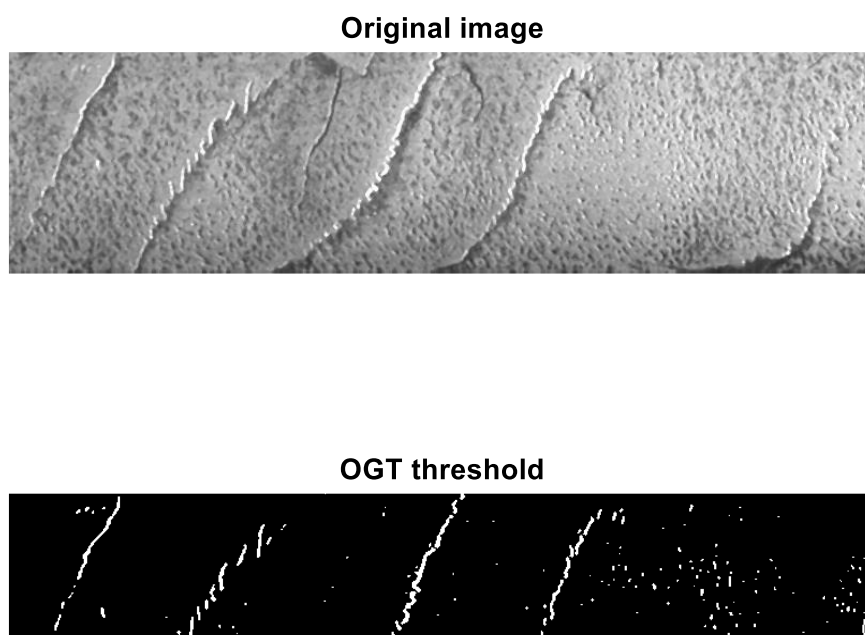
OGT threshold

Figure 3. 10: Shows the performance of Occurrence Global thresholding on image Lab-2.

3.3.3 Defect identification by Feature Extraction (FE)

After successful segmentation of defect regions, the detection of true RCF damage is mainly possible by precise examination of features particular to RCF defects in rails. A variety of methods are explored for this purpose is this section, ranging from Gabor texture analysis, morphological FE, and Standard Deviation Histogram Shape Distance.

3.3.3.1 Texture based feature extraction

Based on the assumption that the presence or absence of a defect is defined by the texture energy, Illonen, J., Kamarainen, J.K., & Kalviainen, H. (2005) explored the application of a 2D Gabor filter, defined as the product of a Gaussian envelope and a sinusoidal wave of specified

frequency and orientation. A thresholding operation is performed based on the energy of the extracted texture suspected to be RCF defect. As shown in equation 3.20, the sharpness along the minor and major axis of the Gabor filter is controlled by parameters א and ב respectively.

$$\varphi(x, y; f_0, \theta) = \frac{f_0^2}{\pi ב א} e^{-\frac{f_0^2}{ב^2}x'^2 + \frac{f_0^2}{א^2}y'^2} e^{j2\pi f_0 x'} \tag{3.20}$$

Where f_0 is the maximum frequency and θ is the orientation of the gabor filter and $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$

The associated energy $E(x, y)$ of each texture is dependent on maximum allowable frequency and orientation of the Gabor filter estimated in accordance with equation 3.21 as the square of Gabor filter response $R(x, y)$ after convolution with the input image.

$$E(x, y) = R(x, y)^2 \tag{3.21}$$

The filter maximum frequency and orientation that corresponds to the maximum energy is further binarised according to the thresholding function in equation 3.22.

$$B(x, y) = \begin{cases} 1, & if \ T > E(x, y) \\ 0, & otherwise \end{cases} \tag{3.22}$$

Figure 3.11 below, shows the result of applying a Gabor filter on Lab-2 sample image. The filter frequency and maximum energy are set to 50Hz and 0.1 respectively for most precise extraction of crack feature. It is worth mentioning that equation 3.20-3.22 are locally implemented as functions with block size set to be the same as that of the LN method.

Original image

Texture based feature extraction
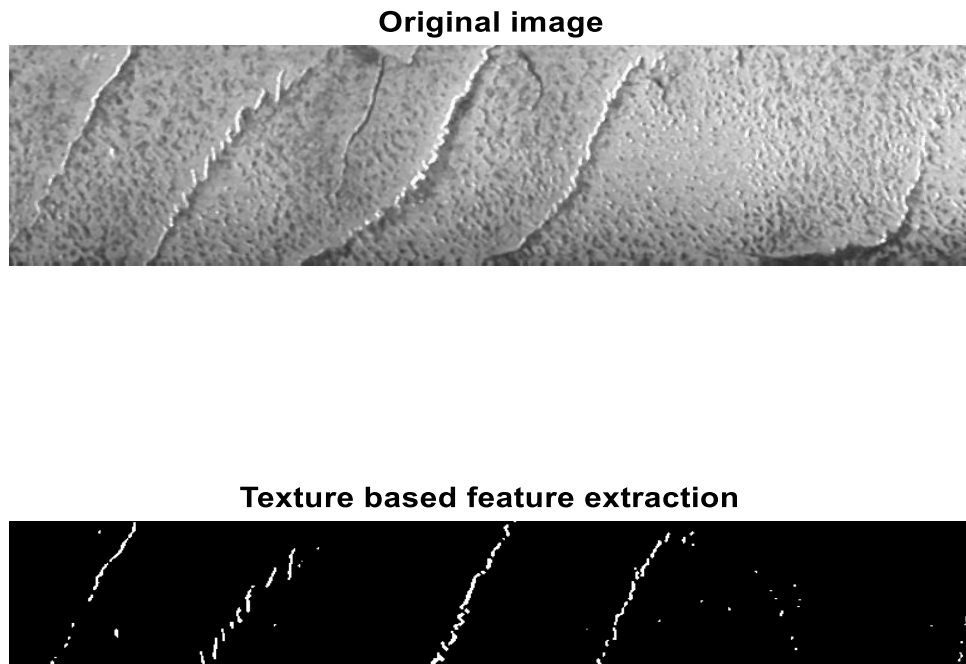
Figure 3.11: Shows the result of Texture based feature extraction on Lab-2 image. Figure

3.3.3.2 Standard deviation of shape histogram distance

In the works of Zhang, W., et al (2014) successful classification of crack feature was presented as a measure of spread within boundary coordinates of the segmented defect under investigation. This method proposes that distance distribution of irrelevant objects are heterogeneous and as a result the estimates of the standard deviation of such an irregular object is observed to significantly be larger than those of actual defects. For each blob (connected objects in a binary image) its corresponding central pixel location (centroid) can be estimated as described in equation 3.23.

$$x_c = \frac{1}{Nb}\sum_{i=1}^{Nb} x_i \; ; \; y_c = \frac{1}{Nb}\sum_{i=1}^{Nb} y_i \qquad (3.23)$$

Where Nb is he total number of boundary coordinates for each blob.And x_i, y_i are the boundary coordinates for $i = 1,2,3,...,Nb$

By computing the Euclidean distance between the centroid and each boundary coordinate (d_i), the shape distance histogram of the blob is evaluated according to equation 3.24.

$$p_i = \frac{N_{di}}{Nb}$$
(3.24)

Where N_{di} is the number of pixels with the same di to the centroid. p_i is the corresponding probability. And Nb is the total number of boundary coordinates of the bolb.

The standard deviation of the shape distance histogram is evaluated according to equation 3.25, and the result of coins.jpg image is presented in Figure 3.8. The simulated results confirm that a standard deviation of less than one corresponds to an irregular shape while above this threshold signifies a near perfect circular blob.

$$\sigma = \sqrt{\frac{1}{Nb}(p_i - \mu_p)}$$
(3.25)

Where μ_p is the average of all pixel boundary pixels of the blob in consideration.

Original image
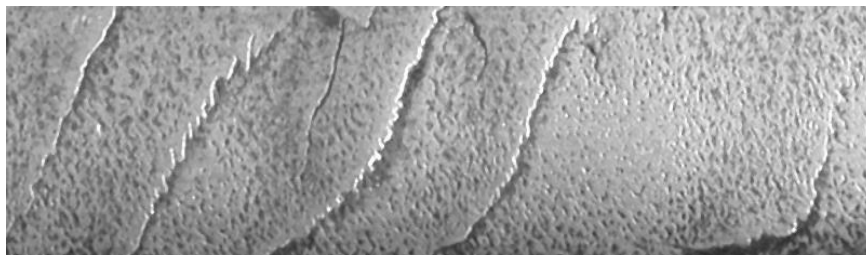
SDSHD based feature extraction
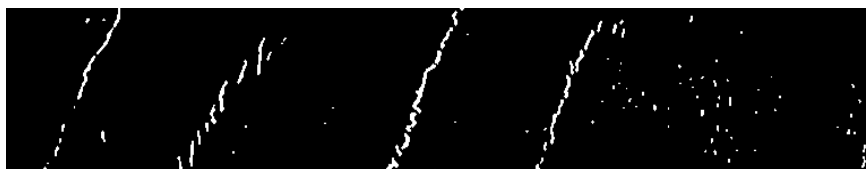
Figure 3. 12: Shows the output of SDHSD feature extraction on Lab-2 image.

3.3.3.3 Geometrically unacceptable (true false detection) based feature extraction

In Li, Q., & Ren, S. (2012), investigation of the extraction of defect based on the morphological features of RCF in rail has been analysed. This entails the elimination of irrelevant objects by considering a combination of geometrical and/or morphological features. In this thesis the minimum area, minimum length and orientation of the defect being investigated are used to establish a criterion for true detection defined in equations 3.26-3.29.

$$\text{FE}_{Area} = \begin{cases} 1, & if\ Area > Area\_threshold \\ 0, & otherwise \end{cases} \tag{3.26}$$

$$\text{FE}_{Orientation} = \begin{cases} 1, & if\ Orientation\ != 0^o, 90^o, 180^o, 360^o \\ 0, & otherwise \end{cases} \tag{3.27}$$

$$\text{FE}_{Length} = \begin{cases} 1, & if\ Length > Length\_threshold \\ 0, & otherwise \end{cases} \tag{3.28}$$

For each defect being considered a true positive detection is only established when equation 3.29 is stratified.

$$\text{FE}_{Area} \times \text{FE}_{Orientation} \times \text{FE}_{Length} = 1 \tag{3.29}$$

3.3.4 Generating geometrical statistical data of true defects

Of interest to this research are some geometrical data related to each detected defect. These include maximum surface length, area, and orientation. The significance and application of such geometrical data in predictive maintenance will be explored in subsequent chapter(s) of the thesis. The mathematical methods related to these morphological operations responsible for generating geometrical statistical data are readily available in MATLAB image processing toolbox, and can be traced to much earlier works of Hort, T as discussed in Nixon, M., & Aguado, A. (2008).

3.3.4.1 Area

Depending on the total number of pixels contained within each detected defect, MATLAB morphological 'bwarea' returns a scalar value that specifies the actual number of pixels in a particular region of interest of a binary image by weighing different patterns of pixels differently.

3.3.4.2 Orientation

Based on the elliptical equivalent representation of each detected defect (i.e. ellipse with the same second-moments as detected defect), 'bworientation' returns a scalar that specifies the angle between the horizontal *x*-axis and the major axis of the equivalent ellipse. Typical values of the 'bworientation' function are in the range of -90 to 90 degrees as depicted in Figure 3.13 below.
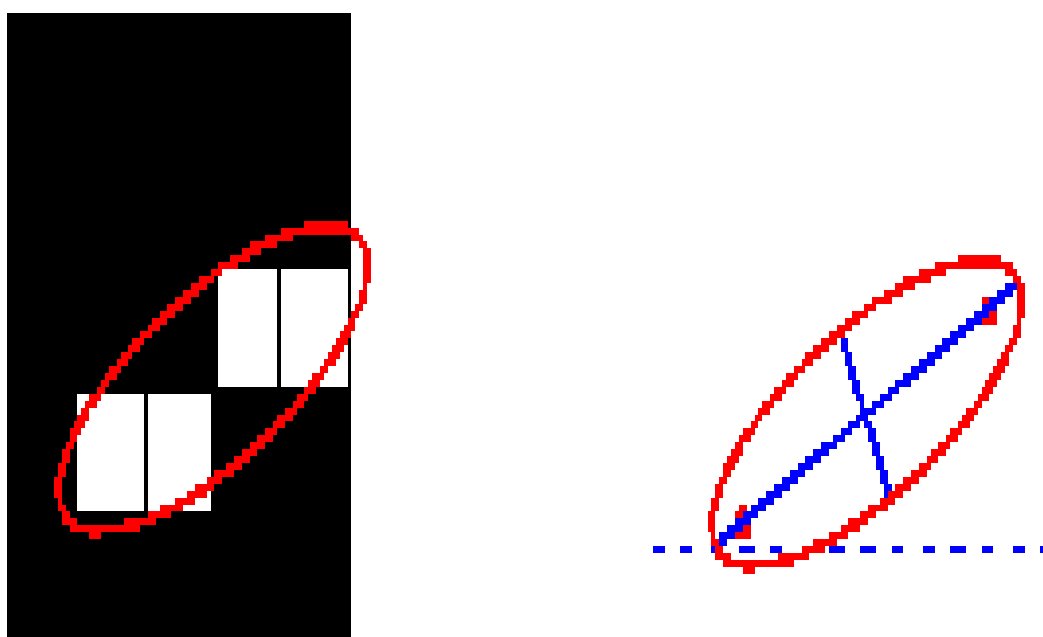
Figure 3. 13: Determining morphological orientation of blob. (Mathworks,2017).

3.3.4.3 Maximum surface length

The maximum length of each blob is computed based on its extracted boundary coordinates according to equation 3.30 describing the distance between any two Cartesian coordinate points.

$$d_i = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \qquad (3.30)$$

Where N is total number of boundary co-ordinates for a defect, and $i = 1,2,3 \dots N$

3.4 Comparative study and analysis of data set 1 and 2

In this section, a comparative study of both field and laboratory acquired images is presented, highlighting the performance of each method within the enhancement, segmentation, feature extraction stage of image processing; based on a quantitative measure of the Peak Signal to Noise Ratio (PSNR) for indicating the performance of each enhancement stage. For the segmentation stage, consistency in number of segmented regions with validation data for lab acquired image samples 1 and 2 is used as indicator of best performance. A similar approach has been adopted in the case of feature extraction, with emphasis on the morphological details related to RCF defects. These morphological considerations include consistency in terms of irregularity in shape of RCF defects, and exclusion of false defects (geometrically false).

3.4.1 Enhancement of laboratory acquired images (data set 1)

3.4.1.1 Smoothing by median filtering

The significance of smoothing an image, besides noise reduction include the actualisation of less pixelated image. Among the various methods of performing smoothing, the statistical approach is mostly preferred for replacing the intensity value of a seed pixel by a single intensity value commonly evaluated by the mean or median operator. The median filter has been adopted in this study mainly because of the following reasons; no blurring effect on details of the input image, faster speed compared to average or Gaussian filter equivalent, and easier implementation and optimisation compared to Low pass filter application for the same purpose. It is observed that in the case of laboratory acquired images using a local window of size $1 \times N$ (be it parallel or perpendicular to rolling direction) yields little or no observable differences in enhancement performance. But in the case of an $N \times N$ window, controlled samples are observed to exhibit diminishing enhancement with increasing values of N. For N ranging between $2 - 5$, this gives highest Peak Signal to Noise Ratio (PSNR) signifying improved enhancement peformance.

Furthermore, a smooth image with better perception of details/information is obtained especially in the case of input images with unfirm reflection property (samples Lab-1 and

Lab-2 Figure 3.14 a and b). However in the case of images with poor illumination properties such as samples Lab-3 and Lab-4 (see Figure 3.14 c and d), a significant amount of blurring is observed especially within the vicinity of the of excess light reflection.

Original image   **Median filtered image**

(a)

Original image

Median filtered image

(b)
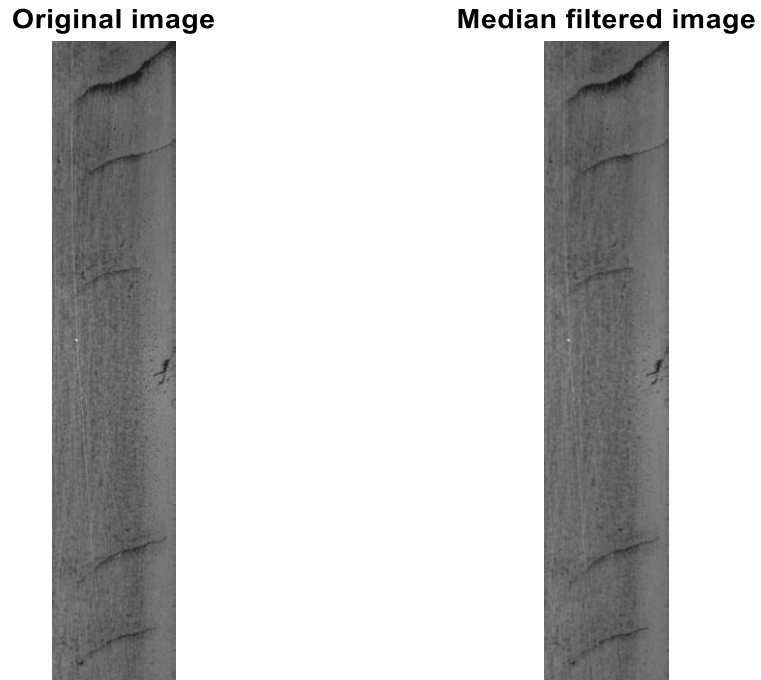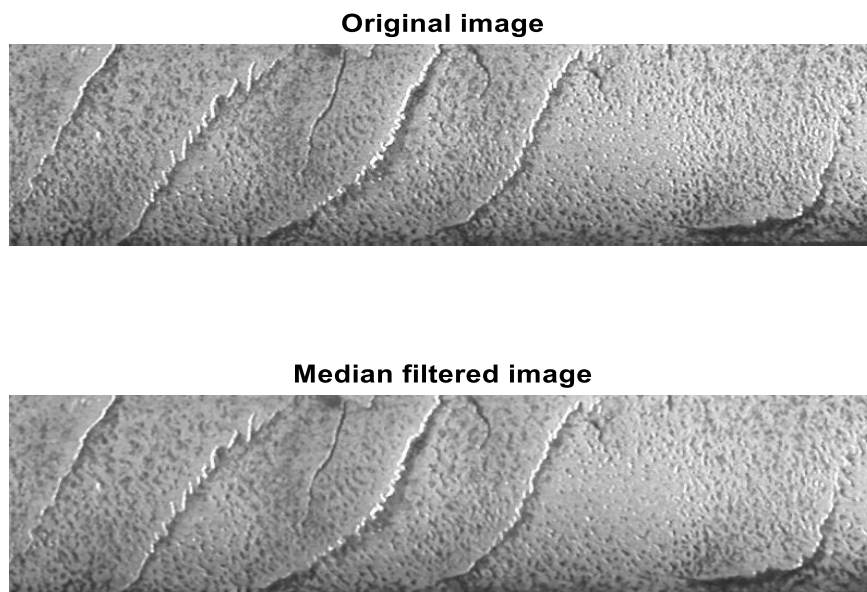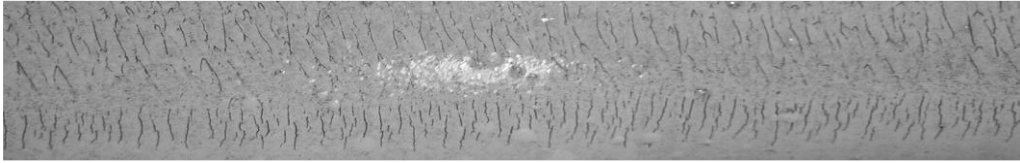
Original image

Median filtered image
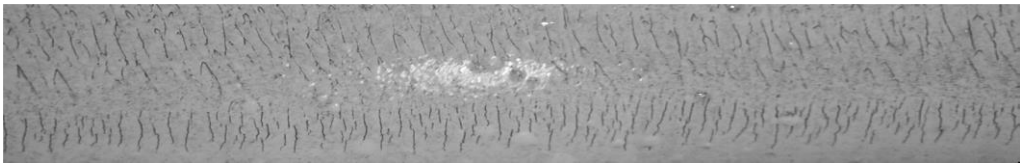
(c)

Original image
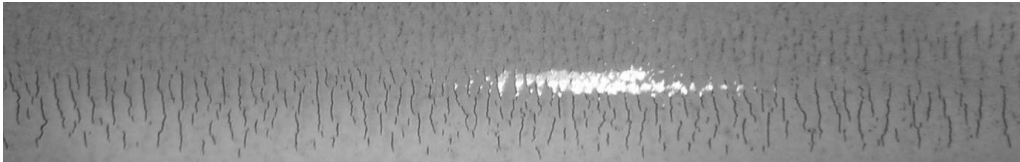
Median filtered image
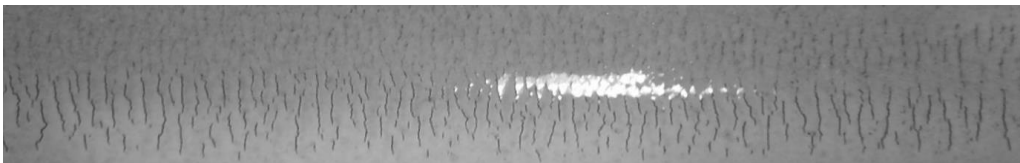
(d)

Figure 3. 11: Presents simulated results for median smoothing of all laboratory acquired images Lab-1 to Lab-4 denoted by a-d, using optimised wind size N=3.

In the case of field acquired images, the median filter simulated results for the subset of data set 2 (containing one image from each distinct level of damage) is observed to respond in a similar manner to laboratory acquired images. For block size ($N > 10$), it is observed to cause bluring effect which corresponds to PSNR saturation levels shown in Figure 3.15. Simulated results from investigating this subset suggest an optimum value of N=5 for all field acquired images. From the Figure below the tolerance of increasing N on PSNR is observed to be less sensitive for moderately and lightly damaged rails. This can further be related to the dominant defect type for the image being investigated, such that clusters of fatigue damage in image samples (FA-M1 and FA-L1) are less sensitive to increase in block size compared to spall type in FA-H1.

It is obvious that the influence of excess light reflection (at the top left corner of all samples of data set 2) is still persistant even after smoothing. This observation (see Figure 3.15) suggests that more sophisticated illumination invariant methods are required for both laboratory and field acquired image enhancement.

Original image **Median filtered image**
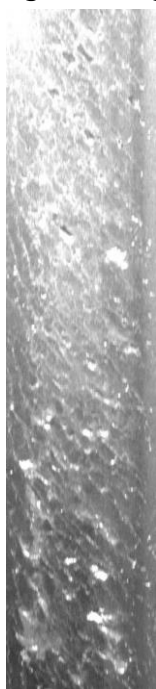
(a)

Original image   **Median filtered image**

(b)
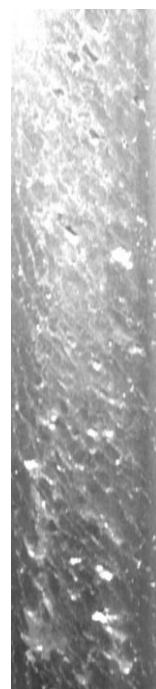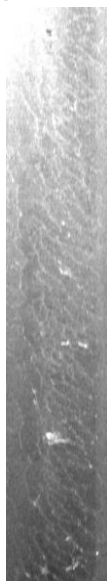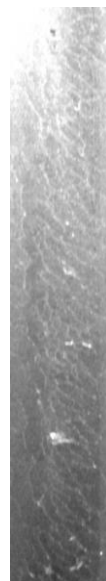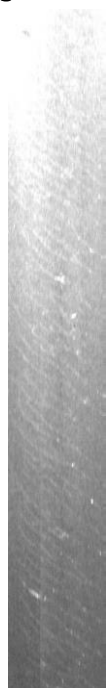
Original image   **Median filtered image**
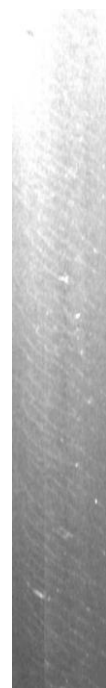
(c)

Figure 3. 12: Median filter response for (a) FA-H1, (b)FA- M1, and (c) FA-L1, data set 2.

3.4.1.2 Local Normalisation (LN)

The appropriate choice of the desired mean (RM), and variance (RV) of the output image in addition to optimal window size (bs_LN) are the influential parameters of the LN function investigated in this section of the study. The optimal values of these influential parameters are also observed to vary depending on the data set in consideration. In general, it was observed that an increase in RM corresponds to overall increase in brightness of the image while increasing RV corresponds to increased sharpness of edges. The window size of the of the LN function is initially observed to correlate with the severity of damage, that is the more damaged the sample is, the larger block size required for good enhancement. For each of the images investigated, the performance of this approach is observed to be influenced by a more complex interaction between dominant defect type, severity of damage, and presence/absence of grind marks, scratches, and rust. In the case of Lab-1, Lab-3, and Lab-4 (all samples mostly dominated by fatigue damage like defects), a range for the mean and variance which provides best enhancement is observed to be $RM \cong 0.9 - 1.2$, and $RV \cong 2.0 - 2.1$ respectively. A block size ranging between $N = 22 \ and \ N = 44$ for all samples of data set 1 is also established. However, an exception is in the case of Lab-2, which requires modification of RM and RV to 2.1 and 1.7 respectively to improve the performance of LN function on this sample due to the presence of rust. This modification can be justified in order to increase the overall brightness of the image to exclude effect of rust (dark spots) and simultaneously lowering the variance to avoid considering false edges. In Figure 3.16, the simulated results of all samples in data set 1, confirm that local dim regions correspond to suspect defect regions, while local bright regions correspond to rail head.

Original image **Locally normalised image**
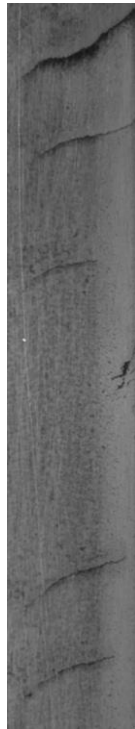
(a)

Original image

Locally normalised image

(b)
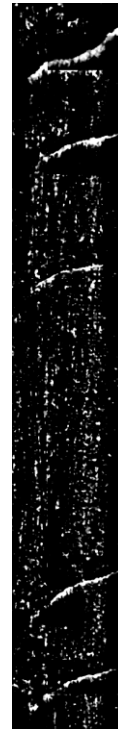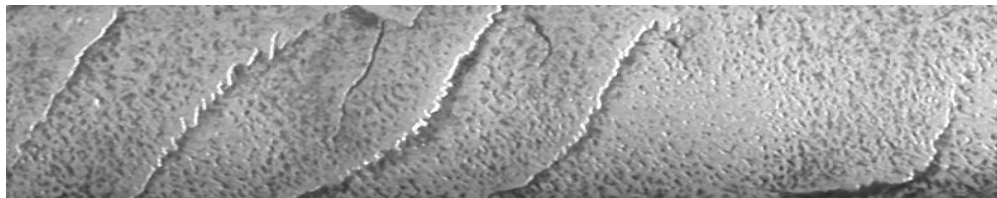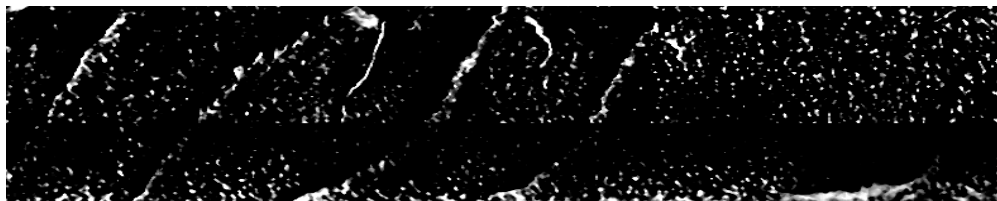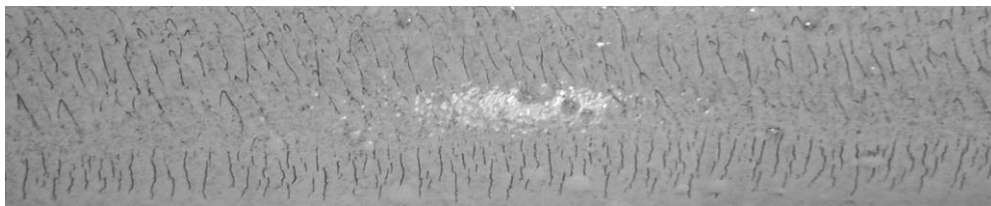
Original image

Locally normalised image
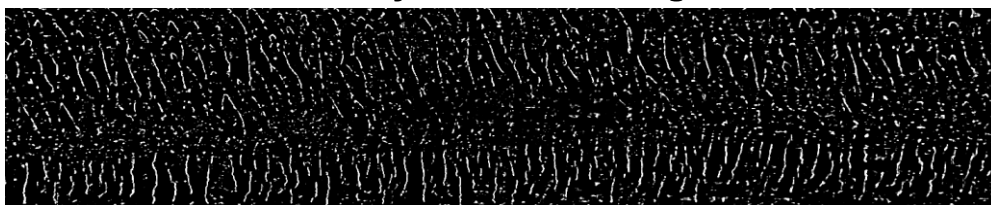
(c)

Original image
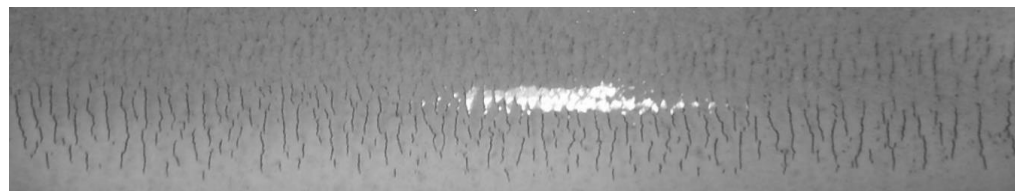
Locally normalised image

(d)

Figure 3. 13: Presents the simulated results for LN method on data set 1.

In the case of field acquired images, the influence of RM, RV, and N have been concluded to be very similar to the case of laboratory acquired images especially the relationship between dominant type of defect and appropriate window size (N). However, an observable difference for data set 2 compared to controlled images of data set 1 is that the RM value is optimal at 0. This difference is considered to be as a result of the excess illumination of light at the top left corner of the samples in addition to rail steel reflection property for data set 2. Higher RV corresponds to sharper defect enhancement, with optimum values in the range of 1-2, while RV>5 is observed to enhance grind marks, scratches, even in regions of excess light illumination. In Figure 3.17, the simulated results confirm that local dim regions correspond to rail head, while local bright regions correspond to defect region. The illumination invariant nature of this method has encouraged investigating its combination with other enhancement methods to be discussed in subsequent sections of this thesis.

Original image   **Locally normalised image**
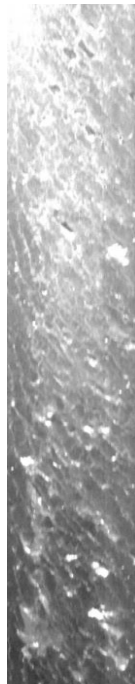
(a)

Original image **Locally normalised image**

(b)
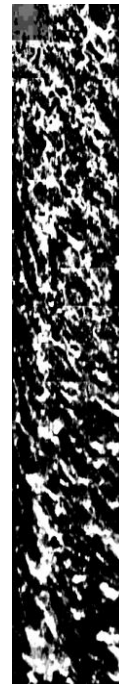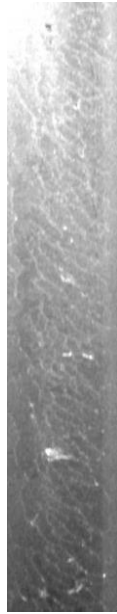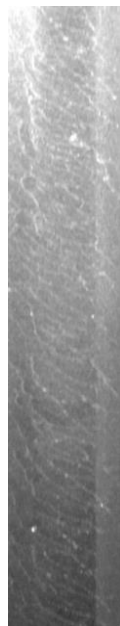
Original image **Locally normalised image**

(c)

Figure 3. 14: Presents the simulated LN results of (a) H1, (b) M1, and (c) L1.

3.4.1.3 Linear Moving Average Filter (LMAF)

Employing LMAF for enhancement of laboratory acquired images, it can be seen that this method's performance is strongly dependent on the spatial distribution of intensities in the input image and dominant type of defect. In general, the even distribution of light reflection across the rail head supports improved enhancement of suspect defect region in this method. It was also observed that for Lab-1 image, defects are enhanced as local dim regions, while Lab-2, Lab-3, Lab-4 result in defect regions enhanced as local bright regions. This clarifies the influence of rust and regions of excess light illumination in the latter samples. The block size of LMAF function is observed to be most influenced by the type of defect within the image such that all laboratory acquired images (which are dominated by fatigue cracks) can be effectively enhanced with a block size of no more than 32×32 window size. It is also worthy to note that for sufficiently large averaging window size, the computational time of the LMAF (see Table 3.8) method is significantly reduced to almost 50% (around 30 minutes) at the expense of depreciating PSNR values. Further examination of the mean intensity and variance also recommend that low values of RM combined with high RV results in better identification of defect regions in data set 1. The optimum RM and RV values for data set 1 have been determined by trial and error and summarised in the comparative study of section 3.4. Considerations of the high computational requirement of the LMAF method (see section 3.6 for summary of recorded execution time) suggests that it is not beneficial for the intermediate objective of this thesis (automatic detection of RCF). The simulated results for all data set 1 samples are presented in Figure 3.18 for a block size of N=32. The suspect defect regions are identified as local dim regions for Lab-1, Lab-3, Lab-4 and vice-versa for Lab-2 (due to increase in RM).

Original image   **LMAF enhanced image**
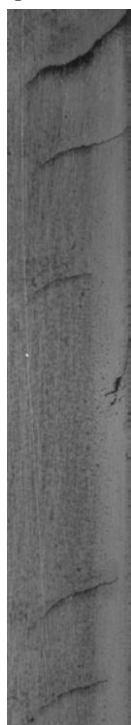
(a)

Original image

LMAF enhanced image

(b)
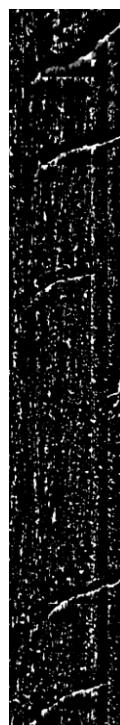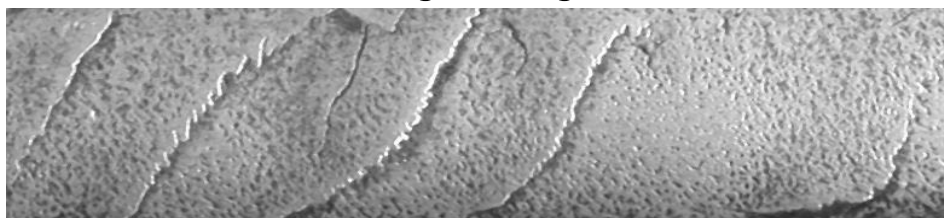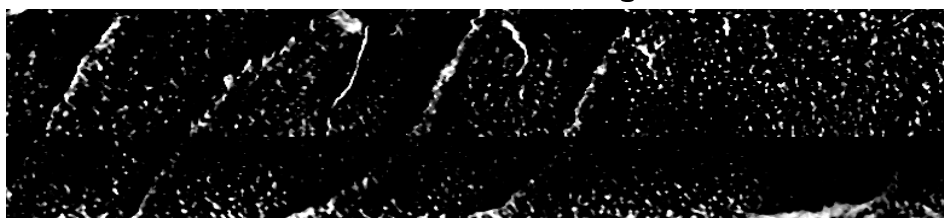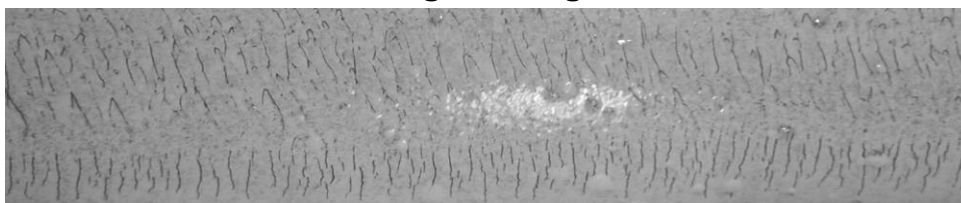
Original image

LMAF enhanced image
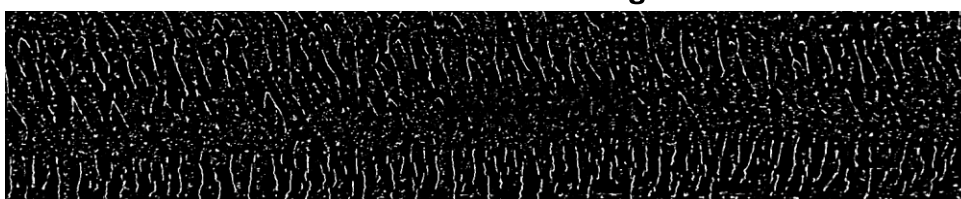
(c)

Original image
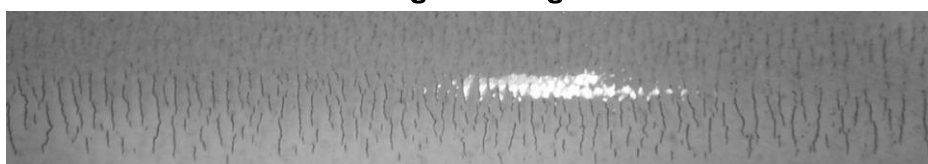
LMAF enhanced image

(d)

Figure 3.18: Simulated LMAF results for (a) Lab-1, (b) Lab-2, (c) Lab-3 and (d) Lab-4.

The LMAF approach to enhancing field acquired images is concluded to be similar to that of controlled samples. It is also observed that the method's efficiency is highly correlated to the type of defect within the input image. Such that, damage as a result of removal of rail head material (spalling) is best enhanced by larger window sizes of no more than 64×64 pixel block, while moderately, and lightly damaged images correspond to 32×32, and 16×16 pixel block sizes respectively. The high computational time of the method (as detailed in section 3.6) can be minimised by increasing the size of the averaging window at the expense of diminishing PSNR values. For all images of data set 2, an RM of 0 and RV of 1 has been utilised which corresponds to suspect defect regions represented by local bright regions and background as local dim regions, as presented in Figure 3.19 below. As in the case of data set 1 the consideration of computational requirement has discouraged further analysis of the LMAF method on field acquired images.

**Original image**   **LMAF enhanced image**
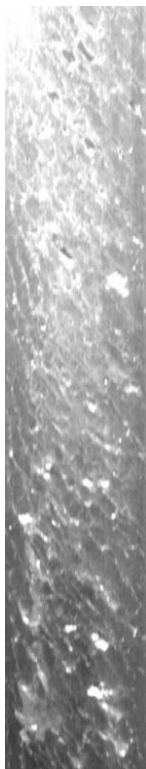
(a)

Original image   **LMAF enhanced image**

(b)
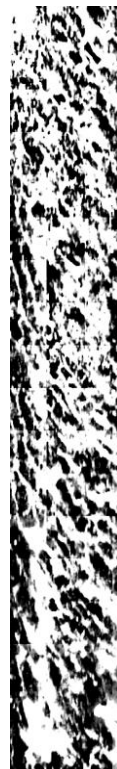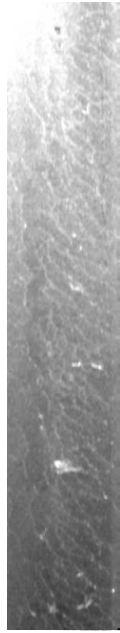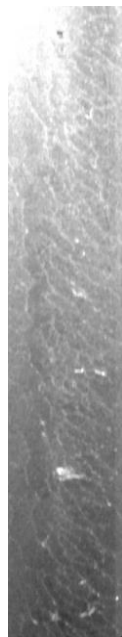
Original image   **LMAF enhanced image**

(c)

Figure 3.19: Simulated LMAF results for (a) FA-H1, (b)FA- M1, and (c) FA-L1.

3.4.1.4 FFT enhancement

In the case of FFT enhancement of laboratory acquired images, the most influential factor is a combination of the appropriate window size and enhancement factor k. These influential parameters have been observed to be consistent for all laboratory acquired images as bs_FFT=2, while k=0.1. Further examination of the influence of RM and RV on the performance of FFT enhancement suggests that the value of k is proportional to RM of the input image, such that increasing the RM would require a greater k value. In Figure 3.20, the simulated results of input images Lab-1 to Lab-4 are initialised with RM and RV as 0.9 and 2.1 respectively, and this consistently enhances suspect defect regions as local bright regions. For this method it is worthy to note that the influence of excess illumination of light on the rail head is also observed to cause blurring effect within such vicinities leading to possible false defect detection.

Original image   **Fast Fourier Transform enhanced image**
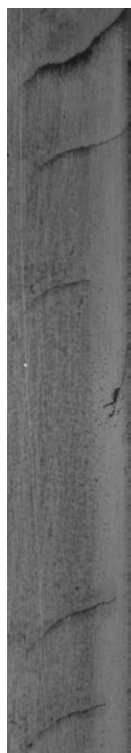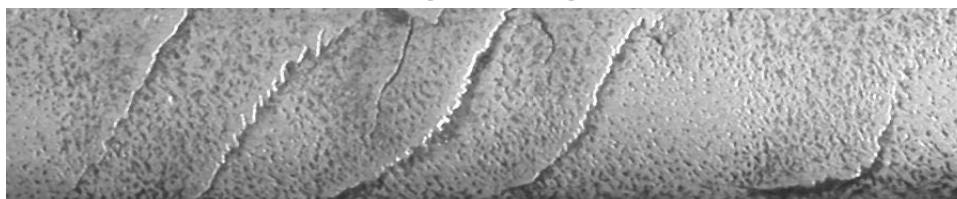
(a)

Original image

Fast Fourier Transform enhanced image

(b)
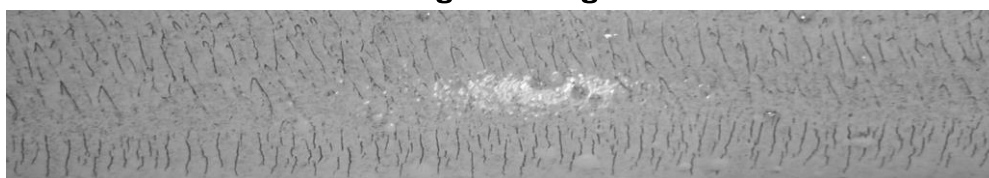
Original image

Fast Fourier Transform enhanced image

(c)

Original image
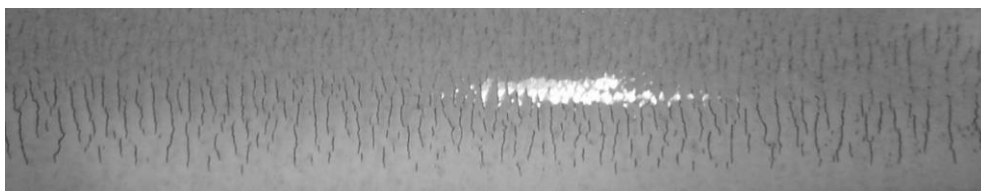
Fast Fourier Transform enhanced image
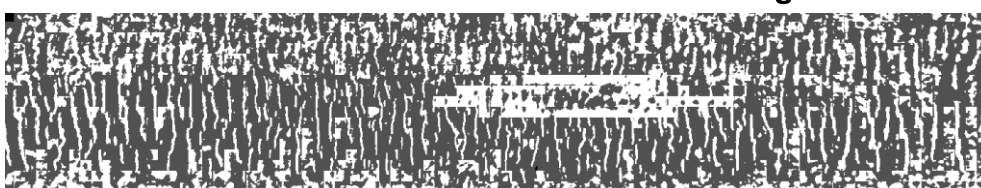
(d)

Figure 3. 15: Presents the simulated FFT results of data 1.

In the case of FFT enhancement of field acquired images, only the block size of the frequency enhancement is observed to differ as a function of dominant defect type. The optimal value of k for all samples has been determined by trial and error as 0.1975. However, for each image, the appropriate widow size varies as a function of severity of damage. Considering the FA-H1, FA-M1, and FA-L1 images are best enhanced using the FFT window size of 16, 8, 4 pixel block size respectively, the same enhancement factor of k=0.1975 is consistently applied for all samples. Further examination of the influence of RM and RV on the performance of FFT enhancement of field acquired images suggests that the value of k is also proportional to RM of the input image. That is, a higher value of RM will require a higher value of k. In Figure 3.23, the simulated results of input images H1, M1, and L1 are initialised with RM and RV as 0 and 1 respectively, the output consistently enhances suspect defect regions as local bright regions. The sensitivity of the FFT method to type of defect present within the input image, and also

the severity of damage including and poor illumination property of the rail surface has discouraged its application for automatic enhancement of rail defect images. However, it is worth noting from Figure 3.21, that the influence of excess illumination of light on the rail head is avoided by optimising the RM and RV parameters of each image to a value of 2.1 and 0.9 respectively.

Original image   **Fast Fourier Transform enhanced image**
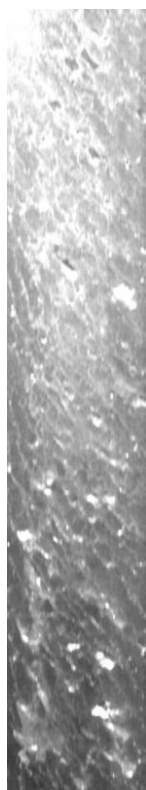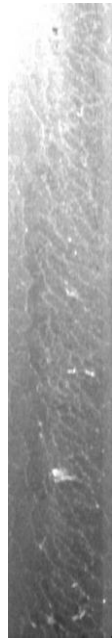
(a)

Original image　　　**Fast Fourier Transform enhanced image**

(b)
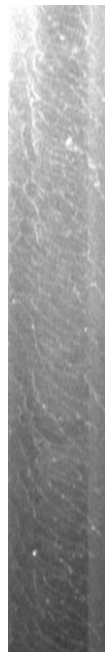
Original image　　　**Fast Fourier Transform enhanced image**

(c)

Figure 3. 16: Presents the simulated FFT enhancement results of data set 2- sub set (one image from each distinct damage severity levels).

3.4.1.5   Visibility measure enhancement

In the case of visibility measure enhancement technique, the most influential parameter is the block size of visibility function (N). And it is observed to diminish enhancement performance of the method with decreasing values of N regardless of severity of damage within the input image being investigated. The block size sensitivity analysis related to this method as depicted in Figure 3.24 confirms the inverse relationship between PSNR and block size N. Optimal values for each of the samples in data set 1 have been decided based on the minimum variation of PSNR values as block size increase (to saturation point). It can be seen that Lab-1, Lab-2, Lab-3, Lab-4 optimum block size range 10-22, 18-22, 16-22, 11-22 with variations less than 0.001% for all samples respectively. Although it was observed that a block size of N≤ 22 results in 40% reduction in computation time, the upper bound $N = 22$ is adopted for automatic defect detection of RCF in rails. To further the investigation on the influence of RM and RV, simulated results generally suggest that increasing values of RM within the range of (0.5-2) with constant RV significantly improves the elimination of background information. For Lab-1 it was also observed (see Figure 3.25a) that adequate enhancement of defect regions is attainable with RM=RV=2. While in the case of Lab-2 to Lab-4 (images with rust or excess light illumination), an optimum value of RM=0.9 and RV=2.1 results in better enhancement of suspect defect regions as depicted in Figure 3.22 b-d.

Original image   **Visibility measure enhanced image**
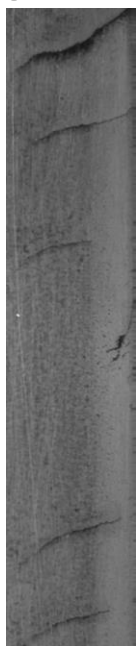
(a)

Original image

Visibility measure enhanced image

(b)
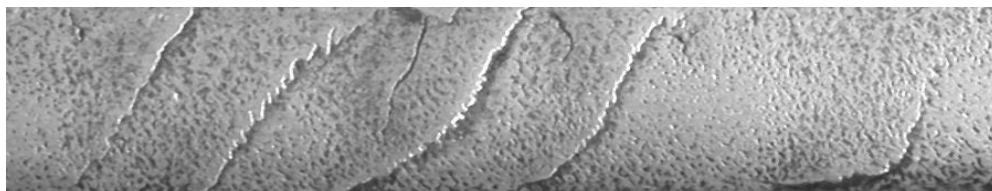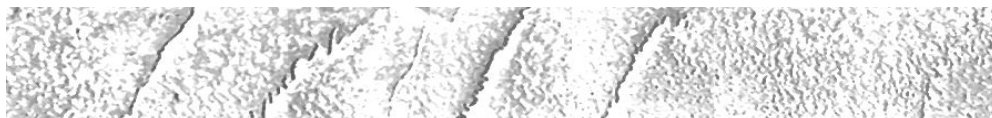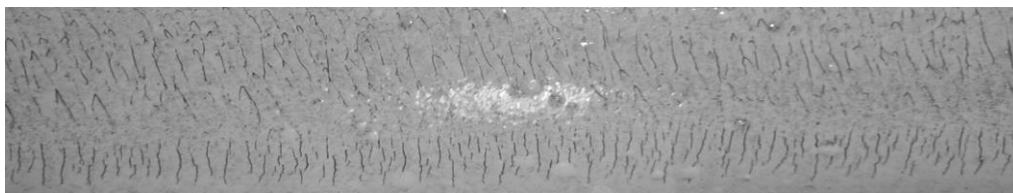
Original image

Visibility measure enhanced image

(c)

Original image
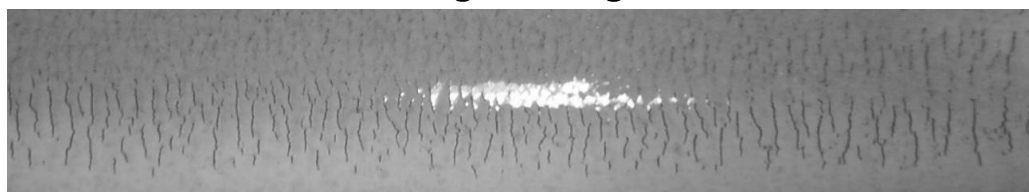
Visibility measure enhanced image

(d)

Figure 3. 17: Presents the simulated visibility measure enhancement results of data 1.

In the case of field acquired images FA-H1, FA-M1, FA-L1 as depicted in Figure 3.23 a-c, a similar observation with the case of controlled samples can be established. The only observable difference in the output image is the opposite characterisation of suspect defect regions as local dim regions while background information as local bright regions. This is attributed to the choice of RM=0 and RV=1 for this data set as discussed in the case of LN method. The appropriate choice of N equals the window size of LN enhancement technique as discussed in the previous section. Further investigation of the influence of RM and RV on the performance of visibility measure enhancement suggests that RM is inversely proportional to PSNR values.

Original image   **Visibility measure enhanced image**
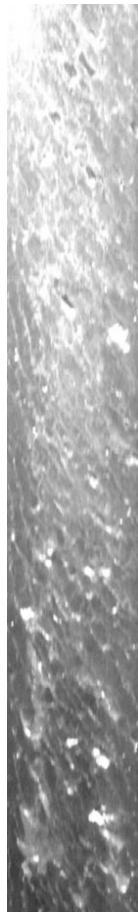
(a)

Original image   **Visibility measure enhanced image**

(b)
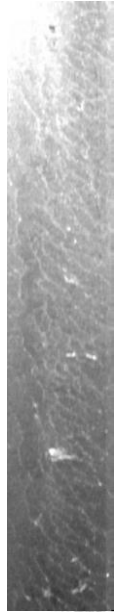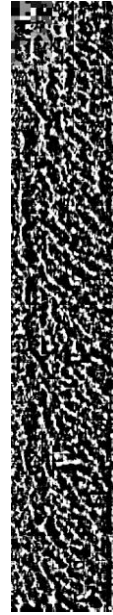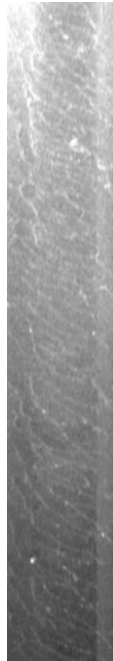
Original image   **Visibility measure enhanced image**

(c)

Figure 3. 18: Presents the simulated visibility measure enhancement results of data set 2- sub set (one image from each distinct damage severity levels).

Table 3.1 below summarises the performance of each enhancement method based on estimates of peak signal to noise ratio. It can be generally concluded that visibility measure enhancement provides better defect region enhancement for all laboratory acquired images, such that an appreciation in PSNR of about 60-70% is recorded compared to median smoothing. While about 2%-9% improvement in enhancement performance is observed compared to LN method. And in comparison to FFT method, the VM method offers 10%- 20% improvement in enhancement performance. An exception is in the case of LMAF, which is observed to actually generate higher PSNR compared to visibility measure by approximately 12% only for Lab-1 sample, while other samples (Lab-2 to Lab-4) fall short by approximately 4%, 1%, and 0.99% respectively.

Image	Median filter (dB)	LMAF enhancement (dB)	Visibility measure enhancement (dB)	Local normalisation enhancement (dB)	FFT enhancement (dB)
Lab-1	15.39	46.81	41.09	40.19	36.86
Lab-2	15.22	45.91	48.20	43.99	37.14
Lab-3	15.01	44.32	44.78	41.66	37.96
Lab-4	15.17	45.87	46.33	43.14	36.52

Table 3. 1: Summarises the PSNR values for data set 1 using optimal influential parameters for each image.

Table 3.2 for field acquired images suggests, (as in the case of Table 3.1) that the visibility measure enhancement method outperforms enhancement by smoothing, LMAF, LN, and FFT by approximately 50%, 0.87%, 0.02%, 13.80% respectively. It can also be deduced from the Table that a negligible shortcoming of the visibility method (0.02%) is evident compared to the illumination invariant LN method. However, the inability of the LN method to enable quantification of its performance for some images (namely FA-H2 and FA-L1) discourages its sole application for automatic enhancement of rail RCF defect images. It is also worth mentioning that higher consistency between samples of data set 2 is observed compared to

data set 1. This is attributed to the consistency in acquisition conditions of the data set 2, which

cannot be said for the controlled images of data set 1.

Image	Median filter enhancement (dB)	LMAF enhancement (dB)	Visibility measure enhancement (dB)	Local normalisation enhancement (dB)	FFT enhancement (dB)
H1	22.01	44.47	44.86	44.87	38.67
H2	21.95	43.19	43.51	NaN	39.37
H3	22.25	43.86	44.35	44.37	36.83
H4	22.47	43.76	43.95	43.97	38.23
H5	23.00	44.41	44.83	44.85	39.49
M1	22.67	44.45	44.34	44.36	38.45
M2	22.37	44.36	44.25	44.28	37.11
M3	22.82	43.52	43.41	43.43	39.02
M4	22.66	44.14	44.03	44.11	36.61
M5	22.04	44.42	44.31	44.34	39.55
L1	22.10	44.81	43.88	NaN	37.73
L2	22.43	43.16	43.24	43.28	38.46
L3	22.77	42.11	42.19	42.20	39.39
L4	24.18	43.71	43.79	43.81	37.82
L5	22.55	43.52	43.60	43.64	39.72

Table 3. 2: Summarises the PSNR values for data set 1 using optimal influential parameters for each image.3.4.2 Segmentation results of laboratory acquired images.

3.4.2 Segmentation of laboratory acquired images (data set 1)

Segmentation by thresholding being one of the simplest methods of separating images into regions of interest, in preparation of better feature extraction is explored in this section of the thesis. The mathematical models detailed in section 3.3.2 are investigated on data set 1 and 2, and the performance of each method is established based on the consistency of number of regions obtained by each method is summarized in Table 3.3 below. It is of interest that Lab-1 and Lab-2 samples (of data set 1) with known number of defects are used as indictors of performance for Wolfs segmentation method, Occurrence probability segmentation method (OGT), and Iterative (AIT) segmentation method. The influential parameters related to the improvements of Wolf's algorithm presented by Feng in equation 3.16 have been simplified based on recommendations of Mostafa, A., et al (2016) as k1=0.01, k2=0.5 and $\gamma = 1$. In the case of OGT method a corresponding primary and secondary windows size of 32×32, and 16× 16 pixel block has been adopted as for all samples as recommended by (Zhang, W., et al 2014). While the AIT method is initialised to use two partitions for the iterative determination of mean intensity value. It is worth mentioning that using small partitions for the AIT method takes longer iterations to converge, even though more accurate.

Image	Feng's algorithm	OGT algorithm	Automatic iterative thresholding
Lab-1	20	7	7
Lab-2	308	26	21
Lab-3	449	253	231
Lab-4	412	229	212

Table 3. 3: Summarises the number of segmented regions for data set 1 using optimal influential parameters for each image.

From Table 3.3 above, it can be concluded that Feng's algorithm consistently generates much higher number of regions compared to OGT and AIT by almost 65%, 40%, and 40% for Lab-

1, Lab-3, and Lab-4 samples of data set 1 respectively. This is as a result of the complex tuning of the three predefined constants. This short coming is most pronounced in the case of Lab-2 with observed excess segmented regions of about 300 blobs. AIT and OGT methods have recorded good agreement with no more than 16% variation compared to the validated number of defects of Lab-1 and Lab-2 samples. Further examination of Lab-3 and Lab-4 suggests a variation in number of segmented regions to be no more than 10%.
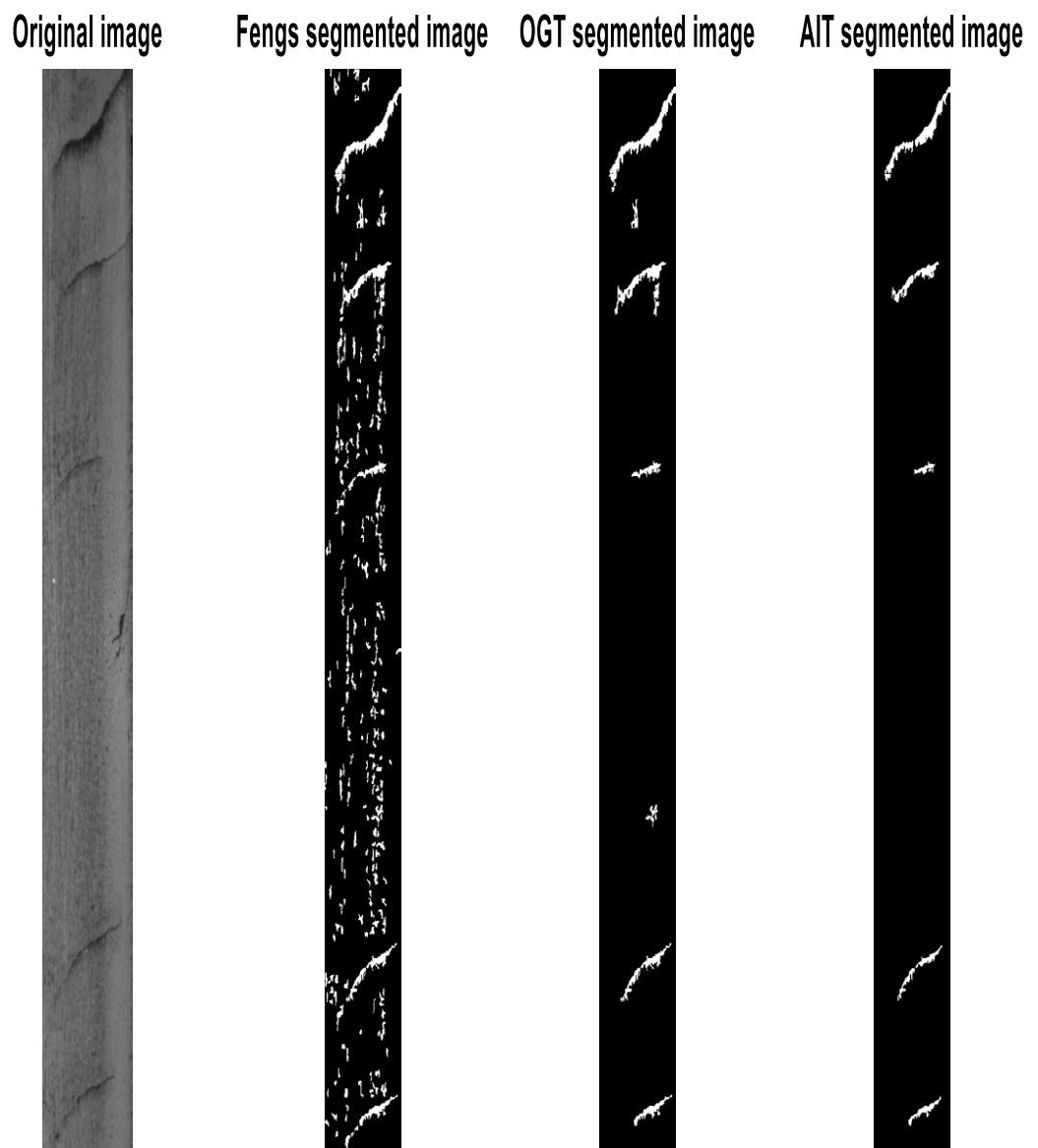
Figure 3. 19: Show the performance of Feng's, OGT, and AIT thresholding on Lab-1 image.

Original image
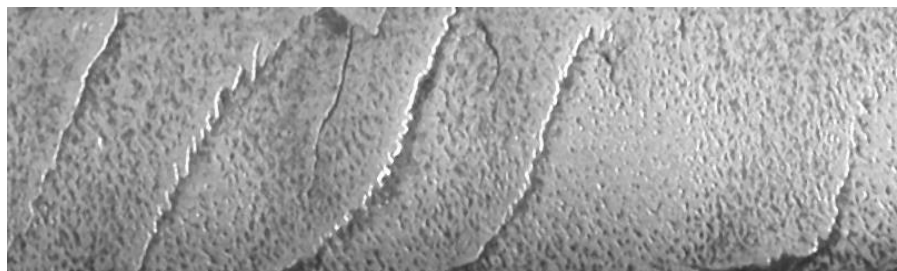
Fengs segmented image
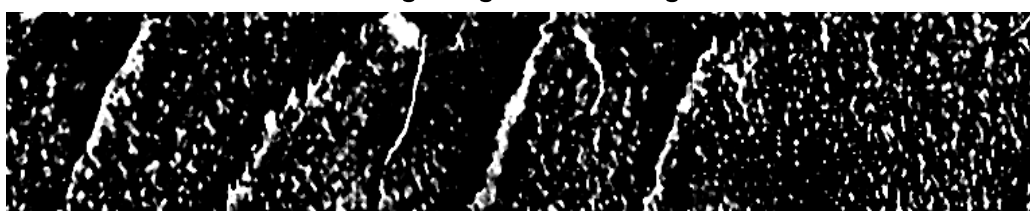
OGT segmented image

AIT segmented image
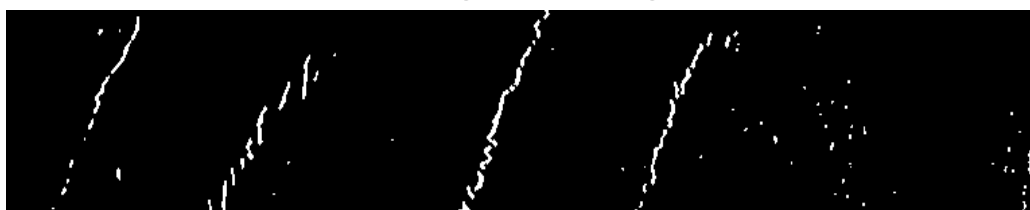
Figure 3. 20: Show the performance of Feng's, OGT, and AIT thresholding on Lab-2 image.

Original image
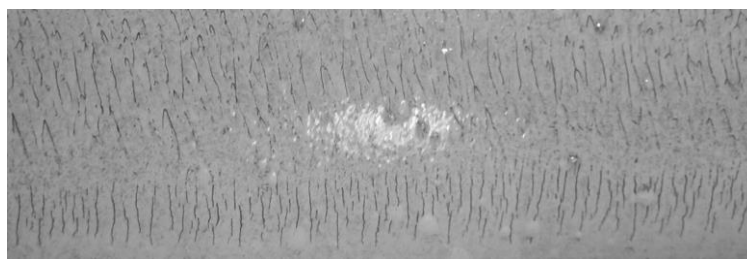
Fengs segmented image

OGT segmented image
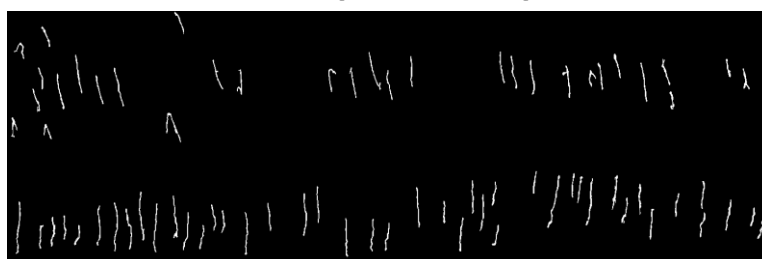
AIT segmented image
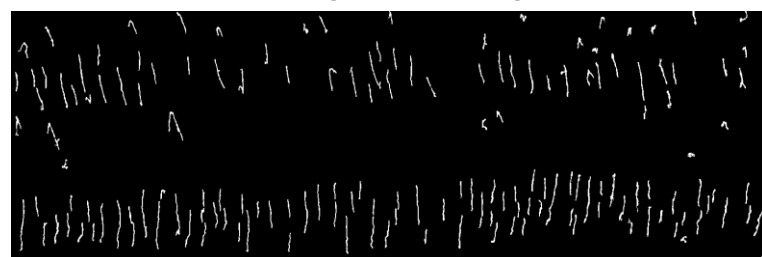
Figure 3. 219: Show the performance of Feng's, OGT, and AIT thresholding on Lab-3 image.

Original image
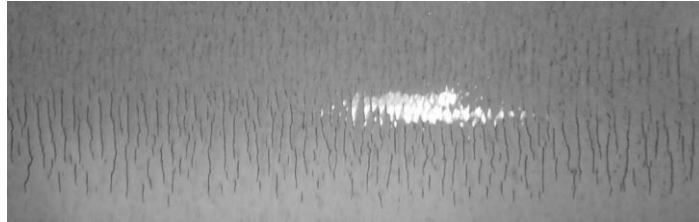
Fengs segmented image
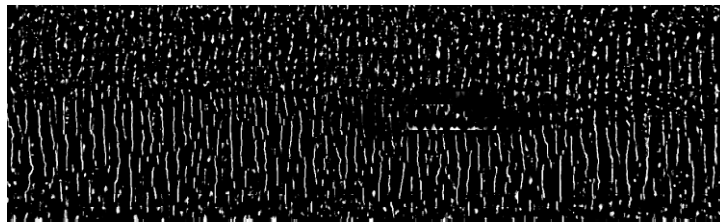
OGT segmented image
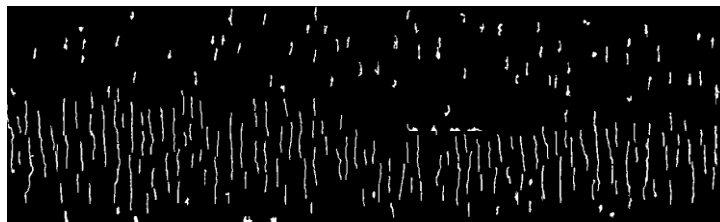
AIT segmented image
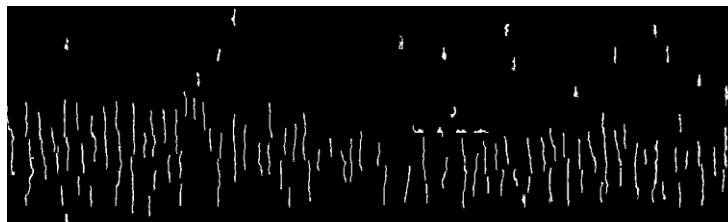
Figure 3. 30: Show the performance of Feng's, OGT, and AIT thresholding on Lab-4 image.

Image	Feng's algorithm	OGT algorithm	Automatic iterative thresholding
H1	561	49	49
H2	874	106	115
H3	708	90	90
H4	1537	176	237
H5	664	100	113
M1	1016	99	90
M2	1131	123	N/A
M3	717	96	101
M4	2000	124	143
M5	614	74	70
L1	799	125	126
L2	1437	182	N/A
L3	1081	131	156
L4	1461	129	167
L5	1654	148	117

Table 3. 4: Summarises the number of segmented regions for data set 2 using optimal influential parameters for each image.

The number of detected defects on field acquired images (subset of data set 2), has also been investigated as summarised in Table 3.4 above. Similar to the previous case (of controlled images) the complexity associated with optimising three variables of Feng's algorithm has discouraged and undermined the results obtained for this method. For simplicity, the predetermined constants k1, k2, and γ have been initialised based on recommendations of Mostafa, A., et al (2016) as k1=0.01, k2=0.5 and $\gamma = 1$. Furthermore, the primary and secondary window size of $N_1 = 64$ and $N_2 = 32$ for samples FA-H1 to FA-H5. And for samples FA-M1 to FA-M5, $N_1 = 45$ and $N_2 = 22$. And in the case of samples FA-L1 to FA-L5, $N_1 = 16$ and $N_2 = 8$, the variation in number of segmented regions for each distinct level of damage is analysed. A range of 0.87-29% is observed for heavily damaged samples. While a range of 2.06%-11.5% and 0.13%-13.8% is observed for moderately, and lightly damaged samples respectively. The consistency in simulated number of segmented regions for OGT and AIT

Incorporating automated rail RCF damage detection algorithms with crack growth modelling

encourages their adoption for automatic detection of RCF. However, the latter is considered

less accurate of the two given that inconclusive results have been realised for FA-M2 and FA-

L2.
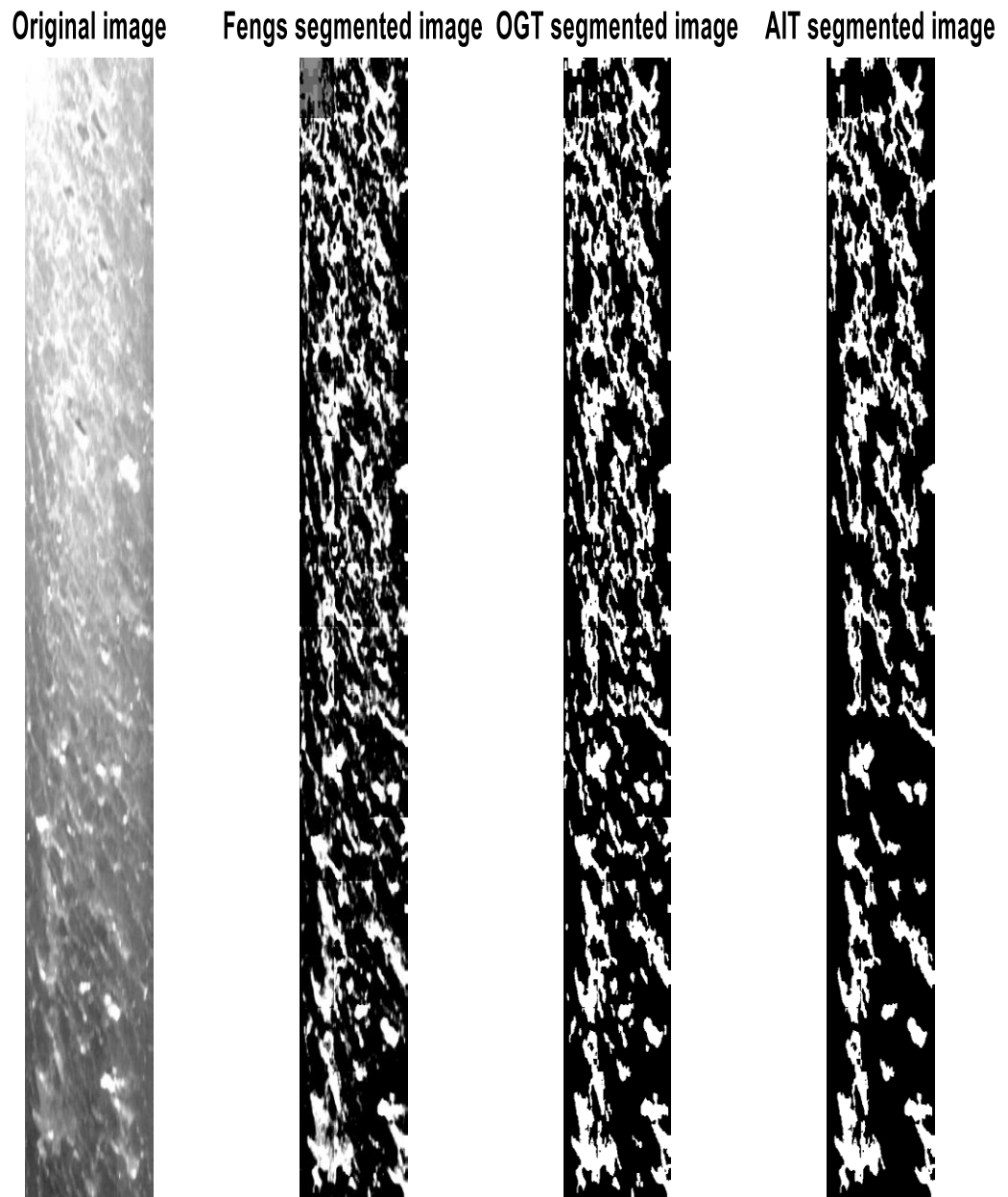
Figure 3.31: Shows the performance of Feng's, OGT, and AIT thresholding on FA-H1 image.
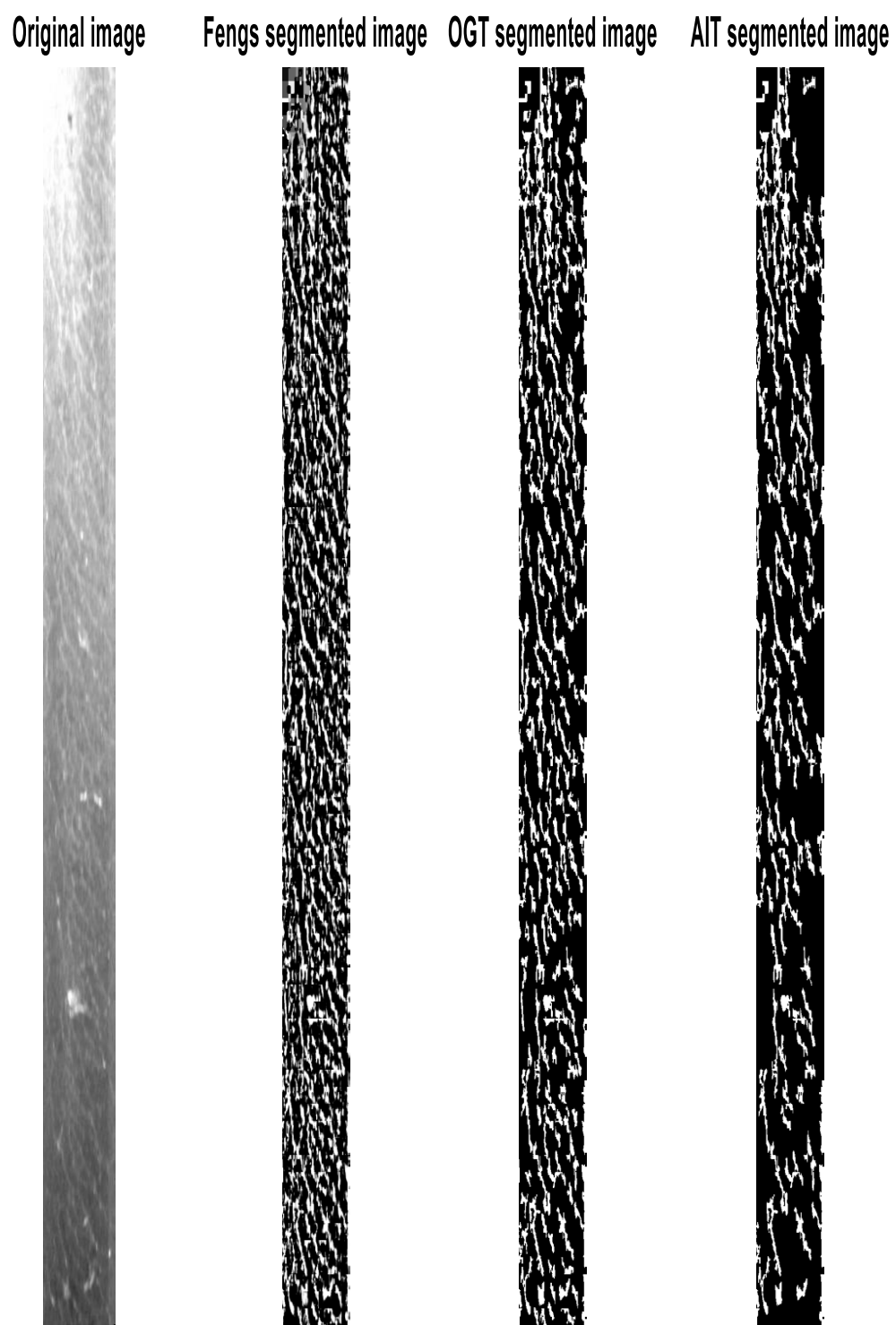
Figure 3.32: Shows the performance of Feng's, OGT, and AIT thresholding on FA-M1 image.
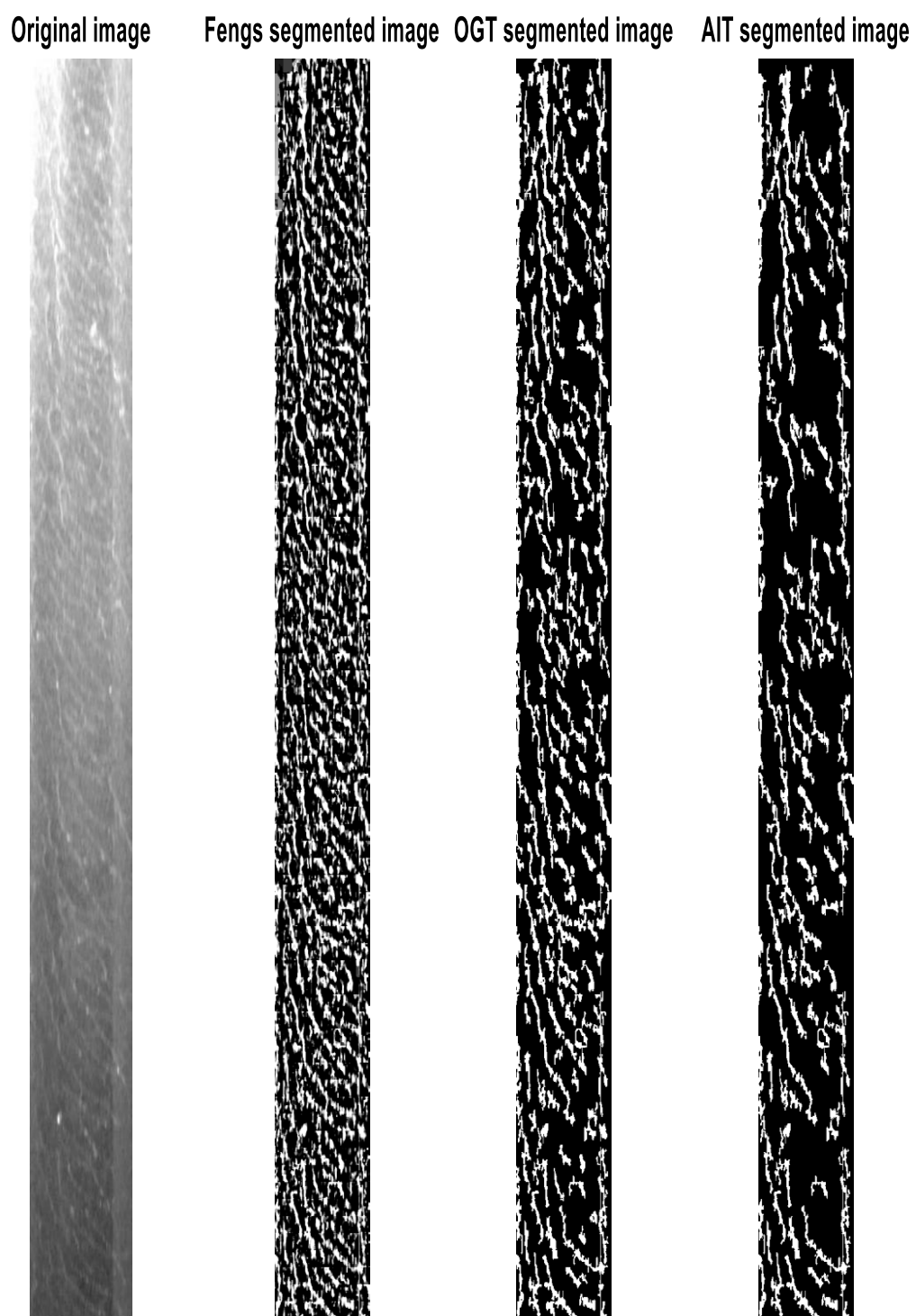
Figure 3. 3322: Shows the performance of Feng's, OGT, and AIT thresholding on FA-L1 image.

3.4.3 Feature extraction

In this section of the thesis the differences in performance of shape distance histogram standard deviation method, morphological characteristic typical to RCF defects, and texture (Gabor) based feature extraction are analysed. The suitability of each method is evaluated as function of the number of suspect RCF features detected. Table 3.5 below, summarises the number of RCF defects detected by each of the aforementioned feature extraction model including the initialised values of influential parameters related to each method. To deepen the understanding of each method's capacity to identify RCF defects, all possible combinations of the enhancement i.e. E1(LN), E2 (VM), E3 (LMAF), and E4 (FFT) with OGT segmentation (S1), and all feature extraction i.e. FE1, FE2, FE3 have been tested on Lab-1, Lab2, Lab3, Lab-4, FA-H1, FA-M1, and FA-L1. The influential parameters related to enhancement (RM, RV, Block size) for each image are also summarised in the table below. Furthermore, in the case of SDHSD method (FE1), a standard deviation threshold value (σ) for each image (see Table 3.5) is recorded. For Gabor method (FE2), a filter frequency of 50Hz and energy threshold of 0.009 is adopted for all image samples. In the case of FE3, the range (Max_area- Min_area) of defect size (pixel unit) is also recorded. It was observed that the difference in number of RCF defects detected is consistent with the performance of the enhancement method adopted regardless of feature extraction method utilised. To this end, Local normalisation and Visibility measure enhancement (E1 and E2) combined with Occurrence global thresholding (S1) in addition to all feature extraction methods (FE1, FE2, and FE3) are observed to give the least variation from expected crack population for both data set 1 and 2. Thus these methods will be considered in section 3.5 for proposing an algorithm for detection of rail surface damage.

Original image   **Texture based feature extraction**   **SDSHD based feature extraction**
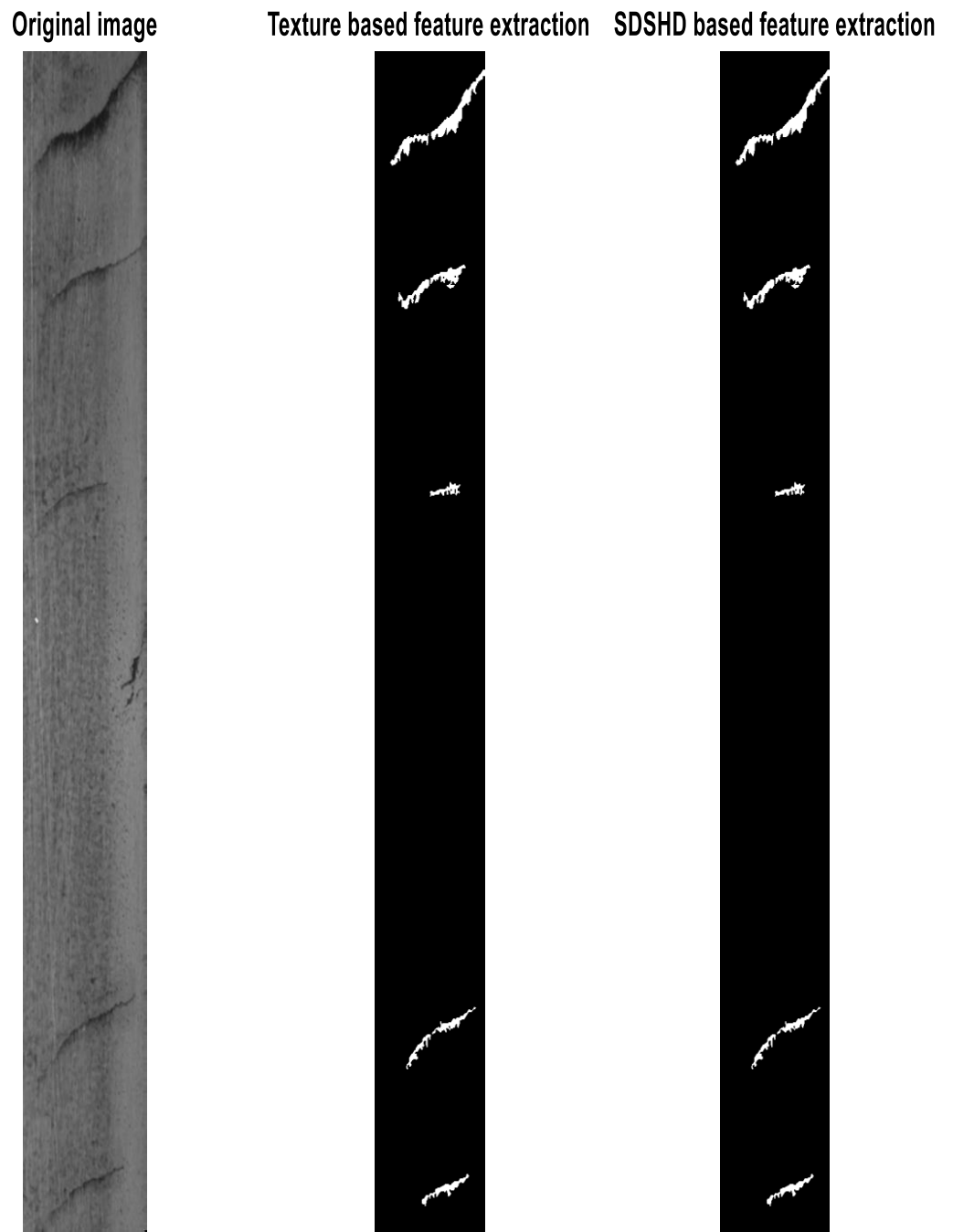
Figure 3.34: Shows the performance of Texture and SDHSD based feature extraction on Lab-1 image.

Original image
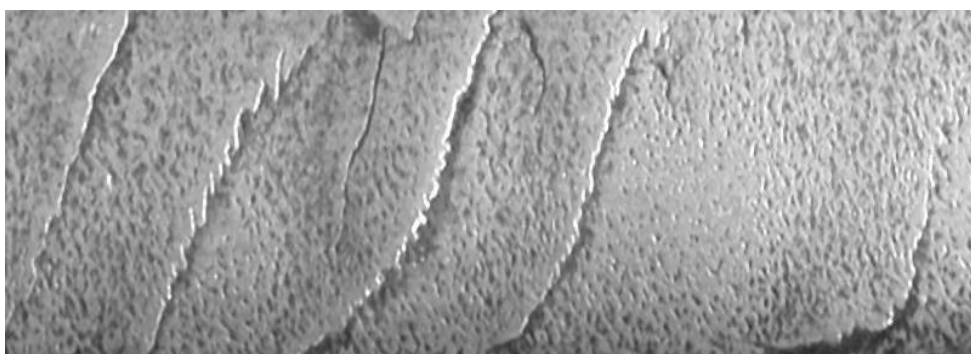
Texture based feature extraction

SDSHD based feature extraction

Figure 3. 35: Shows the performance of Texture and SDHSD based feature extraction on Lab-2 image.

Original image
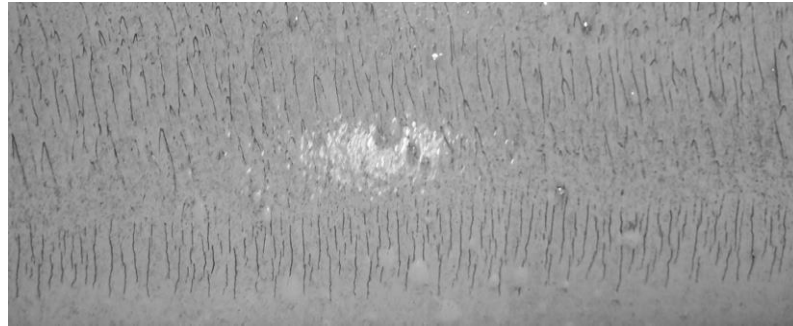
Texture based feature extraction
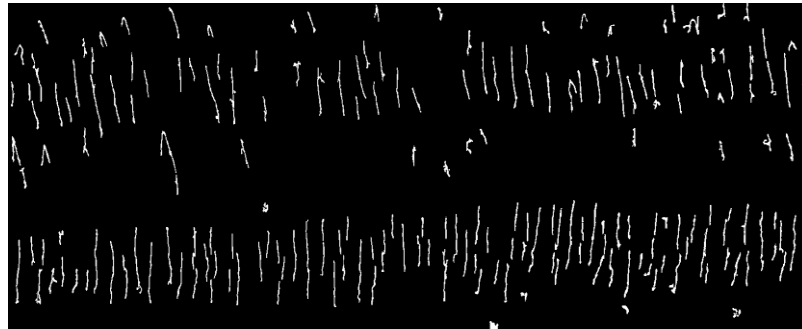
SDSHD based feature extraction
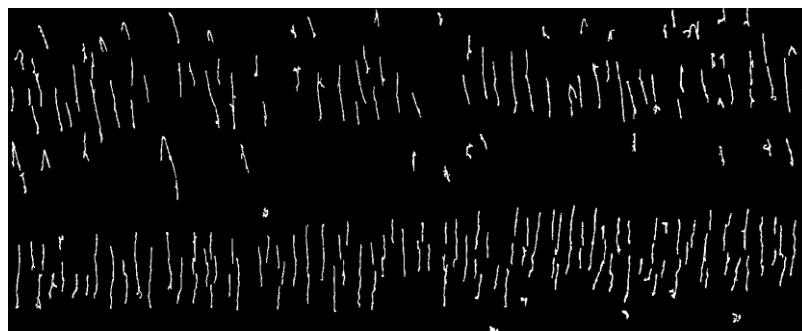
Figure 3.36: Shows the performance of Texture and SDHSD based feature extraction on Lab-3 image.

Original image
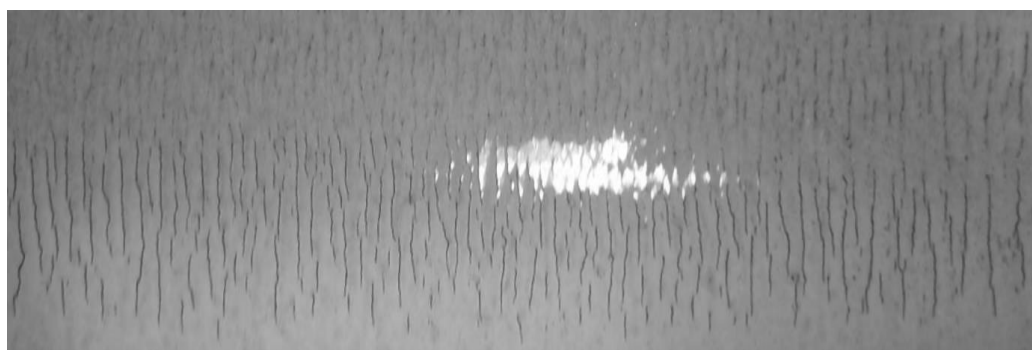
Texture based feature extraction

SDSHD based feature extraction
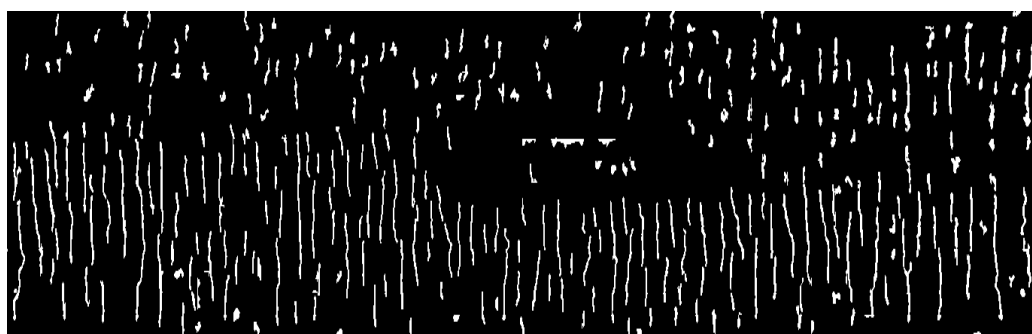
Figure 3. 23: Shows the performance of Texture and SDHSD based feature extraction on Lab-4 image.

Original image   **Texture based feature extraction**   **SDSHD based feature extraction**
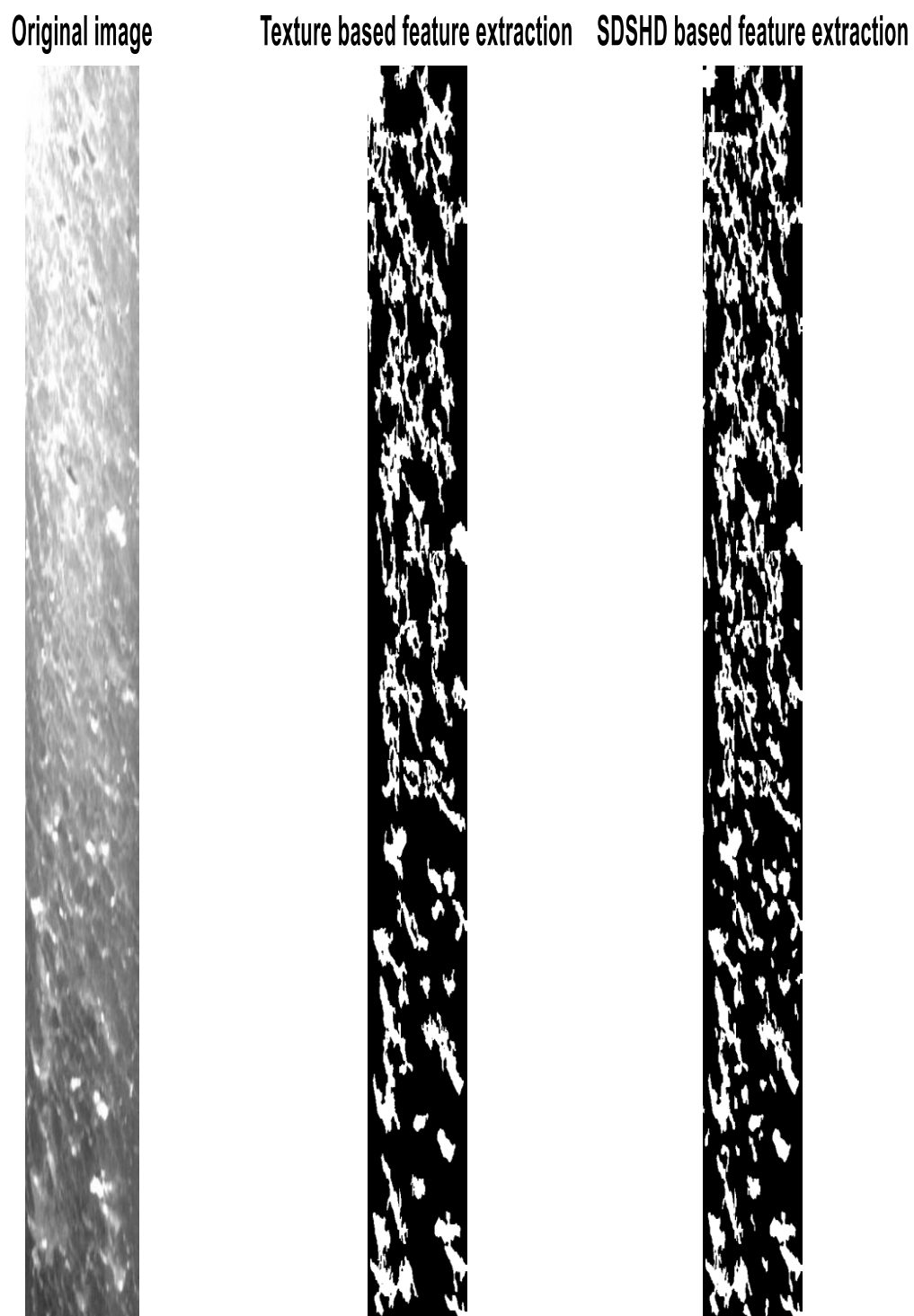
Figure 3. 38: Shows the performance of Texture and SDHSD based feature extraction on FA-H1 image.

Original image   **Texture based feature extraction**   **SDSHD based feature extraction**
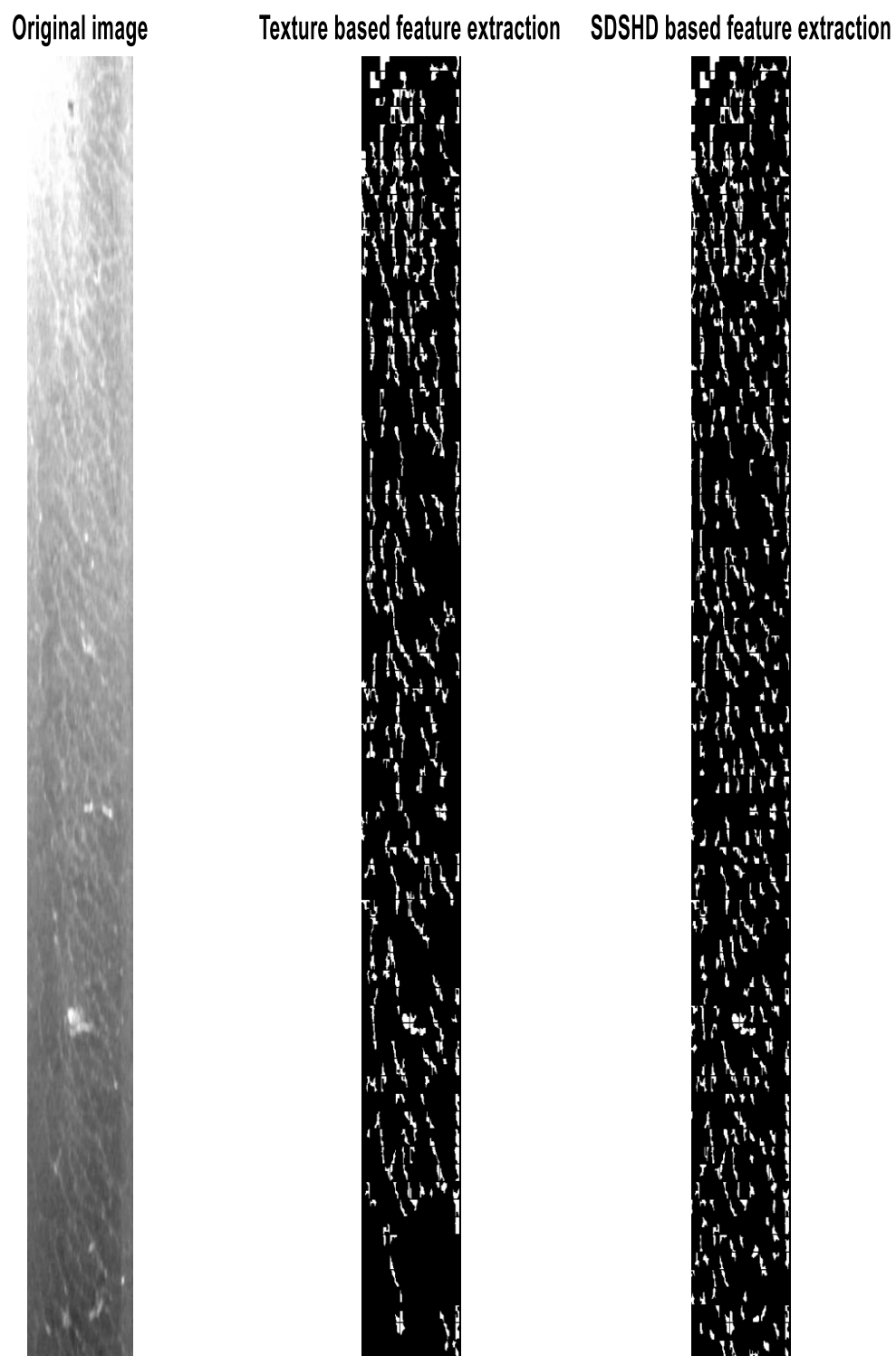
Figure 3. 39: Shows the performance of Texture and SDHSD based feature extraction on FA-M1 image.

Original image   **Texture based feature extraction**   **SDSHD based feature extraction**
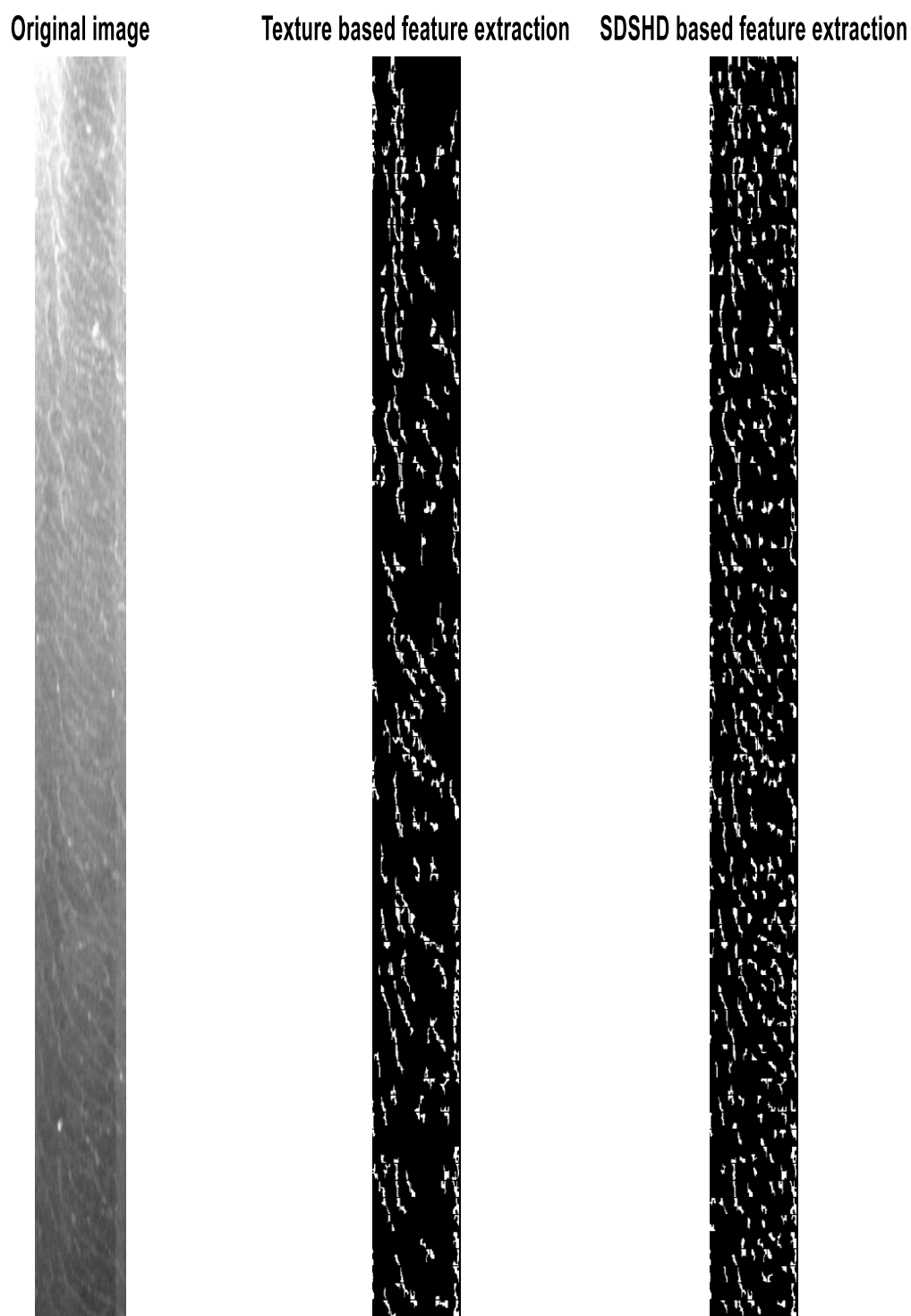
Figure 3. 40: Shows the performance of Texture and SDHSD based feature extraction on FA-L1 image.

Algorithm	Enhancement Parameters: Required mean RM, Required variance RV, and Block size (N)	Feature extraction parameters Standard deviation (FE1) Frequency & energy threshold (FE2) Min & Max area (FE3)	LAB-1 sample # defects detected (Expected crack population =6)	LAB-2 Sample # of defects detected (Expected crack population =5)	LAB-3 sample # of defects detected (Expected crack population $\cong$ 700)	LAB-4 sample # of defects detected (Expected crack population $\cong$ 800)	FA-H1 sample # of defects detected (Expected crack population $\cong$ 27)	FA-M1 sample #r of defects detected (Expected crack population $\cong$ 124)	FA- L1 sample # of defects detected (Expected crack population $\cong$ 128)
E3,S1,FE1	Lab-1:0.9,2.1,32 Lab-2:2.1,1.7,32 Lab-3:0.9,2.1,32 Lab-4:0.9,2.1,32 FA-H1: 0,1,64 FA-M1: 0,1,32 FA-L1:0,1,16	$\sigma <= 5$ $\sigma <= 4,500$ $\sigma <= 12,000$ $\sigma <= 12,000$ $\sigma <= 6,000$ $\sigma <= 7,000$ $\sigma <= 25,000$	16	11	510	454	20	153	144
E3,S1,FE2	Lab-1:0.9,2.1,32 Lab-2:2.1,1.7,32 Lab-3:0.9,2.1,32 Lab-4:0.9,2.1,32 FA-H1: 0,1,64 FA-M1: 0,1,32 FA-L1: 0,1,16	50,0.009 50,0.009 50,0.009 50,0.009 50,0.009 50,0.009 50,0.009	13	11	120	119	21	133	156

E3,S1,FE3	Lab-1:0.9,2.1,32	550-2937	14	10	226	314	19	148	143
	Lab-2:2.1,1.7,32	600-4014							
	Lab-3:0.9,2.1,32	4-150							
	Lab-4:0.9,2.1,32	4-180							
	FA-H1: 0,1,64	200- 22565							
	FA-M1: 0,1,32	10-12644							
	FA-L1: 0,1,16	10-25765							
E2,S1,FE1	Lab-1: 0.9,2.1,32	$\sigma <= 2$	15	8	208	276	25	128	159
	Lab-2: 0.2,1.7,32	$\sigma <= 4$							
	Lab-3: 0.9,2.1,22	$\sigma <= 15$							
	Lab-4:0.9,2.1,22	$\sigma <= 15$							
	FA-H1: 0,1,64	$\sigma <= 3,000$							
	FA-M1: 0,1,32	$\sigma <= 20,000$							
	FA-L1: 0,1,16								
E2,S1,FE2	Lab-1:0.9,2.1,32	50,0.009	15	8	214	245	26	114	221
	Lab-2:0.2,1.7,32	50,0.009							
	Lab-3:0.9,2.1,22	50,0.009							
	Lab-4:0.9,2.1,22	50,0.009							
	FA-H1: 0,1,64	50,0.009							
	FA-M1: 0,1,32	50,0.009							
	FA-L1: 0,1,16	50,0.009							
E2,S1,FE3	Lab-1: 0.9,2.1,32	140-722	13	7	206	240	24	121	158
	Lab-2: 0.2,1.7,32	300-2300							
	Lab3: 0.9,2.1,22	90-618							

	Lab-4: 0.9,2.1,22 FA-H1: 0,1,64 FA-M1: 0,1,32 FA-L1: 0,1,16	85-579 0.2e3-43006 550-21978 20-21978 10-31301						
E1,S1,FE1	Lab-1:0.9,2.1,22 Lab-2:0.2,1.7,32 Lab-3:0.9,2.1,22 Lab-4:0.9,2.1,22 FA-H1: 0,1,64 FA-M1: 0,1,32 FA-L1: 0,1,16	$\sigma <= 5$ $\sigma <= 4$ $\sigma <= 90$ $\sigma <= 90$ $\sigma <= 400$ $\sigma <= 600$ $\sigma <= 800$	16	11	566	566	39	134	171
E1,S1,FE2	Lab-1:0.9,2.1,32 Lab-2:0.9,2.1,32 Lab-3:0.9,2.1,22 Lab-4:0.9,2.1,22 FA-H1: 0,1,64 FA-M1: 0,1,32 FA-L1: 0,1,16	50,0.009 50,0.009 50,0.009 50,0.009 50,0.009 50,0.009 50,0.009	14	11	488	474	44	138	138
E1,S1,FE3	Lab-1:0.9,2.1,32 Lab-2:0.9,2.1,32 Lab-3:0.9,2.1,22 Lab-4:0.9,2.1,22	100-4944 50-2703 62-1962 64-1911	5	5	566	565	38	133	170

	FA-H1: 0,1,64	150-8234							
	FA-M1: 0,1,32	71-1501							
	FA-L1: 0,1,16	54-1353							
E4,S1,FE1	Lab-1: 0.9,2.1,64,2,0.1 Lab-2: 0.9,2.1,64,2,0.1 Lab-3: 0.9,2.1,64,2,0.1 Lab-4: 0.9,2.1,64,2,0.1 FA-H1: 0.9,2.1,64,16,0.1 FA-M1: 0.9,2.1,32,4,0.1 FA-L1: 0.9,2.1,16,8,0.1	$\sigma <= 1$ $\sigma <= 1$ $\sigma <= 1$ $\sigma <= 1$ $\sigma <= 1$ $\sigma <= 1$	12	9	3164	3164	26	97	36
E4,S1,FE2	Lab-1: 0.9,2.1,64,2,0.1 Lab-2: 0.9,2.1,64,2,0.1 Lab-3: 0.9,2.1,64,2,0.1 Lab-4: 0.9,2.1,64,2,0.1 FA-H1: 0.9,2.1,64,16,0.1 FA-M1:	50,0.009 50,0.009 50,0.009 50,0.009 50,0.009 50,0.009 50,0.009	1358	214	2913	2913	19	96	21

	0.9,2.1,32,4,0.1 FA-L1: 0.9,2.1,16,8,0.1								
E4,S1,FE3	Lab-1: 0.9,2.1,64,2,0.1 Lab-2: 0.9,2.1,64,2,0.1 Lab-3: 0.9,2.1,64,2,0.1 Lab-4: 0.9,2.1,64,2,0.1 FA-H1: 0.9,2.1,64,16,0.1 FA-M1: 0.9,2.1,32,4,0.1 FA-L1: 0.9,2.1,16,8,0.1	4-415 6-185 62-1962 64-1911 150-8234 71-1501 54-1353	1204	369	2818	2818	25	96	19

Table 3. 5: Summary of sensitivity analysis detected RCF damage to feature extraction models (F1-F3) including the influence of the different enhancement methods.

The feature extraction models differ in terms of characterisation of RCF defects in the following detailed manner. In the case of SDHSD method (FE1) the estimated standard deviation of the shape histogram (σ) varies depending on the enhancement methods adopted. Weng, W. & Chen, H. (2015) suggests $\sigma \leq 1$. However, from Table 3.5 above, this assumption is only applicable in the case of FFT enhanced images.

The performance of texture based feature extraction method (denoted by FE2 in Table 3.5) is mainly dependent on the appropriate choice of Gabor filter frequency and orientation corresponding to the maximum energy dissipated with the region of interest. For this method it is worth mentioning that regardless of the input image, a constant frequency of 50Hz and energy threshold of 0.009 is satisfactorily adopted for all samples. This observation has encouraged favourable application of the method for automatic RCF detection in image processing algorithms. As discussed in the works of Li, Q., & Ren, S. (2012) applying morphological criteria (denoted by FE3 in Table 3.5) for excluding false defects requires experienced knowledge or site data information, specifically the area, length, and orientation of identified damage. In this thesis, RCF defects are assumed to have a minimum length of 1mm, and it is impossible to have perfectly vertical or horizontal orientations. For the area of defects, it is mainly dependent on the type of defect be it spall or fatigue damage in nature. In general, FE3 is concluded to be mostly influenced by the enhancement method applied and dominant defect type within the image being analysed.

As an overall assessment of the performance of FE1-FE3, the variation in number of RCF defects detected especially in the case Lab1 and Lab2 samples, is a reasonable indicator of best performance for the algorithms presented in Table 3.5. Furthermore, preference is given to the less tedious nature of optimising threshold values related to standard deviation of shape distance (σ) and defect area threshold values. A cascaded application of two or more of the investigated enhancement and feature extraction methods will be explored in proposing an automatic RCF detection algorithm in later section of the thesis (see section 3.5). In the case of sample Lab-1 and Lab-2, the texture based method is observed to be susceptible to falsely identifying rust on the rail head as suspect RCF defects.

The Receiver Operating Characteristic (ROC) curve is presented as a method to analyse the performance of extracted rail damage by comparing the ground truth image with the feature extraction response traced image (for Lab-1 image with known expected crack population of 6) as shown in Figure 3.41 below.

Ground truth image **Feature extraction response** **Traced image**
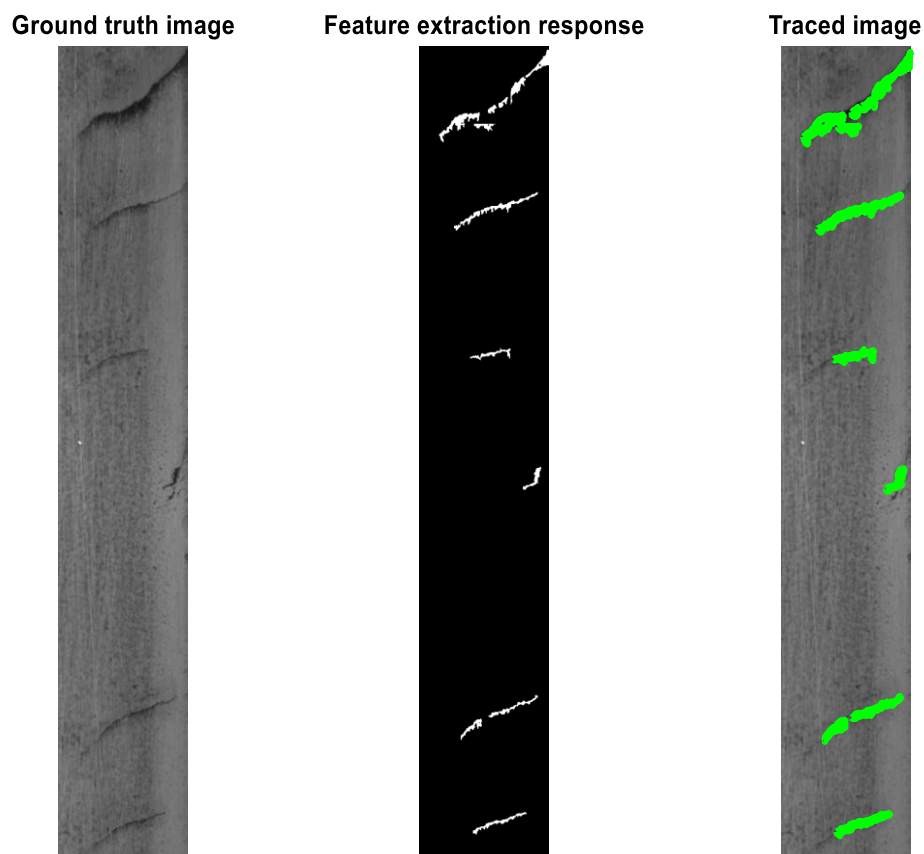
Figure 3.41: Shows the Original image (left), extracted feature (centre) and traced image of Lab-1 image.

From the plot of sensitivity against (1-specificity) in Figure 3.43 below, results obtained for changing the feature extraction threshold applied on Lab-1 indicates that the probability of incorrectly detecting a defect is less than 2% when its true detection state is negative and the probability of correctly detecting rail damage is approximately 98%.

ROC curve for Lab-1
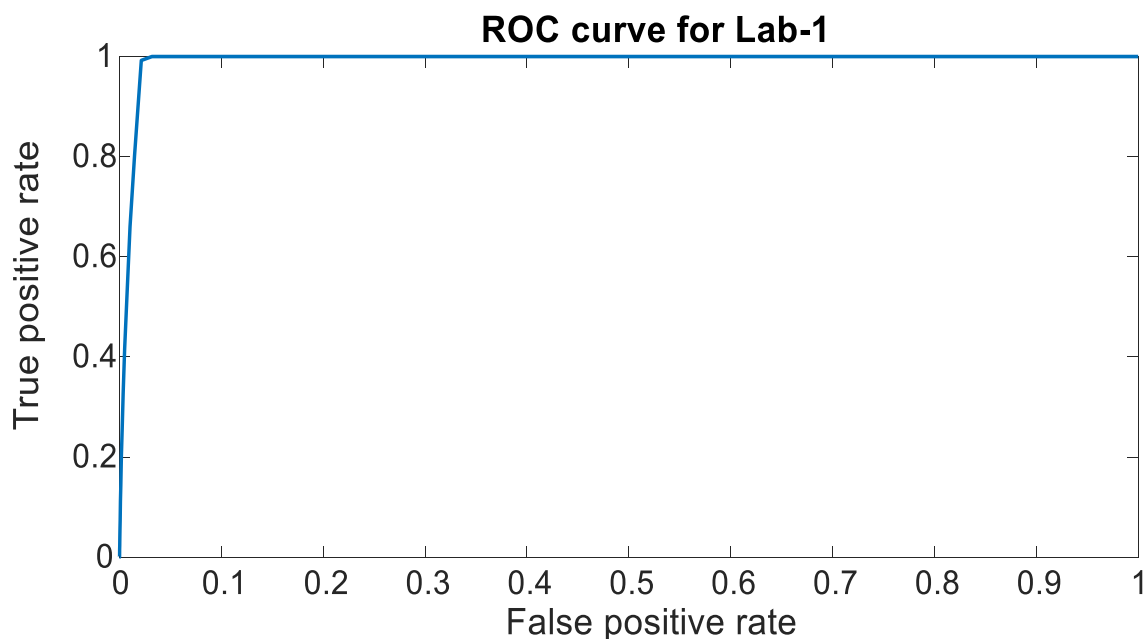
Figure 3.42: Shows the probabilities of true and false detection for Lab-1 image.

The sensitivity/specificity plot shows the true positive rate and true negative rate probabilities defined over a 100 sample points between minimum and maximum pixel intensity value of the feature extraction response image. The optimal threshold identified from the crossover in Figure 3.43 is approximately equal to -5.208.

Sensitivity/Specificity decision plot
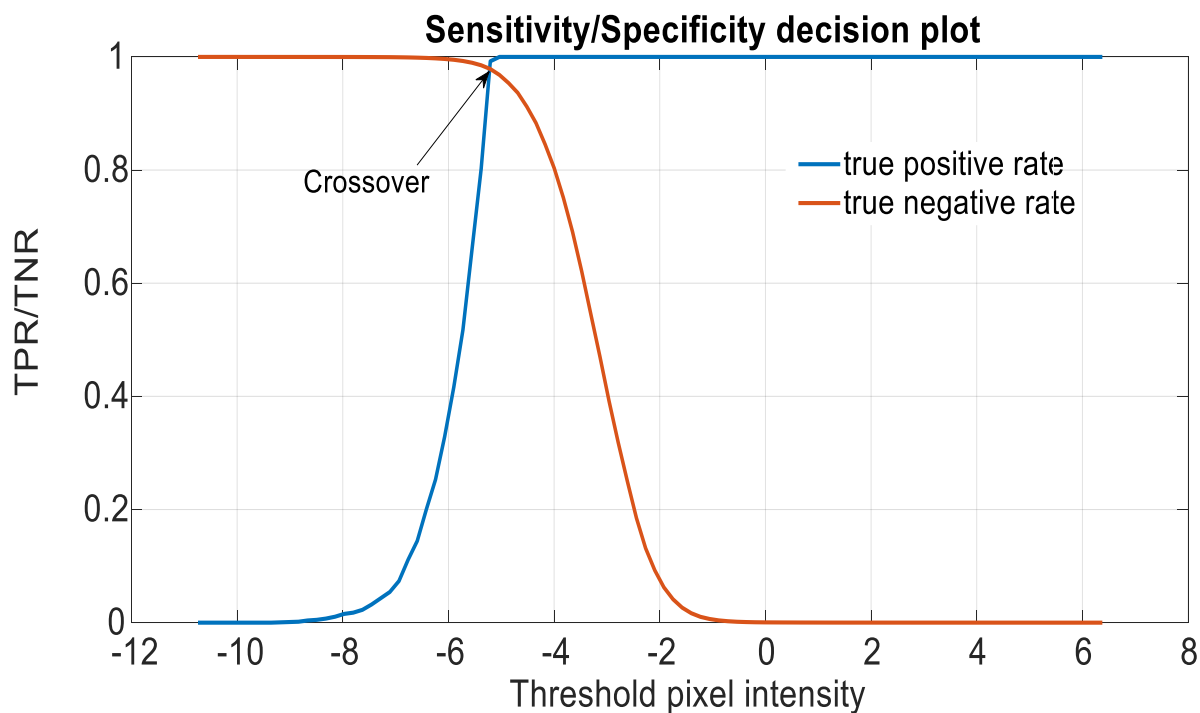
Figure 3. 43: Shows the cross over between the sensitivity and specificity for Lab-1 image.

In the case of Lab-2 depicted in Figure 3.44 below, the performance of extracted rail damage by comparing the ground truth image with the feature extraction response showing the traced image for Lab-2 (with known expected crack population of 5).

Original image

Feature extraction response
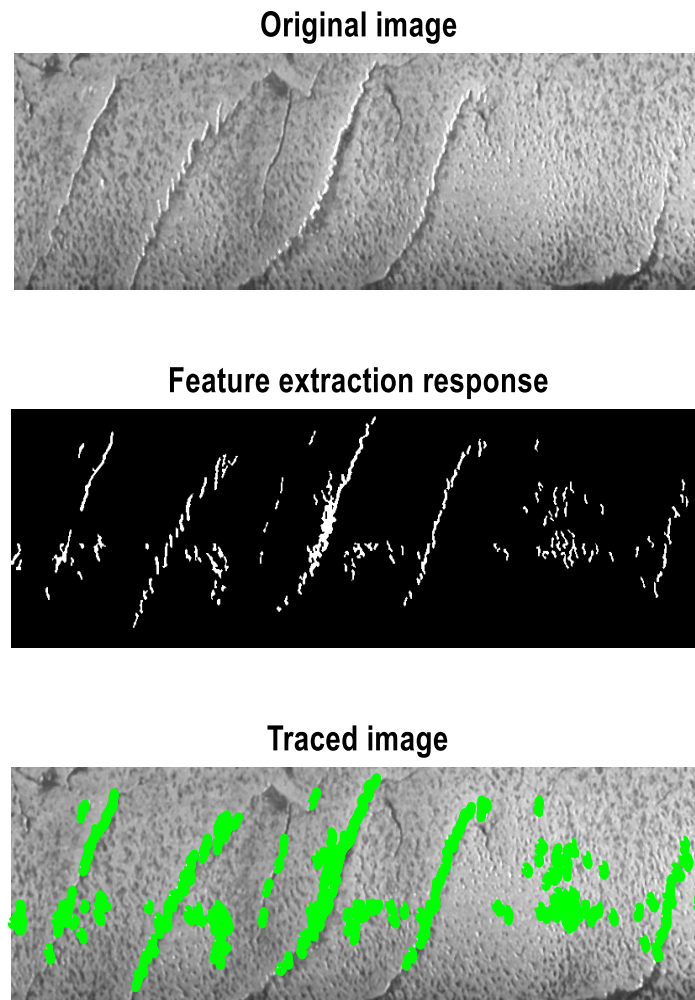
Traced image

Figure 3. 44: Shows the Original image (left), extracted feature (centre) and traced image of Lab-2 image.

From the plot of sensitivity against (1-specificity) in Figure 3.45 below, results obtained for changing the applied feature extraction threshold on Lab-2 indicates that the probability of incorrectly detecting a defect is less than 18% when its true detection state is negative, and the probability of correctly detecting rail damage is approximately 72%.
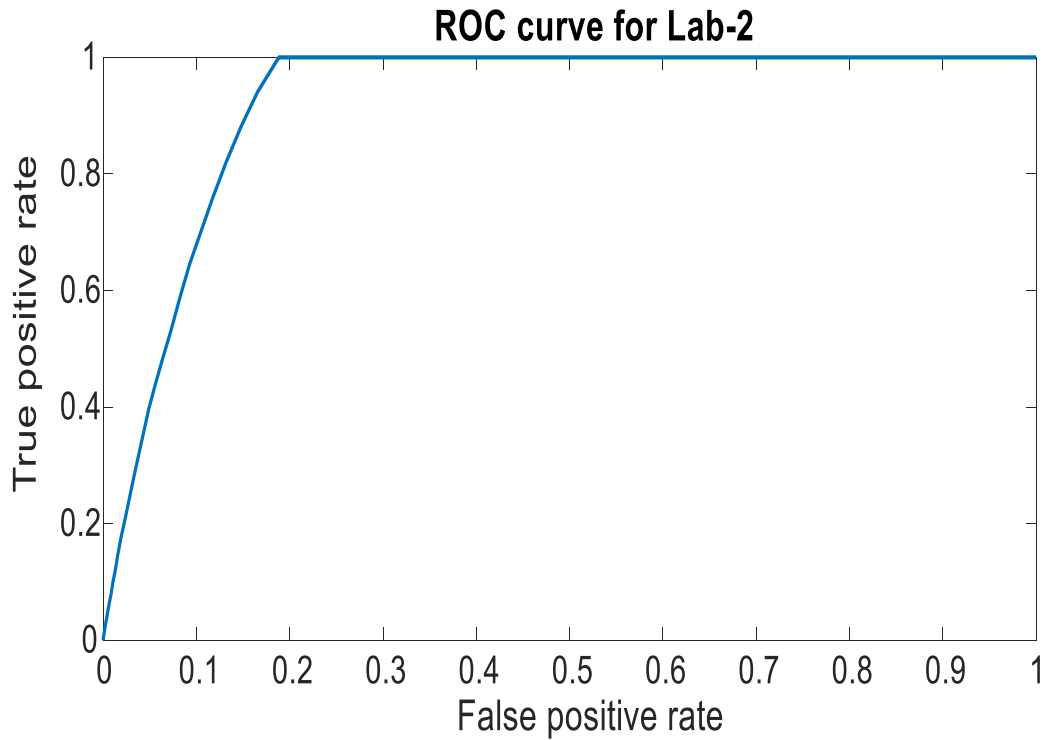
Figure 3. 45: Shows the probabilities of true and false detection for Lab-2 image.

The sensitivity/specificity plot confirms that these probabilities are defined over a threshold ranging from minimum to maximum pixel intensity value of the ground truth image and the optimal threshold identified from the crossover in Figure 3.46 is approximately equal to -1.725
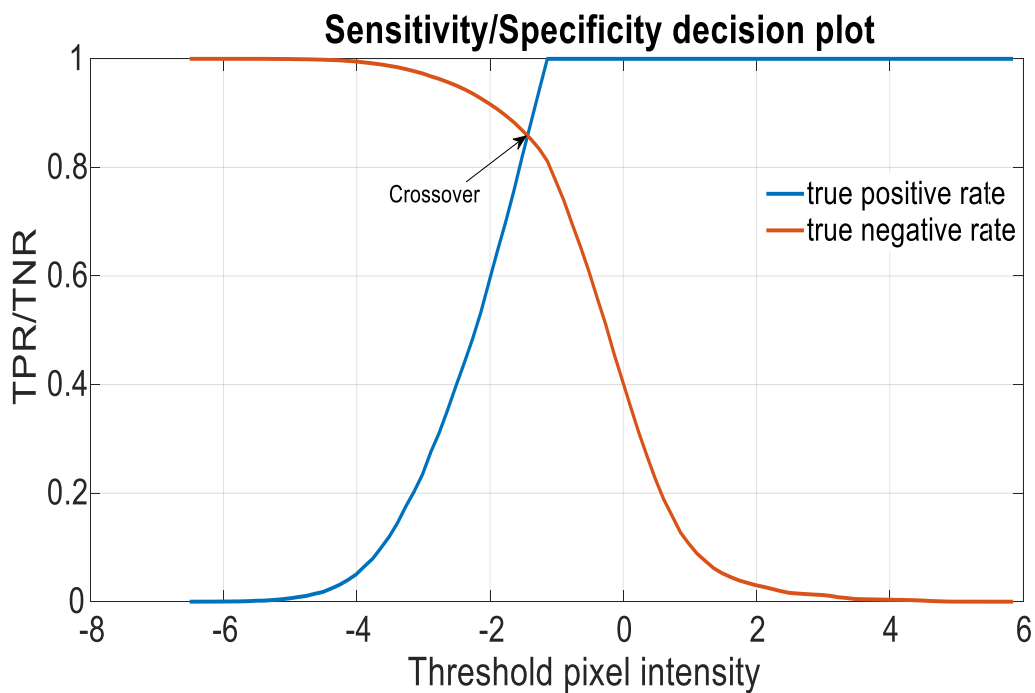
Figure 3.46: Shows the cross over between the sensitivity and specificity for Lab-2 image.

The ROC curve analysis is presented for FA-H1 image with multiple expected crack population as shown in the Figure 3.47 below.

Ground truth image **Feature extraction response** **Traced image**
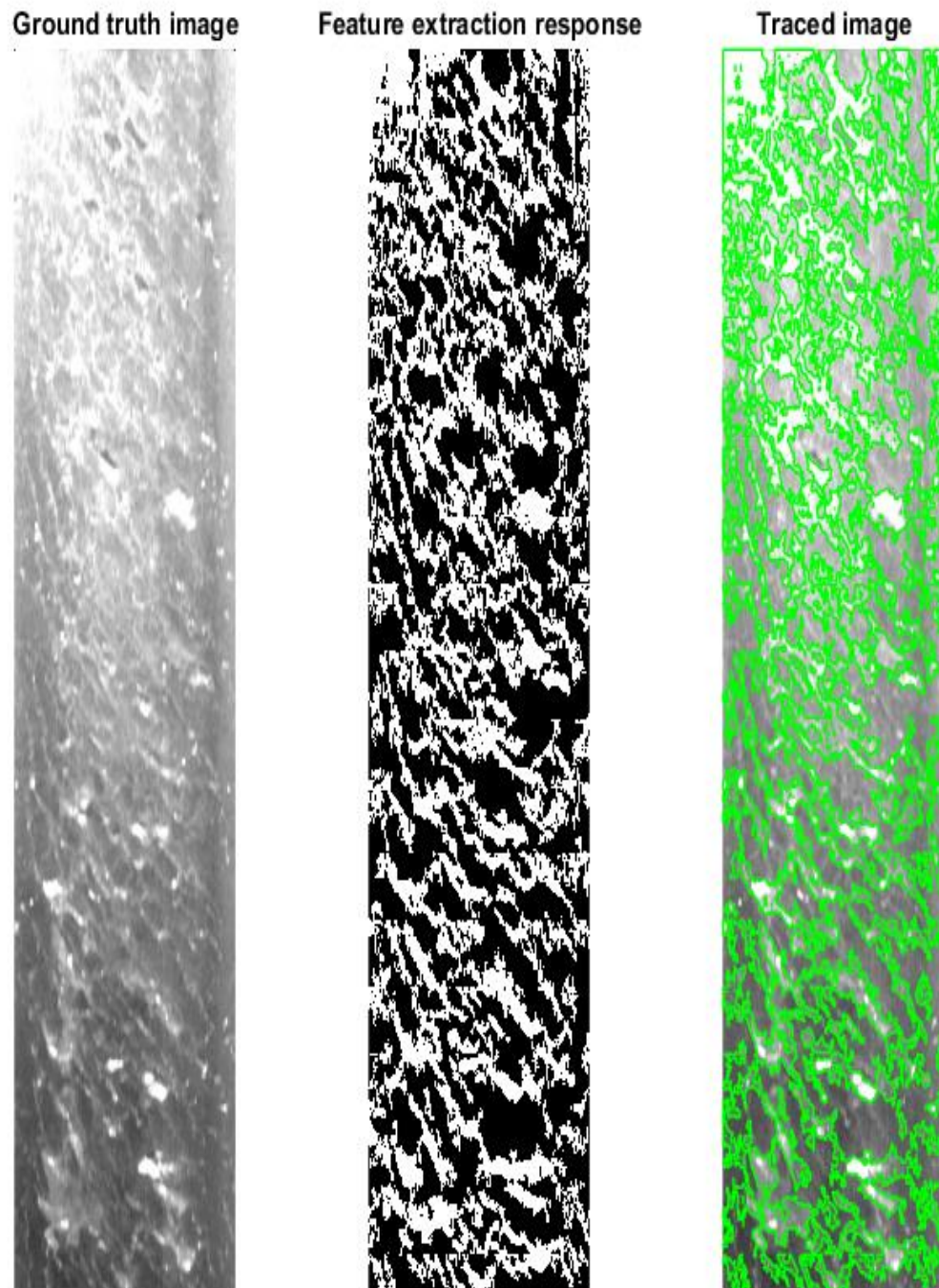
Figure 3.47: Shows the Original image (left), extracted feature (centre) and traced image of FA-H1 image.

From the plot of sensitivity against (1-specificity) in Figure 3.48 below, results obtained for changing the applied feature extraction threshold on FA-H1 indicates that the probability of incorrectly detecting a defect is less than 12% when its true detection state is negative, the probability of correctly detecting rail damage is approximately 88%.
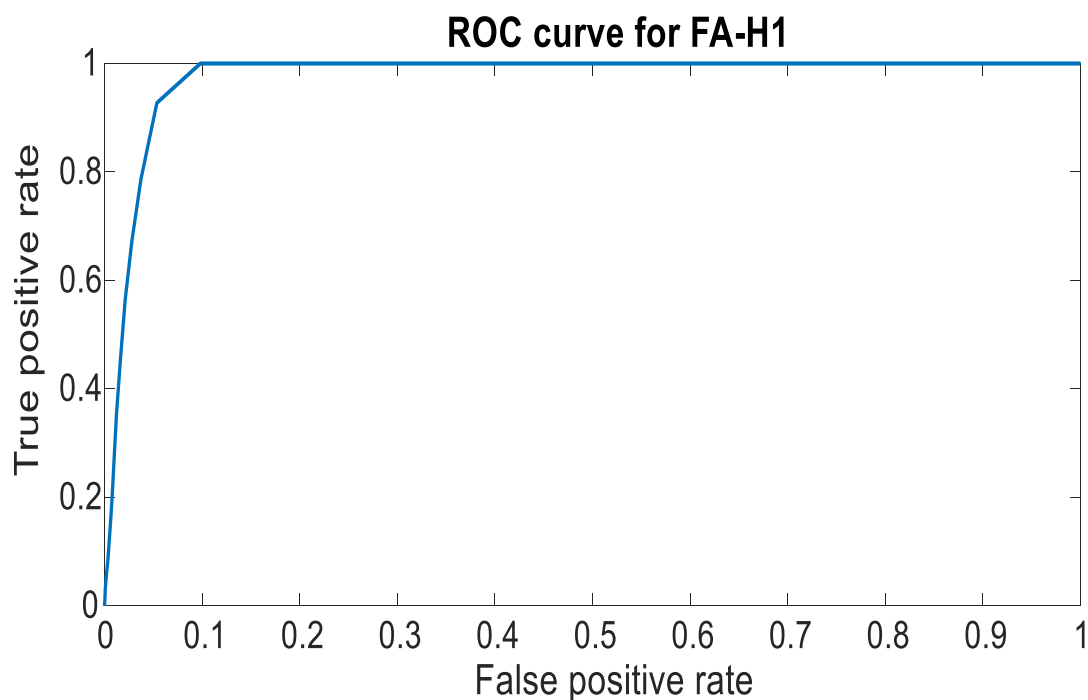
Figure 3.48: Shows the probabilities of true and false detection for FA-H1 image.

The sensitivity/specificity plot confirms that these probabilities are defined over a threshold ranging from minimum to maximum pixel intensity value of the ground truth image and the optimal threshold identified from the crossover in Figure 3.49 is approximately equal to -1.208
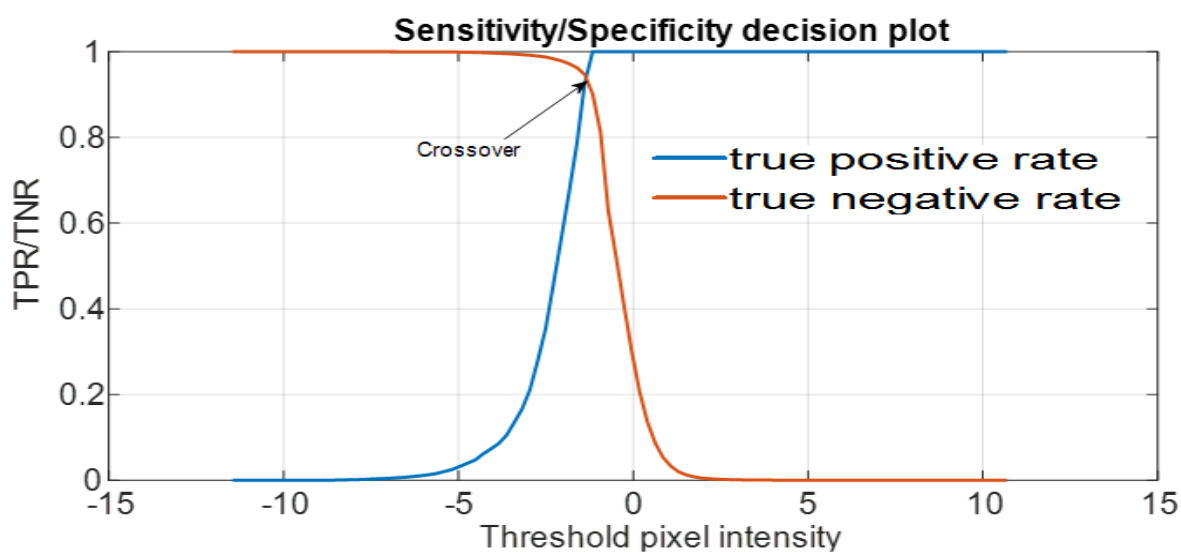
Figure 3. 49: Shows the probabilities of true and false detection for FA-H1 image.

In the case of FA-M1, the performance of extracted rail damage by comparing the ground truth image with the feature extraction response showing the traced image for FA-M1 (with multiple expected crack population) as shown in Figure 3.50 below.

Ground truth image   **Feature extraction response**   **Traced image**
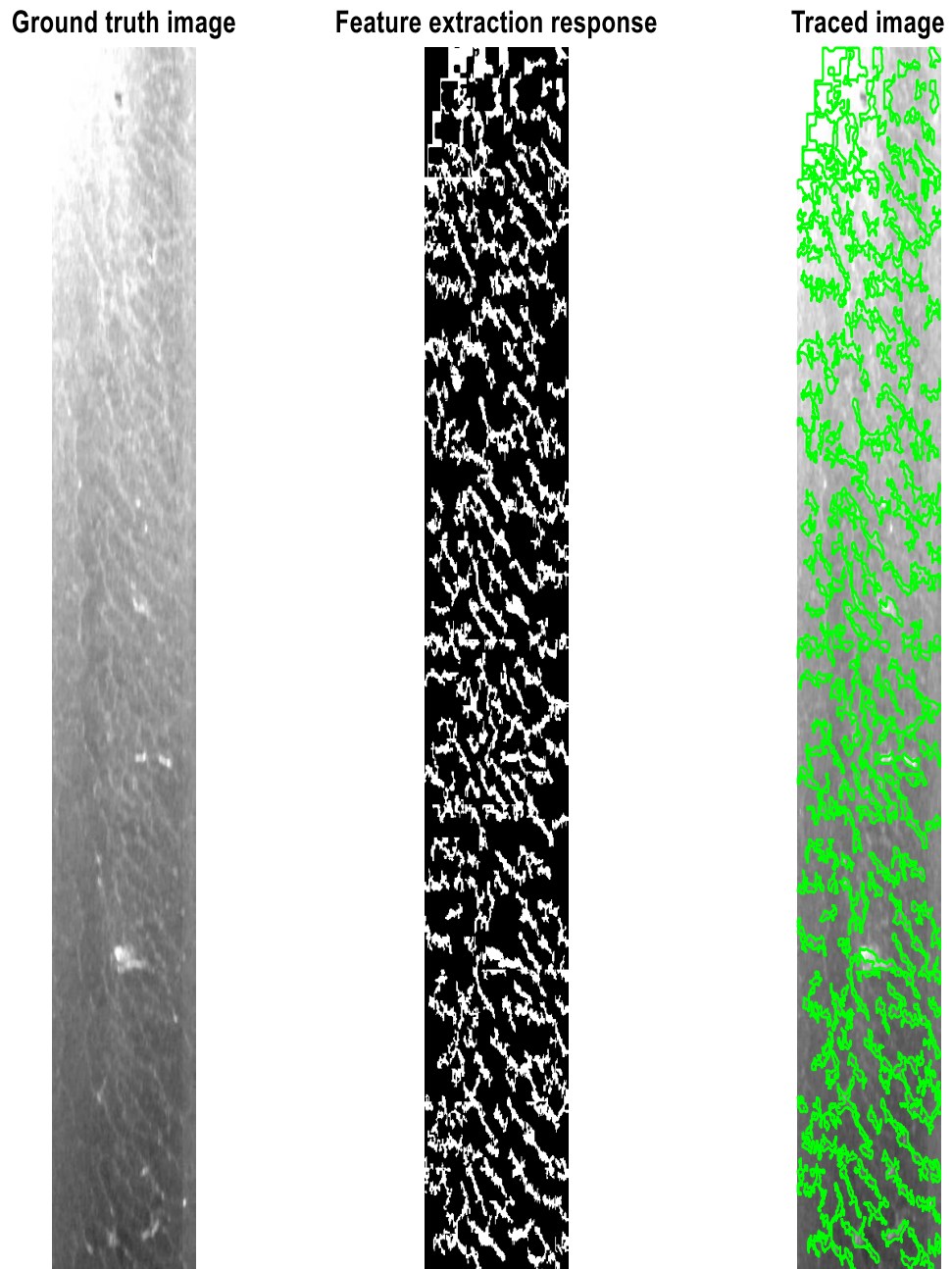
Figure 3. 50: Shows the Original image (left), extracted feature (centre) and traced image of FA-M1 image.

From the plot of sensitivity against (1-specificity) in Figure 3.51 below, results obtained for changing the applied feature extraction threshold on FA-M1 indicates that the probability of incorrectly detecting a defect is less than 8% when its true detection state is negative, the probability of correctly detecting rail damage is approximately 92%.

ROC curve for FA-M1
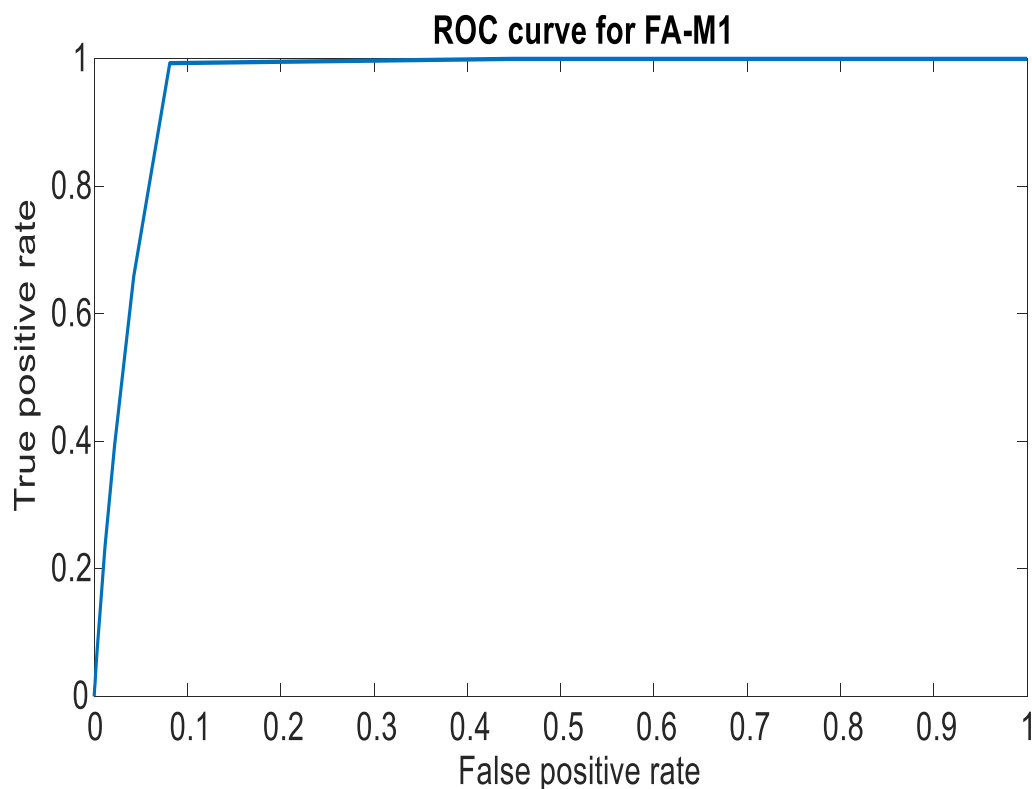
Figure 3. 51: Shows the probabilities of true and false detection for FA-M1 image.

The sensitivity/specificity plot confirms that these probabilities are defined over a threshold ranging from minimum to maximum pixel intensity value of the ground truth image and the optimal threshold identified from the crossover in Figure 3.52 is approximately equal to -32.300
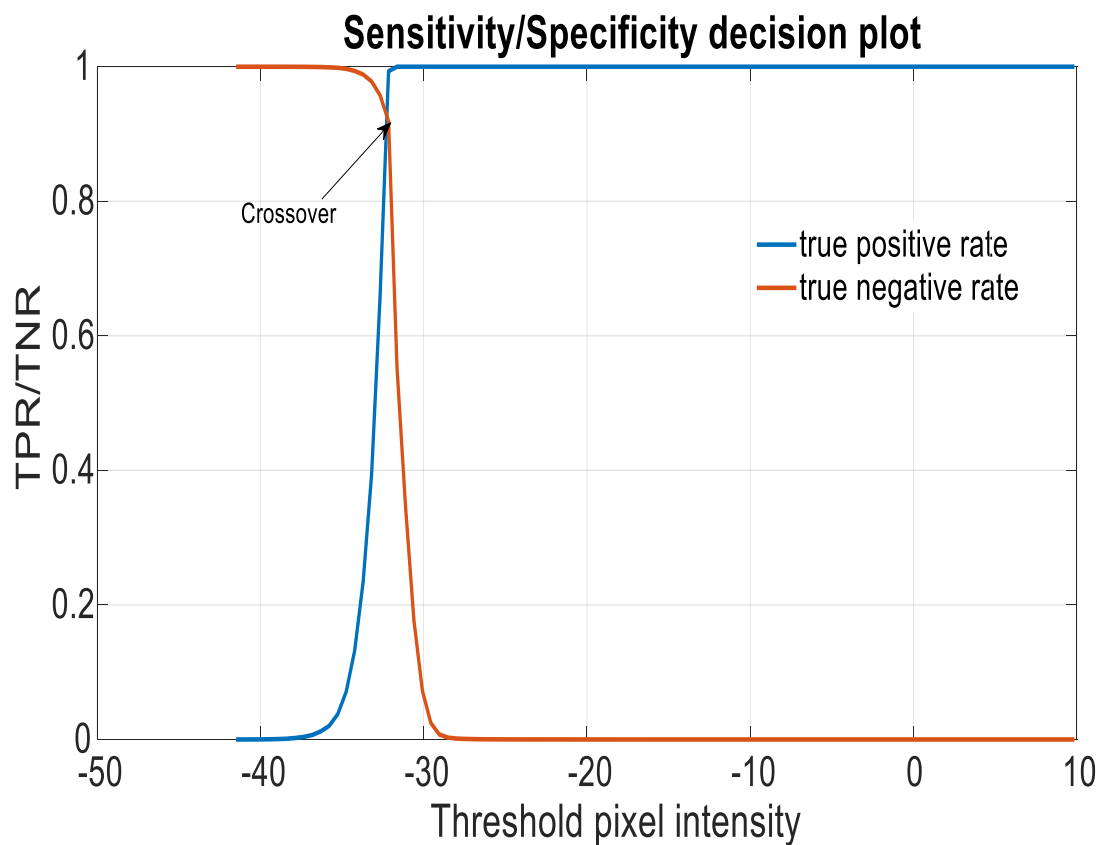
Figure 3. 52: Shows the probabilities of true and false detection for FA-M1 image.

For FA-L1, the performance of extracted rail damage by comparing the ground truth image with the feature extraction response showing the traced image for FA-L1 (with multiple expected crack population) as shown in Figure 3.53.

Ground truth image   **Feature extraction response**   **Traced image**
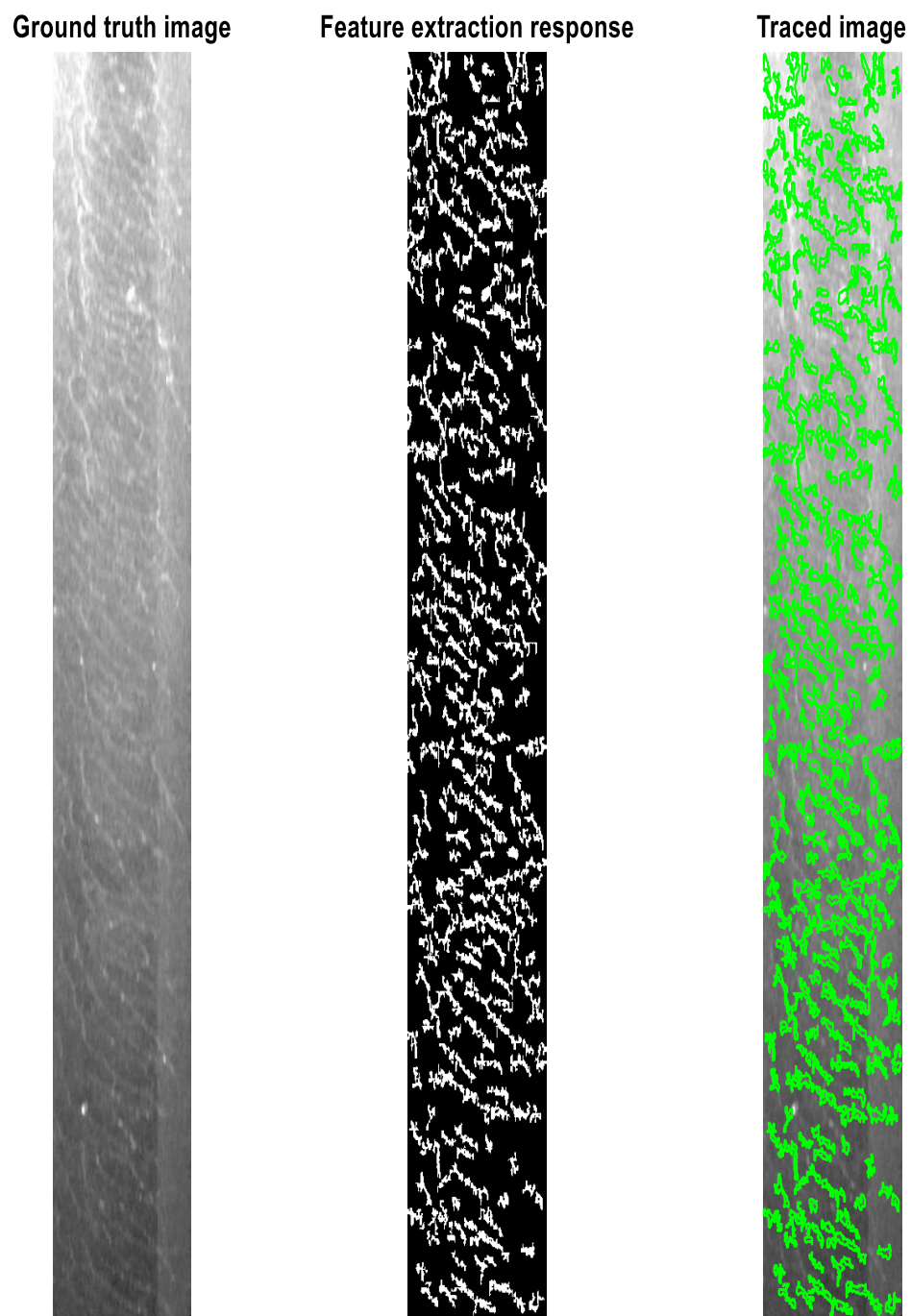
Figure 3.53: Shows the Original image (left), extracted feature (centre) and traced image of FA-L1 image.

From the plot of sensitivity against (1-specificity) in Figure 3.54 below, results obtained for

changing the applied feature extraction threshold on FA-L1 indicates that the probability of

incorrectly detecting a defect is less than 18% when its true detection state is negative, the probability of correctly detecting rail damage is approximately 82%.

ROC curve for FA-L1
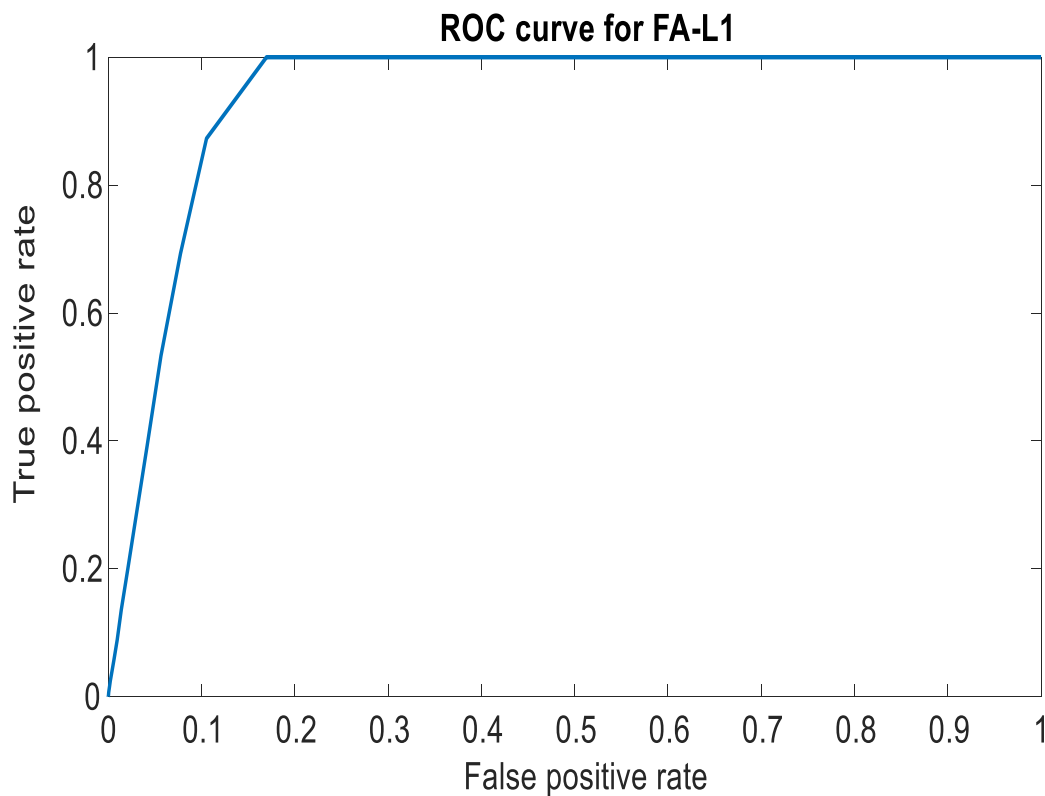
Figure 3. 54: Shows the probabilities of true and false detection for FA-L1 image.

The sensitivity/specificity plot confirms that these probabilities are defined over a threshold ranging from minimum to maximum pixel intensity value of the ground truth image and the optimal threshold identified from the crossover in Figure 3.55 is approximately equal to -8.590
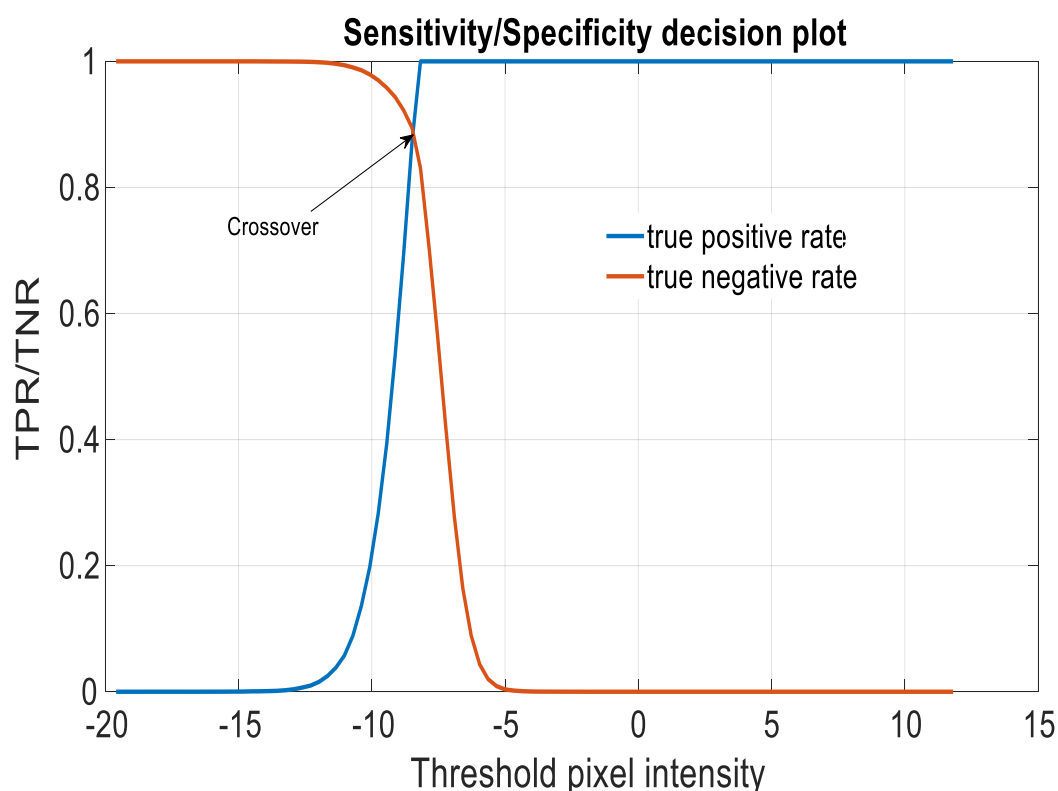
Figure 3. 24: Shows the probabilities of true and false detection for FA-L1 image.

3.4.4 Calibration

Further to successful feature extraction of RCF defects, it is essential that image calibration is performed for the purpose of generating defect statistical data that readily supports maintenance decision making. In this thesis, the calibration of rail defect images has been performed using the dimension of rail head or that of the rail foot and the performance of each of the approaches is investigated. It was observed as depicted in Figure 3.56 below, that using the dimensions of the rail head is susceptible to error due to deformation, wear, and effect of imaging on curved tracks at different curve radiuses. It is therefore considered more accurate to utilise the dimensions of the rail foot, by linear interpolation of the pixel count corresponding to the actual dimension of rail foot, rail head, and rail gauge to gauge width (150mm, 72mm, and 80mm respectively for 60E1 rail profile). A calibration factor of 0.162 has been established contrary to a value of 0.3 and 0.27 as suggested by using the dimension of rail head width and gauge to gauge width respectively.
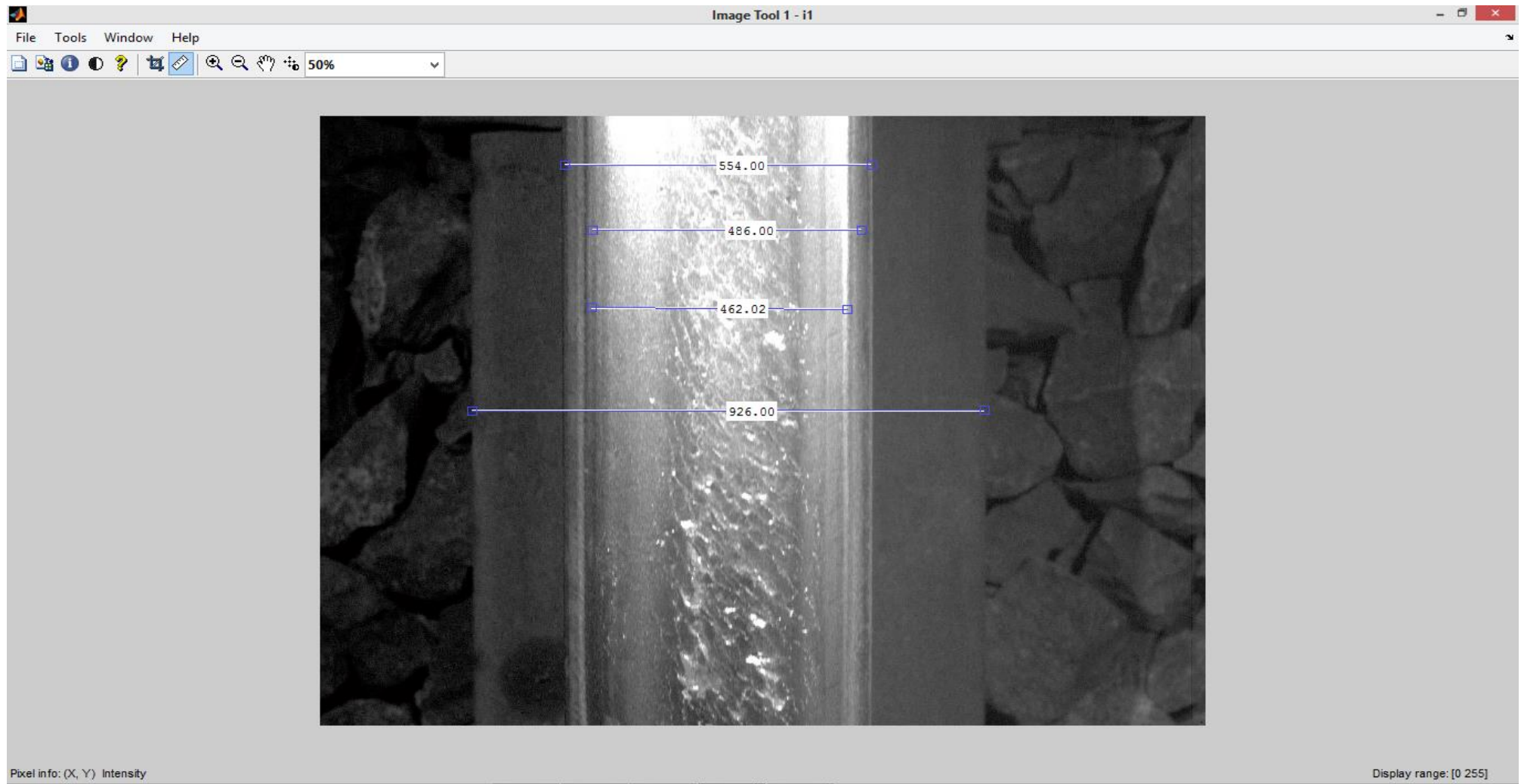
Figure 3. 5625: Shows the rail head and rail foot methods of calibrating rail damage images.

3.4.5 Generating defect statistical

After successful calibration of the defect images, MATLAB toolbox functions and the mathematical models discussed in section 3.3 were used to estimate the maximum length, and orientation of detected RCF defects. Figure 3.57 below shows the comparison of generated statistical data related to FA-H1, FA-M1, and FA-L1. These results have been computed for the most prominent enhancement methods, i.e. LMAF (E1) and visibility measure (E2). A single segmentation algorithm (OGT segmentation method) is adopted for this section of the thesis as S1. All feature extraction methods (FE1, FE2, FE3) are explored. It was observed that so long as the influential parameters (RM and RV) are kept constant regardless of feature extraction method, then a considerable agreement between statistical data is obtained for all combinations of the algorithms.

In the case of Visibility Measure enhancement (denoted by E2 in Figure 3.57) it is observed in Figure 3.57, that for FA-H1 sample a variation in total number of defects of 4% is recorded. Considering FE1, a maximum variation of 67% in measured crack length for defect number 6 is attributed to measurement of clustered defect which cause a spike in defect length. In Figure 3.57b, FA-M1 sample is observed to vary in terms of number of detected defects by about 7.7% while measured data varies by no more than 1.5% for FE1, FE2, and FE3. In the case of FA-L1 sample, Figure 3.57c confirms a variation in total number of detected defects of about 30%, while the measured data is observed to vary by no more than 55%. In-spite of the fast computational speed of VM method (less than 1.5 seconds as shown in Table 3.8), it can be concluded that this approach detects clusters due to measured crack lengths in the range of 130-40mm.
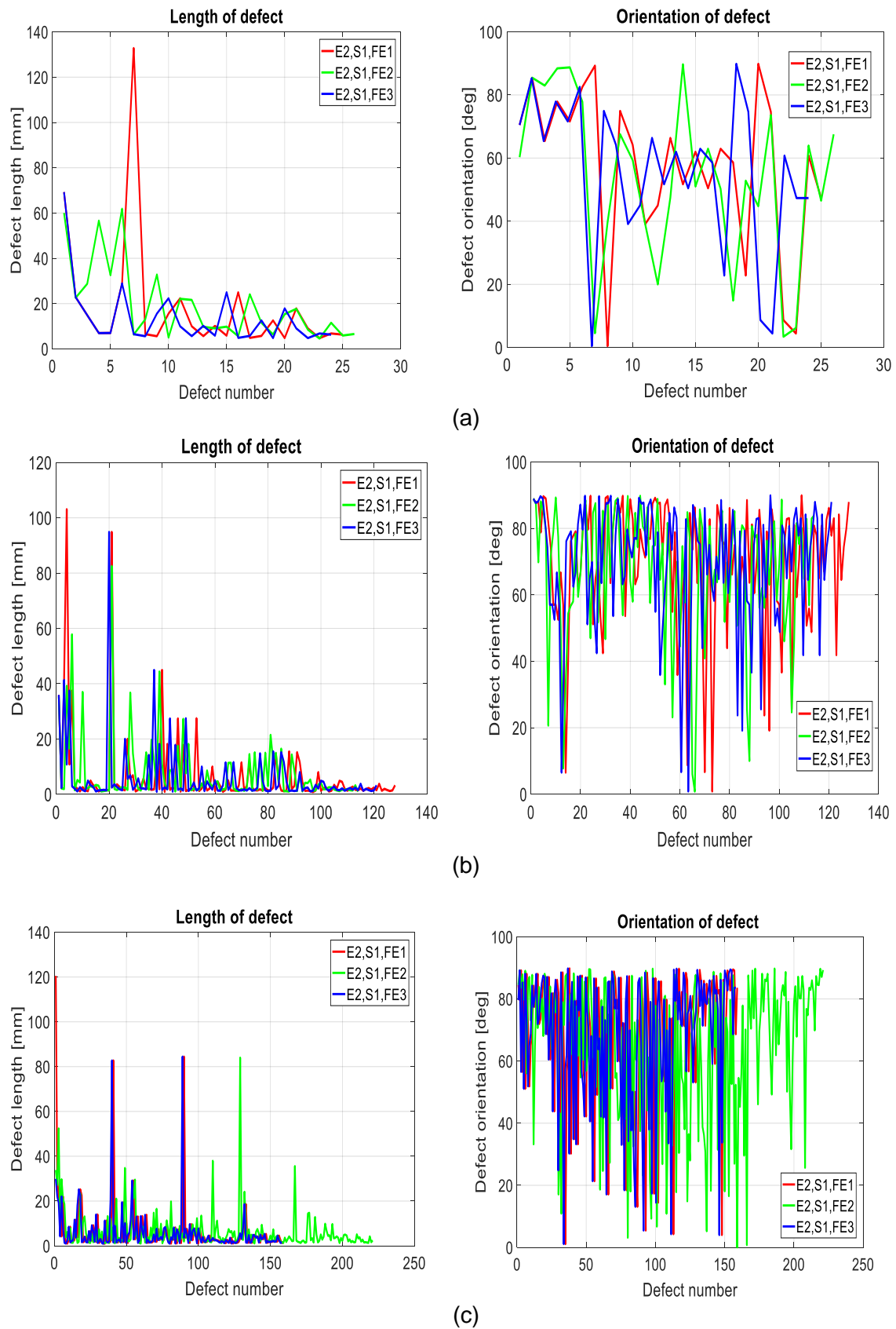
Figure 3. 57: Simulated geometry data for (a) H1, (b) M1, and (c) L1, using VM (E2) and Feature extraction (FE1-FE3).

Using LMAF enhancement (denoted by E1 in Figure 3.58) of FA-H1 field acquired image, as observed in Figure 3.58a, a variation in total number of defects of 5% is recorded, while 25% variation in measured statistical data for all feature extraction models (FE1 to FE3) is observed. In Figure 3.58b, FA-M1 sample is observed to vary by about 6.6% while measured data varies by no more than 0.1% for FE1, FE2, and FE3. In the case of FA-L1 sample, Figure 3.58c confirms a variation in total number of detected defects of about 20%, while the measured data is observed to vary by no more than 55%. The improved consistency recorded in this method is however overshadowed by the high computational time of the LMAF method (see Table 3.8). Furthermore, the measured defect length below indicates detection of clusters (because of defect length in the range of 70-40mm), which needs further pre-processing for more accurate results.

(a)

(b)

(c)
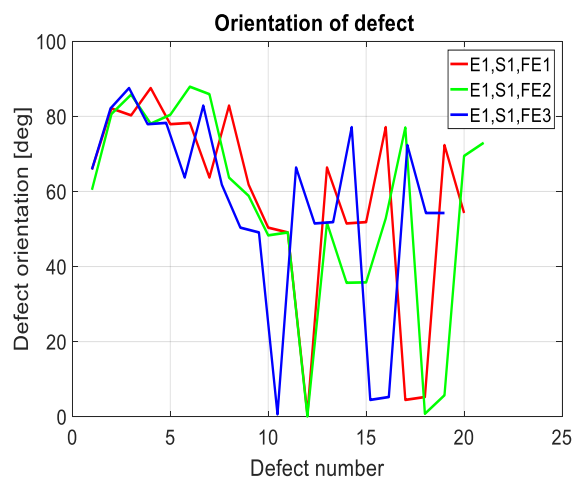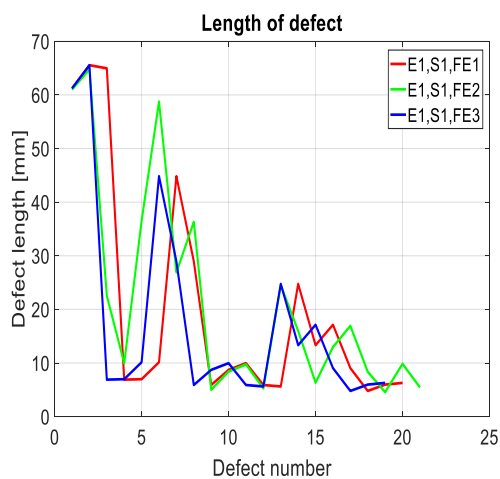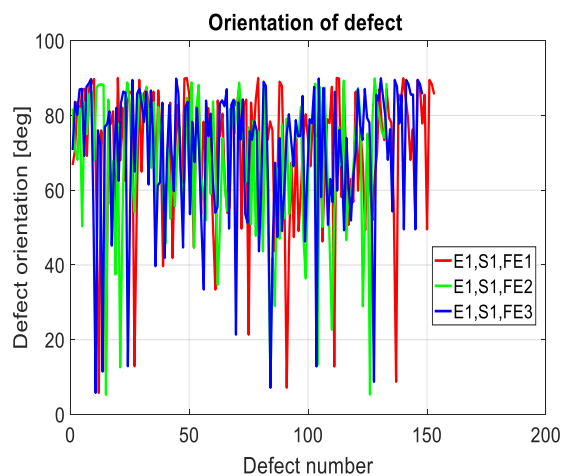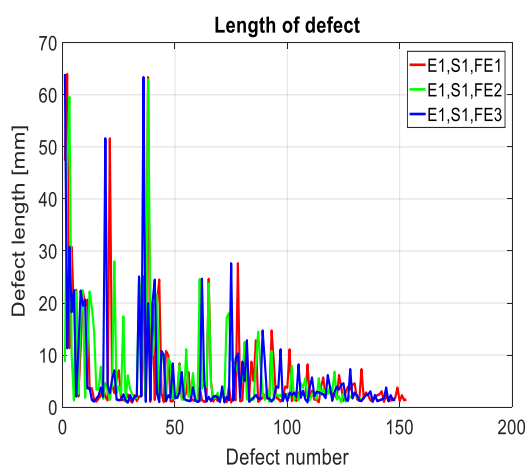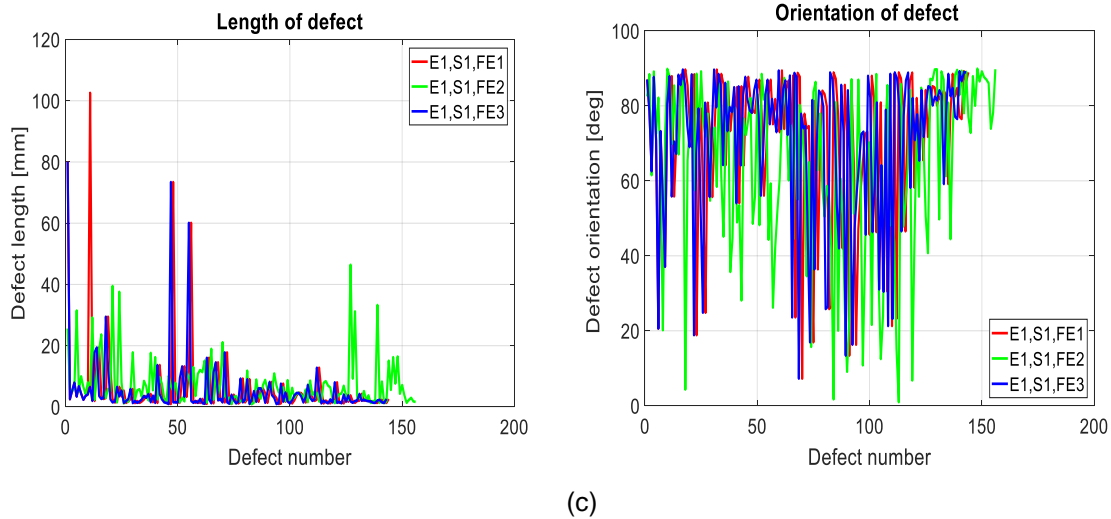
Figure 3. 58: Simulated geometry data for (a-b) H1, (c-d) M1, and (e-f) L1. Using LMAF (E1) and Feature extraction (FE1-FE3).

As a remedy to clustered defect detection, Figure 3.59 confirms that combining LN method with both VM, and LMAF (E1 and E2) and also cascading all feature extraction models (F1+F2+F3=FE*) improves agreement in the simulated results. The effect of this modification in FA-H1 as depicted in Figure 3.31a, is observed to account for 76 critically grown defects (surface length above 5mm) and only 3 newly initiated cracks (with surface length less than 5mm). In the case of sample FA-M1 in Figure 3.59b, 108 critical defects, and 27 newly initiated cracks are recorded. And in the case of FA-L1 in Figure 3.59c, a total of about 801 defects are detected all considered newly initiated (less than 5mm in surface length).

(a)
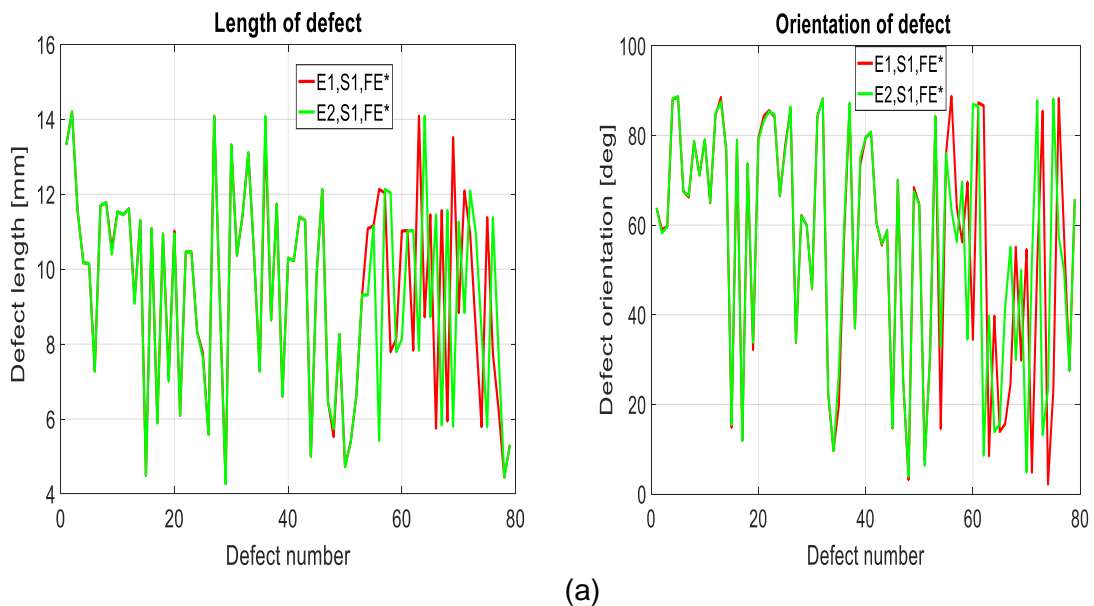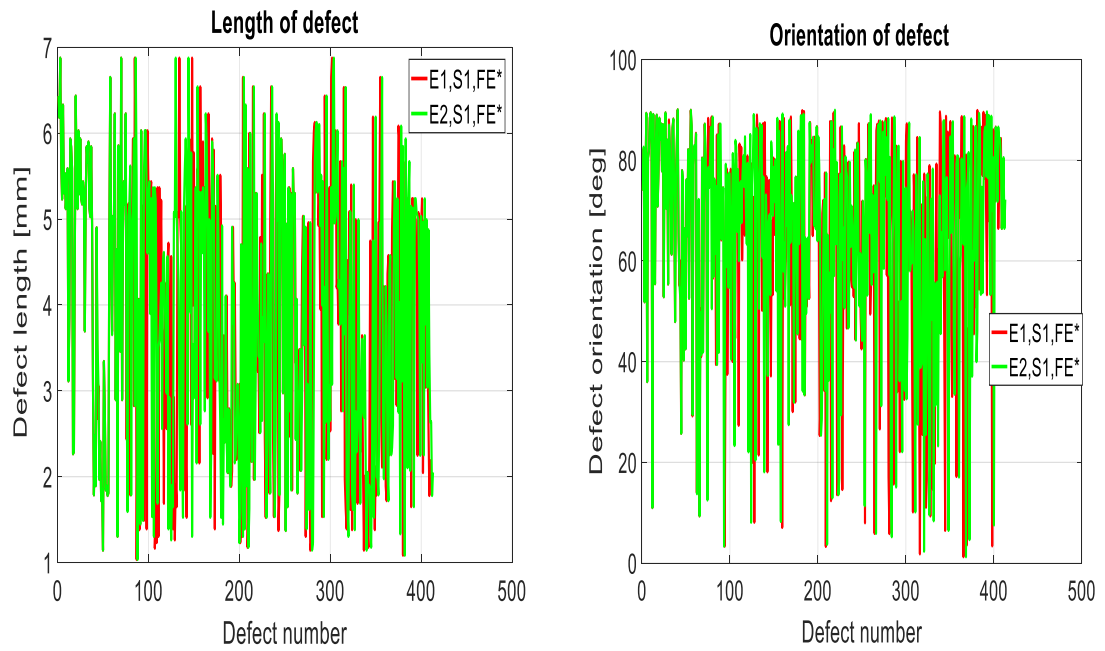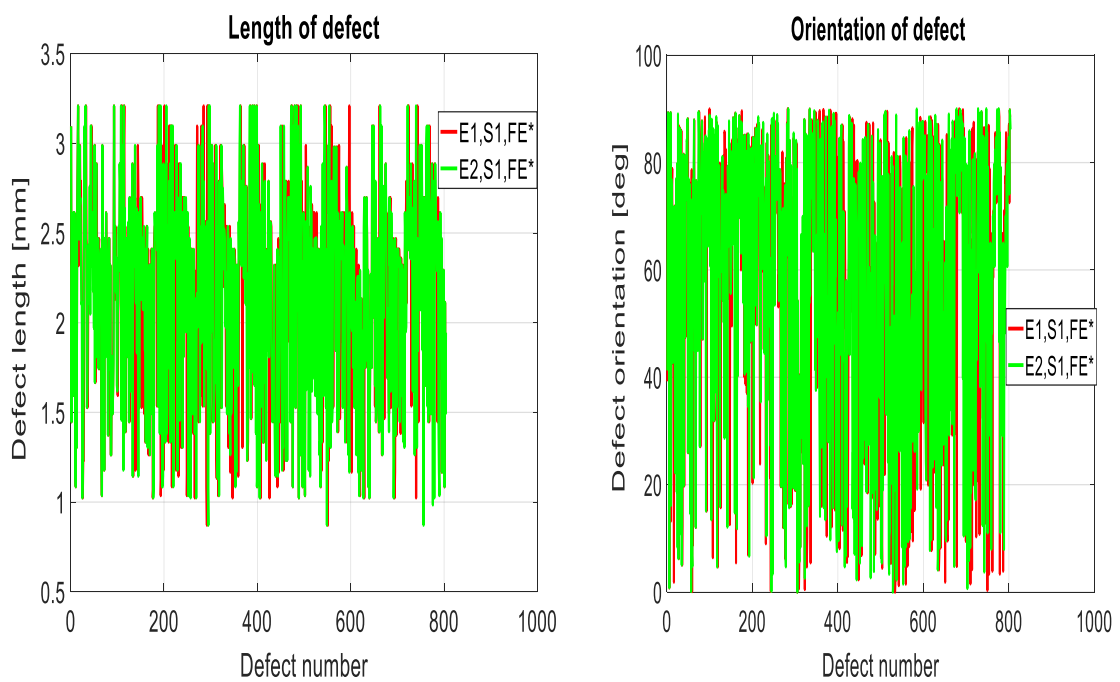
Figure 3. 59: Simulated geometry data for (a) H1, (b) M1, and (c) L1. Using (E1+LN+F1+F2+F3) and (E2+LN+F1+F2+F3).

3.5 Proposed image processing algorithm

After detailed analysis of section 3.4, this section presents a novel application of the algorithms discussed in accordance to Figure 3.32. This algorithm utilises a cascade of LN method and Visibility measure in addition to OGT segmentation and a combination of all feature extraction models. The choice of LN combined with visibility measure is supported by the maximum amplitude of PSNR and low computational time of the method (0.53 seconds as detailed in Table 3.8). Also binary segmentation by occurrence probability thresholding in the form of OGT method is used for crack segmentation, which also offers more consistent number of suspect defect regions with less computational time required of no more than 0.56 seconds (see

Table 3.8). It is generally recommended as a good practice for feature extraction, that a multiplicity of overlapping features is examined for accurate determination of RCF defects. After successful extraction of RCF damage the detected defects are further post-processed with image cleaning operations (such as fill, remove, erosion, and dilation for removal of false defects). The MATLAB '**bwmorph'** command was employed to perform morphological operations on the resulting binary output image (image of extracted RCF defect). The input arguments are '**fillgap'** (to fill small gaps in edges), '**dilate'** (to expand local dim regions), '**erode'** (to shrink local bright regions), and '**remove'** (to remove interior pixels within a suspect RCF defect). As reported in an initial publication of this PhD research (Sambo, B., Bevan, A., & Pislaru, C. 2016) the fill and remove functions are specified for filling or removing a gap as small as a single pixel along the edges or within a defect, while the erode and dilate function are set to expand or contact regions of interest by 0.9 and 0.95 respectively; which is equivalent to area opening. The cleaned image is further utilised for generation of geometrical statistical data for each RCF defect maximum length, and orientation of the defect. And the final stage of the proposed algorithm deals with calibration of geometrical data using the known rail foot dimension, as elaborated in section 3.5, a calibration factor of 0.62 is adopted in this application.
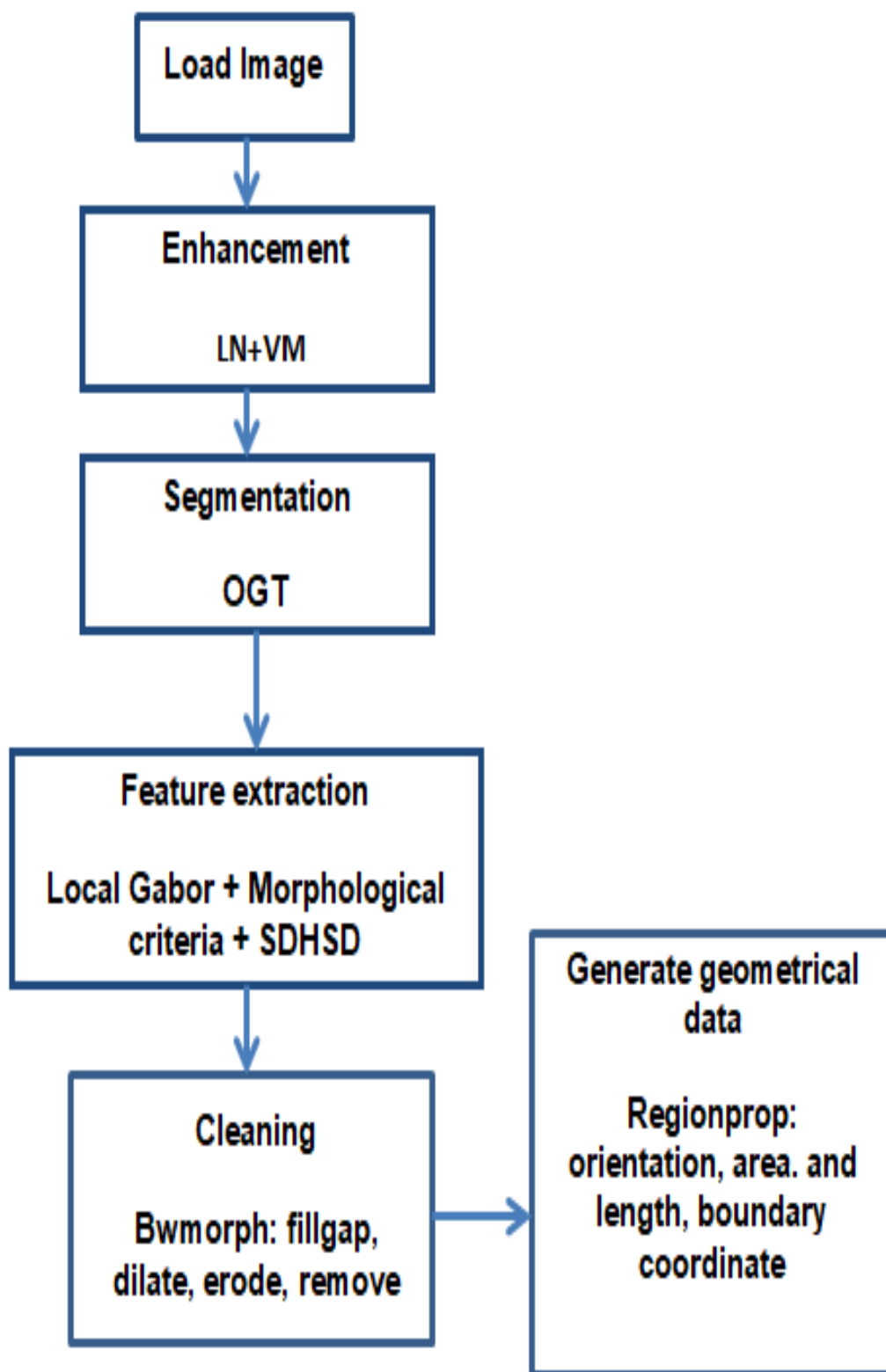
Figure 3. 60: Proposed novel application of image processing algorithm.

3.5.1 Analysis of data set 3using proposed image processing algorithm

Based on the proposition in Figure 3.60, the percentile representation of geometrical data (surface length, area, and orientation) for data set 2 suggest that 90% of detected length of heavily damaged samples FA-H1 and FA-H3 have maximum surface length less than 12mm. However, in the case of fatigue damage dominated image samples e.g. FA-H2, FA-H4, and FA-H5, maximum crack length is less than 4mm. A similar observation is made for the generated area of damage with FA-H1 and FA-H3 having 90% of RCF defects with an area less than 40mm^2 and 28mm^2 respectively, while FA-H2, FA-H4, and FA-H5 correspond to 90% of defects with a maximum area less than 5mm^2. In agreement with the understanding of crack initiation mechanism, and propagation mechanism of RCF defects, the measured orientation confirms a range of 87°-20°.

(a)

(b)

(c)
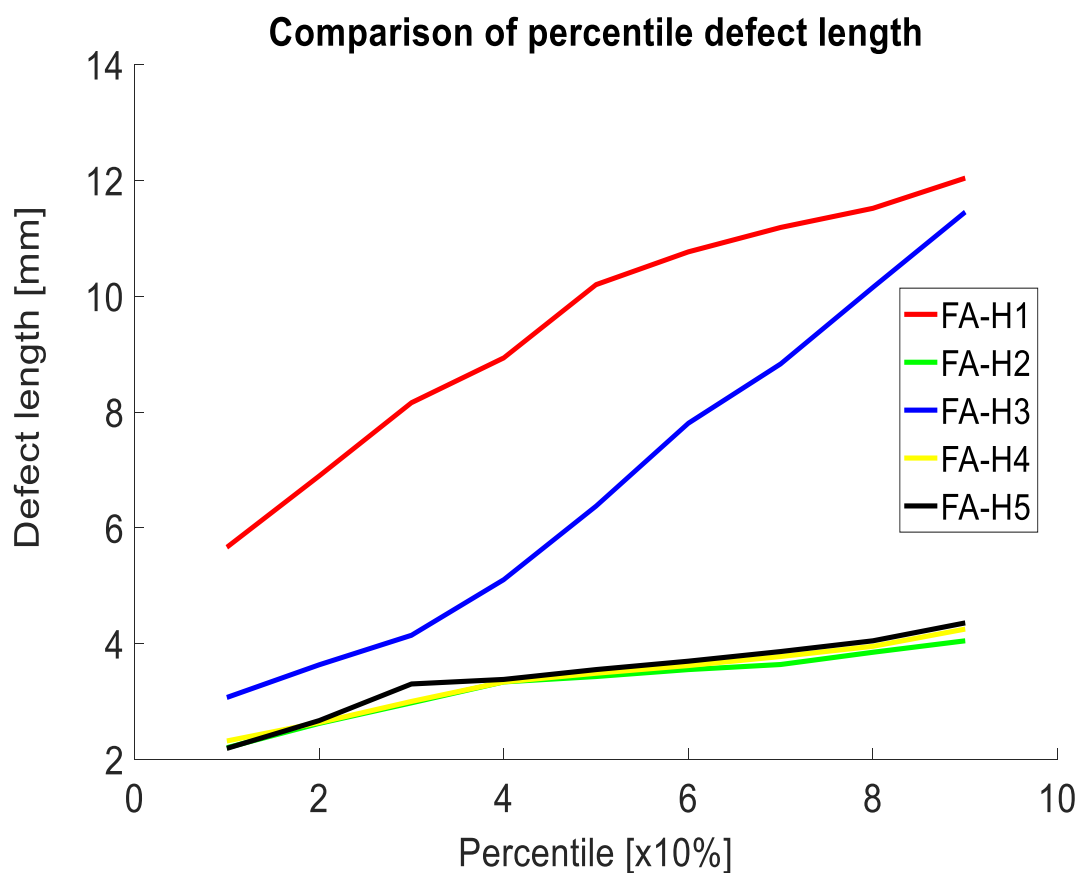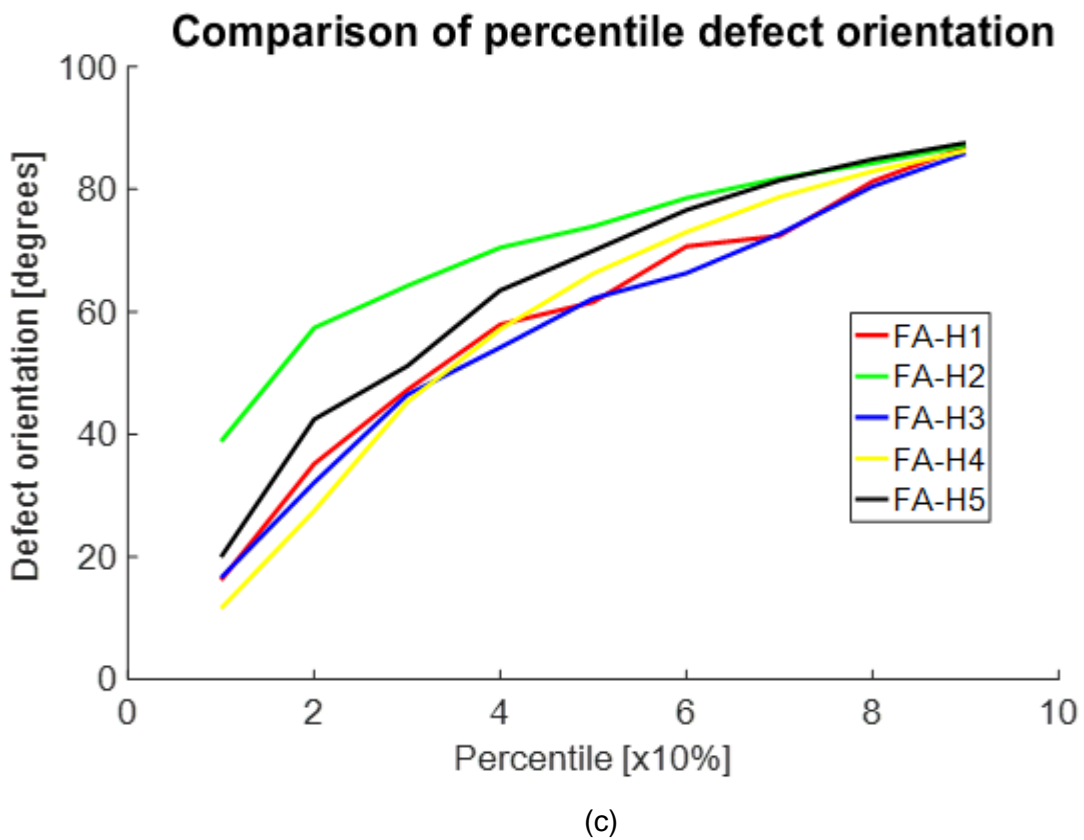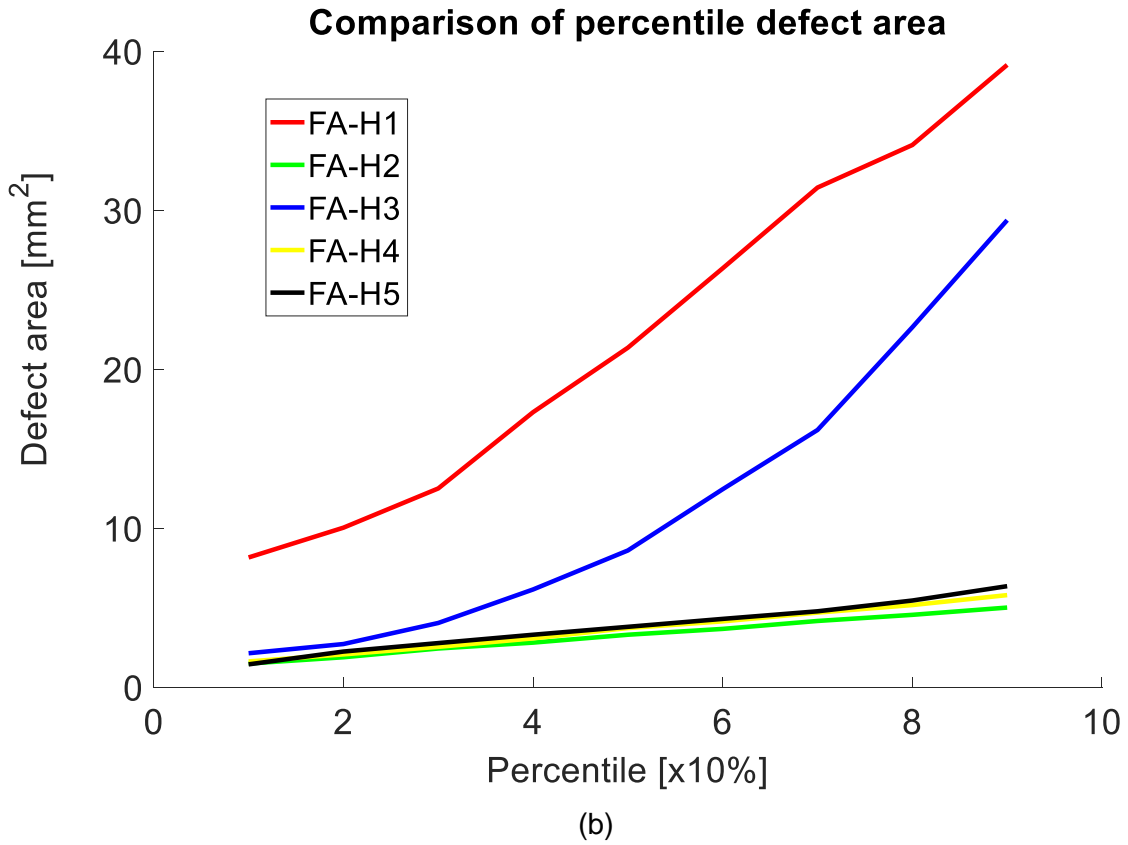
Figure 3. 61: Percentile measured length (a), area (b), and orientation (c) for heavily damaged samples of data set 2.

Analysing the case of moderately damaged field acquired samples, a slightly less than 90th percentile value of 6mm is acquired for FA-M1, FA-M2, and FA-M5 (see Figure 3.62a), while sample FA-M3 and FA-M4 are observed to have 90% of detected defects with a maximum length of no more than 3mm. A similar observation is made in the case of generated defect area and as depicted in Figure 3.62b, 90th percentile of defect area is less than 8mm^2, 8mm^2, 2mm^2, 2mm^2, and 10mm^2 for FA-M1, FA-M2, FA-M3, FA-M4, and FA-M5 respectively. This observation is in agreement with the presence of fatigue damage defects for this level of damage (moderately damaged image sample) as opposed to spall dominated images (heavily damaged samples). Furthermore, a high level of consistency in terms of orientation is in agreement with the understanding of crack initiation mechanism, and propagation mechanism of RCF defects, and the generated orientation confirms a range of 87°-20°.
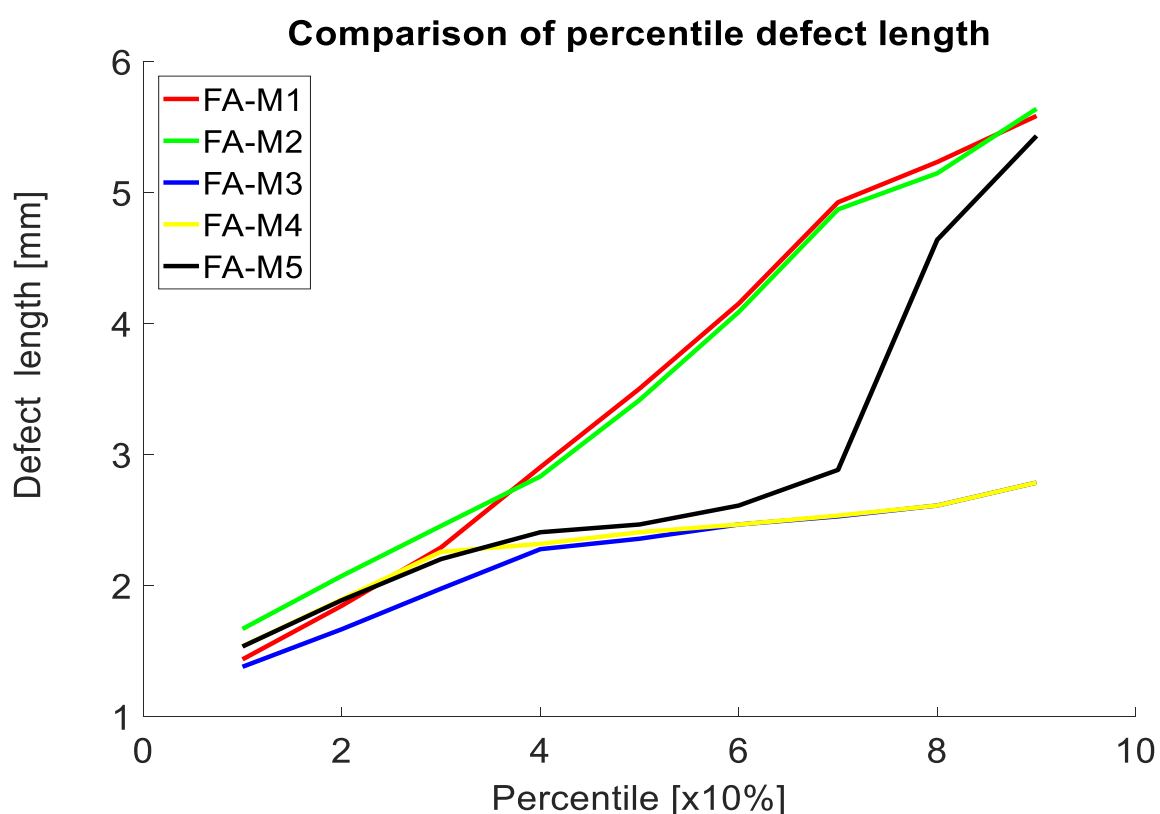
(a)
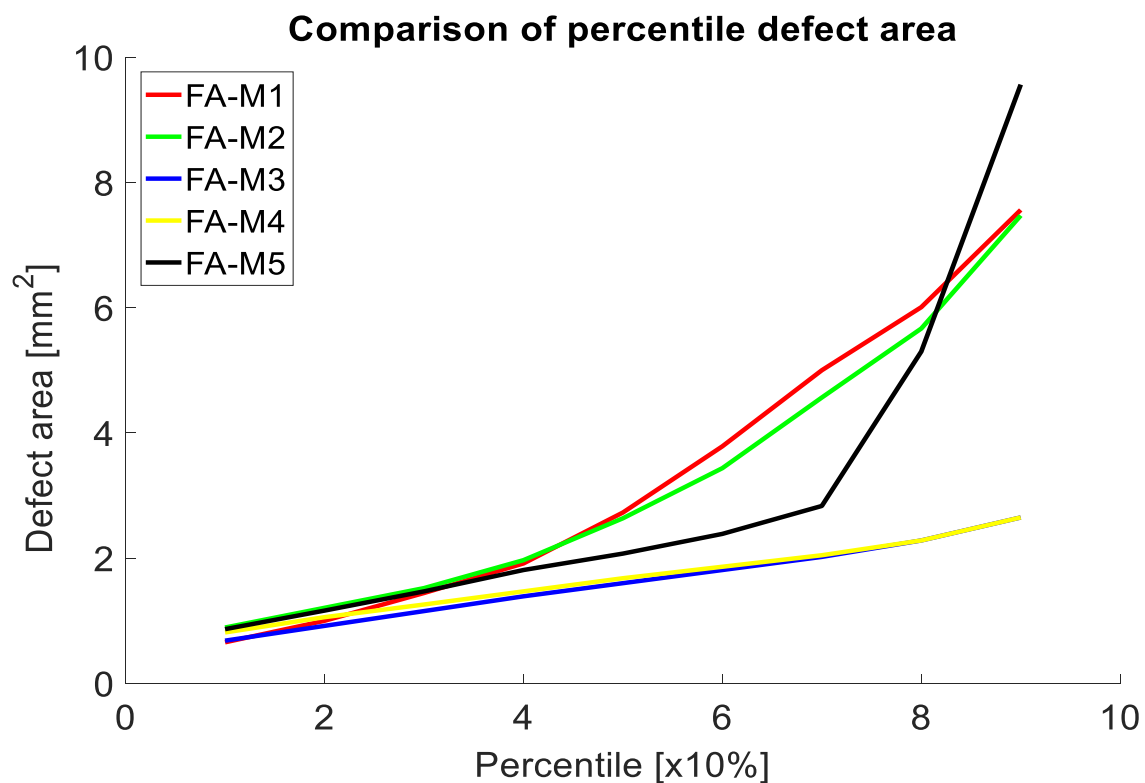
(b)
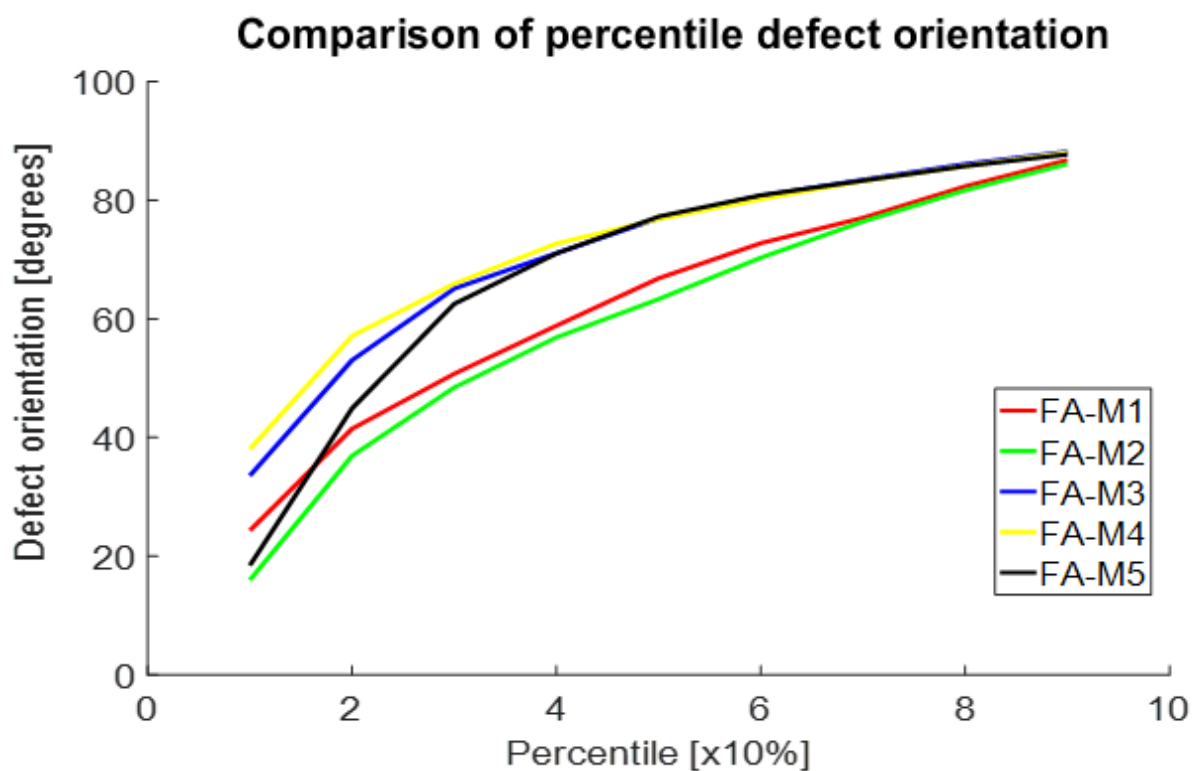
(c)

Figure 3.62: Percentile measured length (a), area (b), and orientation (c) for moderately damaged samples of data set 2.

Considering lightly damaged field acquired samples, 90% of detected defects are observed to have a maximum length of no more than 3mm with corresponding defect area of no more than 3mm^2 for all samples under investigation as depicted in Figure 3.63a and b. It is however worth mentioning that all images of this level of damage demonstrate the highest level of consistency in terms of generated statistical data. Furthermore, considering the maximum defect length and area suggests that the dominant defect type in all samples are fatigue damage of shorter lengths and smaller area as compared to the other levels of damage. As in the previous two levels of damage a similar conclusion can be made with regards the orientation of RCF defect ranging 87°-20° for 90% of detected defects as depicted in Figure 3.63c.
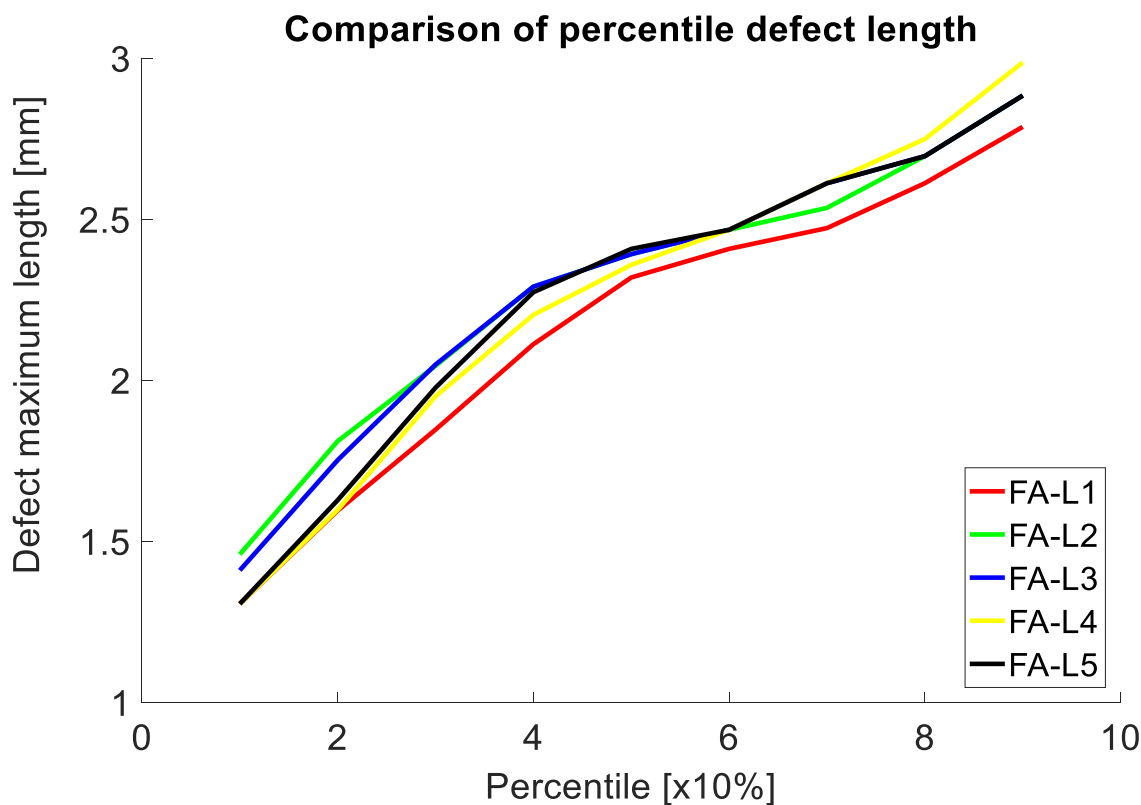
(a)
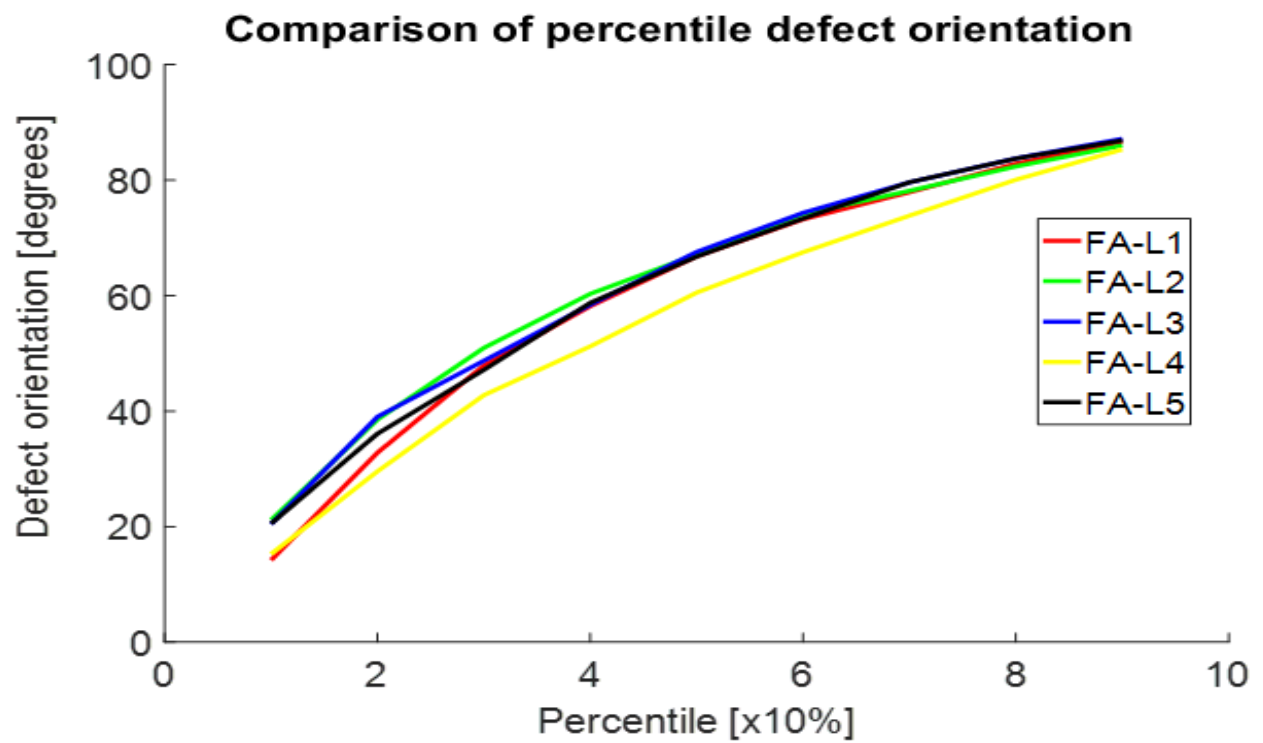
(b)
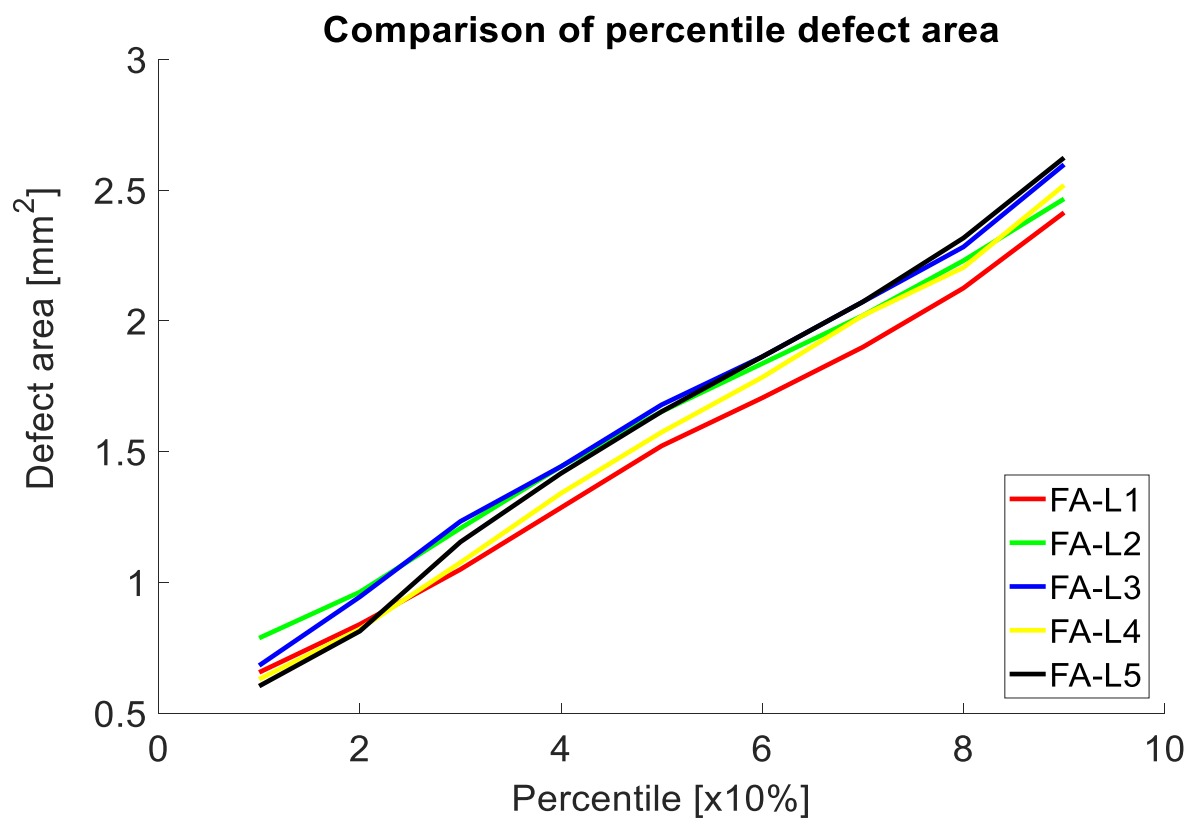
(c)

Figure 3.6326: Percentile measured length (a), area (b), and orientation (c) for lightly damaged samples of data set 2.

3.5.2 Analysis of data set 3 using the proposed algorithm

Based on optimisation by trial and error, the 30 samples of data set 3 have been investigated using the algorithm proposed in section 3.5. The simulated results are summarised as depicted in Table 3.6. The following observations can be made with regards to left rail images of the track. For the influential parameters related to enhancement, only RM is observed to be significantly higher with an appreciation of approximately 25% compared to optimum values adopted for data set 1. Even though both data set 2 and 3 are field acquired images an increase of about 120% is recorded which has been attributed to the differences in environmental conditions (i.e. gravel background for data set 2 while snow background in data set 3) including variations in lighting or reflection property of rail steel. With focus on left rail samples, it is observed that the maximum block size of the enhancement stage is no more than 32×32 as opposed to 64×64 in data sets 2. It is also observed that the normalised grey threshold value of segmentation for data sets 1, and 2 varied by no more than 18.26% compared to data set 3. Further cleaning of the defect image was done with a consistent cleaning parameter C=0.0001, for all images in data set 3. In the feature extraction stage a similar observation to data set 1 and 2 has been established, with all images utilising a constant maximum frequency (50Hz), true defect energy threshold (E=0.009), and block size N equal to the block size of LN method i.e. N=32. Simulated RCF geometrical statistical data for samples L-4060ft, L-4063ft, L-4110ft, L-3960ft, L-4563, have been observed to contain some level of spalling, with 90% of these defects observed to have a maximum surface length, area, and orientation of 6.4mm, 7mm^2, and 87° respectively. These samples are best processed with a maximum block size of 32×32 pixels. While samples L-4195ft, L-4270ft, L-3953ft, L-4103ft are observed to contain mostly fatigue damage like defects with 90th percentile of measured length, area, and orientation of defects observed to be no more than 4mm, 2mm^2 and 87° respectively. These samples are best processed with a maximum block size of 22×22 pixels. For samples L-3948ft, L-3943ft, L-3945ft, L-3978ft, L-4025ft, L-4213ft, they are observed to contain mostly fatigue damage like defects with 90th percentile length,

area and orientation of defects no more than 2.7mm, 2mm^2 and 87° respectively. These samples are best processed with a maximum block size of 16×16 pixels. According to the number of RCF defects detected it is right to assume that heavily damaged samples contain less number of defects compared to moderately and lightly damaged samples respectively.

Image	Enhancement parameters Required mean RM, Required variance RV, Local normalisation window bs_Ln, and Visibility measure window bs_Vis	Segmentation parameter (Normalised pixel intensity threshold value)	Feature extraction parameter Filter frequency (Hz), Filter energy threshold, block size	Cleaning parameter (scaling factor to determine size of blob to be cleaned)	Number of RCF detected (With expected crack population =multiple)
L-4060ft	1.2,1,32,32	0.3333	50,0.009,32	0.0001	508
L-4063ft	1.2,1,32,32	0.3843	50,0.009,32	0.0001	442
L-4110ft	1.2,1,32,32	0.3373	50,0.009,32	0.0001	487
L-4195ft	1.1,1,22,22	0.3843	50,0.009,22	0.0001	731
L-4270ft	1.1,1,22,22	0.3882	50,0.009,22	0.0001	740
L-3948ft	1.1,1,16,16	0.3843	50,0.009,16	0.0001	940
L-3953ft	1.1,2,22,22	0.4078	50,0.009,22	0.0001	707
L-3960ft	1.1,2,32,32	0.4039	50,0.009,32	0.0001	470
L-4103ft	1.1,2,22,22	0.4039	50,0.009,22	0.0001	735
L-4563	1.1,2,32,32	0.4039	50,0.009,32	0.0001	446
L-3943	1.1,1,16,16	0.3882	50,0.009,16	0.0001	880
L-3945	1.1,1,16,16	0.3882	50,0.009,16	0.0001	875

L-3978	1.1,1,16,16	0.3843	50,0.009,16	0.00007	958
L-4025	1.1,1,16,16	0.3882	50,0.009,16	0.00007	1046
L-4213	1.1,1,16,16	0.3843	50,0.009,16	0.00007	748

Table 3. 6: Presents the technical details related to each stage of the proposed novel application of image processing algorithms on left rail images.

In the case of examining the right rail equivalent image samples, Table 3.7 below suggests little or no spalling. Accordingly, a block size of no more than 16x16 has been adopted for all images in addition to RM and RV of 1.2 and 1 respectively. As in the case of left rail images, segmentation threshold values are observed to also vary by no more than 18% for all right rail images. Feature extraction has been performed as described in previous case (left rail image sample) using same parameters (maximum frequency of 50Hz, true defect energy threshold E=0.009, and block size equal to LN block size). In a similar manner to left rail image samples, post-processing cleaning has also been performed using parameter 0.0001. In the case of all right rail image samples, generated statistical data have been observed to contain 90% of these defects with a maximum surface length, area, and orientation of 2.7mm, 2mm^2, and 87° respectively.

Image	Enhancement parameters Required mean RM, Required variance RV, Local normalisation window bs_Ln, and Visibility measure window bs_Vis	Segmentation parameter (Normalised pixel intensity threshold value)	Feature extraction parameters Filter frequency (Hz), Filter energy threshold, block size	Cleaning parameter (scaling factor to determine size of blob to be cleaned)	Number of RCF detected (With expected crack population =multiple)
R-4060ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	886
R-4063ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	906
R-4110ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	813
R-4195ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	905

R-4270ft	1.2,1,16,16	0.3843	50,0.009,16	0.0001	847
R-3948ft	1.2,1,16,16	0.3843	50,0.009,16	0.0001	900
R-3953ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	887
R-3960ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	910
R-4103ft	1.2,1,16,16	0.3882	50,0.009,16	0.0001	853
R-4563	1.2,1,16,16	0.3882	50,0.009,16	0.0001	926
R-3943	1.2,1,16,16	0.3882	50,0.009,16	0.0001	868
R-3945	1.2,1,16,16	0.3882	50,0.009,16	0.0001	861
R-3978	1.2,1,16,16	0.3882	50,0.009,16	0.0001	759
R-4025	1.2,1,16,16	0.3882	50,0.009,16	0.0001	870
R-4213	1.2,1,16,16	0.3843	50,0.009,16	0.0001	868

Table 3. 7: Presents the technical details related to each stage of the proposed novel application of image processing algorithms on right rail images.

As a holistic analysis of the simulated result (images presented in Appendix C6) for data set 3 (both left and right rails), it can be said that the detected RCF damage (Table 3.6 and 3.7) is in agreement to field observations and theoretical understanding of curved track rail damage data. Based on the assumption that the investigated damaged samples reported in this section of the thesis are most probably of different curve radiuses, considerations of rail cant deficiency will cause the low rail to contain a higher level of damage (due to lower creep forces resulting to lower T_γ values sufficient enough to generate RCF but low enough to not wear these crack off). Field measurement have suggested a similar conclusion because low wear rates are realised in such a case. On the high rail, it is expected that higher creep forces resulting in higher T_γ will cause severe wear and possible wearing out of initiated RCF defects before they propagate to critical geometry. It is therefore believed that the left rail images of data set 3 correspond to low rail, while right images are for the high rail of a left hand curved

track. This is confirmed from ICRI site data details provided along with the data set 3. It is therefore confirmed that there is good agreement between site data and simulated results of this section of the thesis.

3.6 Conclusion

This chapter highlighted the effectiveness of using image processing algorithms for the automatic detection of RCF damage in rails, with emphasis on generating defect geometry data for rail integrity predictions. A critical requirement for automatic detection of rail damage is the quality of the image acquisition system. More accurate detection of cracks is attainable by using of image acquisition systems equipped with anti-vibration dampers (for blur free images) and light source deflectors/reflectors (for uniform illumination of rail surface).

The analysis of simulated results for all data sets confirms that no single method (at each stage of the algorithm) is equipped to solve the challenges of true positive detection of damage on rails. The computational requirement of each method is recorded as shown in Table 3.8 below.

Method	Recorded time (sec)
Median filter	0.50
LMAF enhancement	3600
Local normalisation enhancement	0.56
FFT enhancement	0.73
Visibility measure enhancement	1.31
OGT segmentation	0.53
AIT segmentation (2 initial partitions)	2.05
Feng's segmentation	0.52
Texture based feature extraction	1.20
SDHSD feature extraction	5.52
Morphological feature extraction	1.04

Table 3. 8: Show the computational time for each algorithm.

Rather, a key finding from this investigation reiterates the benefits of combining one or more methods at each distinct stage. After thorough analysis of simulated results, the significance of illumination invariant methods cannot be neglected. Also it was ensured that even global methods (such as Normalisation) are modified to local versions for improved performance. Critical analysis of the results obtained after applying the algorithms confirm the presence of undesired noise, which is attributed to the susceptibility of defining the existence of a crack as variation in pixel intensity. This assumption leads to the detection of grind marks, scratches, and rust as possible damage on rails and must be addressed using superior definition of true detection of damage. Measurements from high and low rail components of data set 3, confirm theoretical understanding of RCF initiation, propagation, and even the competitive mechanism of wear and crack growth rate in curved track. Given that further investigation of the proposed algorithm is possible, it is therefore of interest for this research to explore (in the next chapter) the potential applications of the generated defect geometrical data related to each defect with the hope of enhancing maintenance team's planning/decision making and predictive maintenance.

Chapter 4 Modelling and simulation of crack behaviour in rails

4.1 Introduction

There exist several models of crack initiation and growth in literature, a variety of which contribute to understanding the concept of crack initiation, propagation, and failure in rails. A wide range of such models have been discussed in the literature review section, which take into account the fundamental equilibrium, deformation, and material equations of solid mechanics. In line with the aim and objective(s) of this PhD research, the mathematical formulations governed by Hertz contact theory are the basic input for estimating the stress concentration at the crack tip, which is further related to the Stress Intensity Factors (SIF) of tensile, shear, and torsion loading cases. Using the data set detailed in section 4.2, this chapter discusses a finite element wheel-rail contact model; henceforth referred to as a global track model, utilised for generating wheel-rail contact stresses and strains responsible for propagating RCF damage. It is designed to accommodate a single sleeper bay with a sleeper distance of 600mm for realistic predictions of field observations. The global track model is initialised based on VAMPIRE generated contact data (these include the wheelset lateral displacement, vertical displacement, roll angle, and yaw angle). The simulated contact stresses and strains realised from the global track model are further processed using a locally incorporated fracture mechanics model.

The local fracture model defined in the mathematical formulations of section 4.3 is adopted to estimate the dominant modes of fracture for surface initiated cracks (i.e. Mode-I and II). The Mode-I and II SIFs are then used to derive equivalent mixed mode SIF. This equivalent SIF in relation to fracture toughness of the rail steel is further processed in accordance with well-understood growth laws for derivation of crack growth rate. The crack growth direction is determined based on maximum tangential stress theory and rail residual life is estimated as a function of the number of loading cycles until failure.

Other influential parameters considered in this chapter include multi-axial loading considerations, rail supporting structure, track curvature, defect geometry (i.e. crack size and orientation/inclination), and relative distance of the crack from peak contact stress.

Due to the complexity associated with stresses generated at the contact patch the mixed mode loading (as opposed to pure mode I and II) is considered as the ideal replica of all wheel-rail contact phenomena.

It is also important while considering fracture mechanics models to acknowledge the impact of rail substructure design, as concluded in (Evans, G., et al 2009); such as sleeper spacing and damping constants to be proportional to reaction forces, i.e. vertical displacements, and contact forces respectively. It is further recognised in the works of Zakeri, A.J., & Xia, H. (2008) that for spring stiffness, there exists an inverse proportionality with rail vertical displacement and negligible increase in support structure interaction (reaction) forces.

Curved track data is of interest to this study because longitudinal and lateral slip can be related to the creep forces directly responsible for the propagation of RCF cracks. The simulated results of curve radius parametric study in section 4.4.3 confirms an inversely proportional relationship between curve radius and contact pressure including creep forces.

It is also common knowledge that the initial length of a defect influences its fracture behaviour. In section 4.4, simulated results of newly initiated (1mm surface length) cracks and critically grown cracks (5mm surface length) have been used to confirm that longer cracks propagate faster than shorter ones under same loading conditions. The global contact model is validated with specialist contact dynamics software, while local fracture mechanics model is compared to laboratory tests by Zafosnik, B., et al (2000) in section 4.5.

4.2 Data set

4.2.1 Contact data

The global track model is a 3D assembly of a revolved and extruded s1002 wheel and 60E1 rail profiles respectively. The profile data have been obtained via scans (on new samples) from a Mini-Prof measurement device. Both tangent and curved track studies are in response to 80KN axle load (typical of in service passenger train axle loads).

In the case of tangent track case study, the VAMPIRE vehicle dynamics simulation package data summarised in Table 4.1 is used to calibrate the global track model.

Lateral displacement [mm]	Roll angle [mrad]
1	-0.053
2	-0.114
3	-0.179
4	-0.252
5	-0.327
6.5	-0.730

Table 4. 1: Tangent track calibration data for global track model.

In the case of curved track case study, the VAMPIRE vehicle dynamics simulation package data summarised in Table 4.2 is used to calibrate the global track model. The parameters of interest to this research include yaw angle, roll angle, vertical displacement, and lateral displacement of the low rail components of the track respectively.

Curve radius [m]	Lateral Displacement [mm]	Roll angle [degrees]	Yaw angle [degrees]	Lateral creep force [kN]	Longitudinal creep force [kN]
200	-7.67	-0.63	3.32	-15.55	-7.50
650	-5.93	-0.09	3.68	-6.77	-14.35
900	-5.14	-0.05	4.18	-3.99	-12.41
1600	-3.74	-0.01	4.15	-1.27	-7.81

Table 4. 2: Curved track calibration data for global track model.

4.2.2 Defect geometry, and location

The influence of the initial size of the defect under investigation is also parameterised as a function of crack surface length, depth, and opening. A case study of defect propagation and branching criteria ranging from a newly initiated crack with surface length of 1mm (initial and final rail crack propagation phase), to that of a critically grown defect of surface length 5mm is presented and analysed. The depth of each defect is determined to be equal to surface length based on the assumption of a unity depth-length ratio. Furthermore, a sensitivity analysis on the influence of the relative distance between the crack centroid and peak contact pressure is performed. The relative distance is defined in Cartesian coordinate system such that (x,y) corresponds to the coordinate of the peak contact pressure at the wheel rail interface and the following crack centroid locations are studied (x-0,y-0), (x-1,y-1), and (x-2,y-2).

4.3 Methodology

The relative motion between the wheel and rail is best described as a combination of rolling and sliding motion such that in the presence of an external normal load, some points may slip and others may stick (adhesion). Iwnicki, S. (2003) explains that it is as result of difference between the tangential strains of the two bodies in the adhesion region that small slip called creepage are realised at the wheel-rail interface. When a wheelset is subjected to acceleration, braking, suspension forces, or curved trajectory, creep forces are generated due to micro slippage in the area of contact. In addition to the normal contact force (N), the creep forces generated within the contact patch that are of interest to this research include, longitudinal creep force (F_x), and lateral creep force (F_y). As a direct consequence of adopting image processing for detection of RCF damage in this study, only N, F_x, F_y components of the creep forces (see Figure 4.1) are influential for surface fracture mechanics (while M_z is neglected).
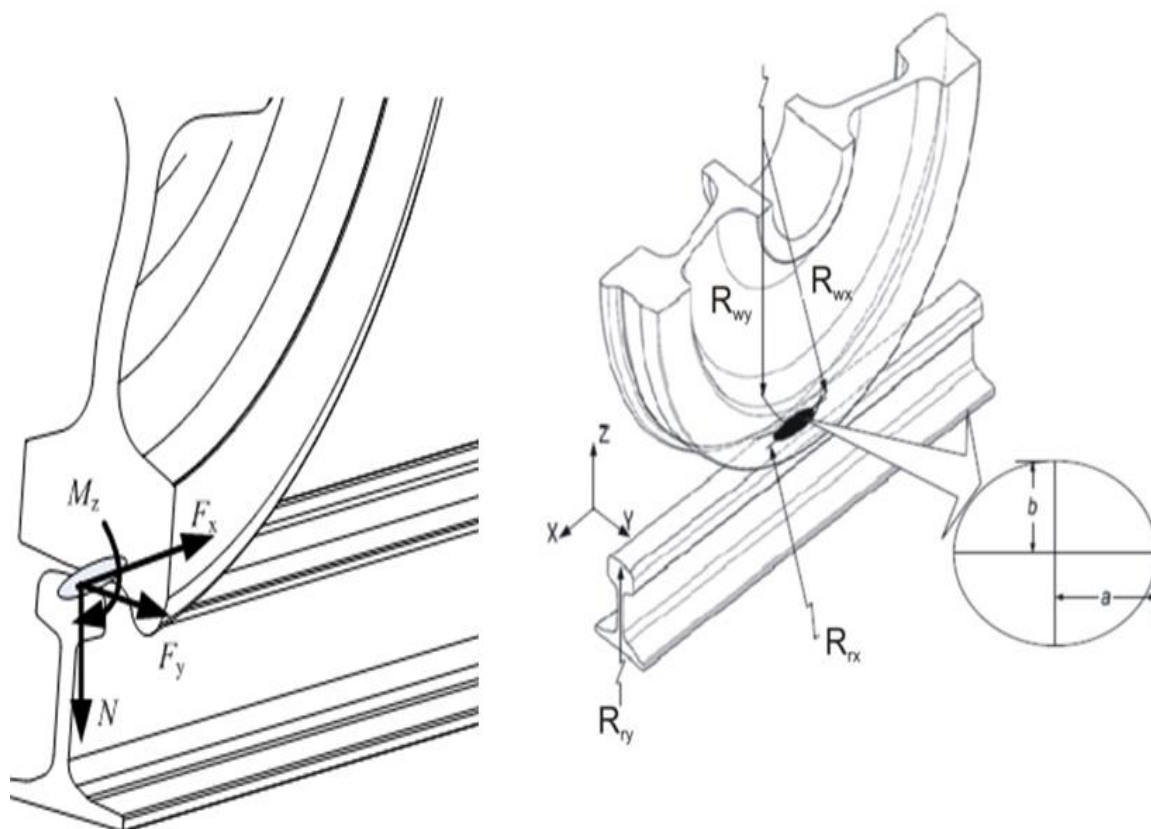
Figure 4. 1: Wheel rail contact showing contact forces and contact patch (Lewis, R., & Olofsson, U. 2009).

4.3.1 Global track model

To accurately implement a multibody contact model, it is necessary to introduce a parametric form of the geometry of the wheel and rail. This definition of the geometry is required to be defined in a global coordinate system since the equations of motion of the contact model are written with respect to global inertial frames capable of representing any spatial configuration of the wheel and rail. The global track Finite Element (FE) model depicted in Figure 4.2, is designed for the s1002 wheel and 60E1 rail profile. This FE approach is preferred in order to account for the initial gap/penetration that exists between the wheel and rail profiles and can be included. Furthermore, the contact model considers linear elastic plastic deformation around the vicinity of the contact patch. The inclusion of extra fine mesh divisions near the contact patch is made in order to improve the accuracy of estimating stress concentration at the crack tip.
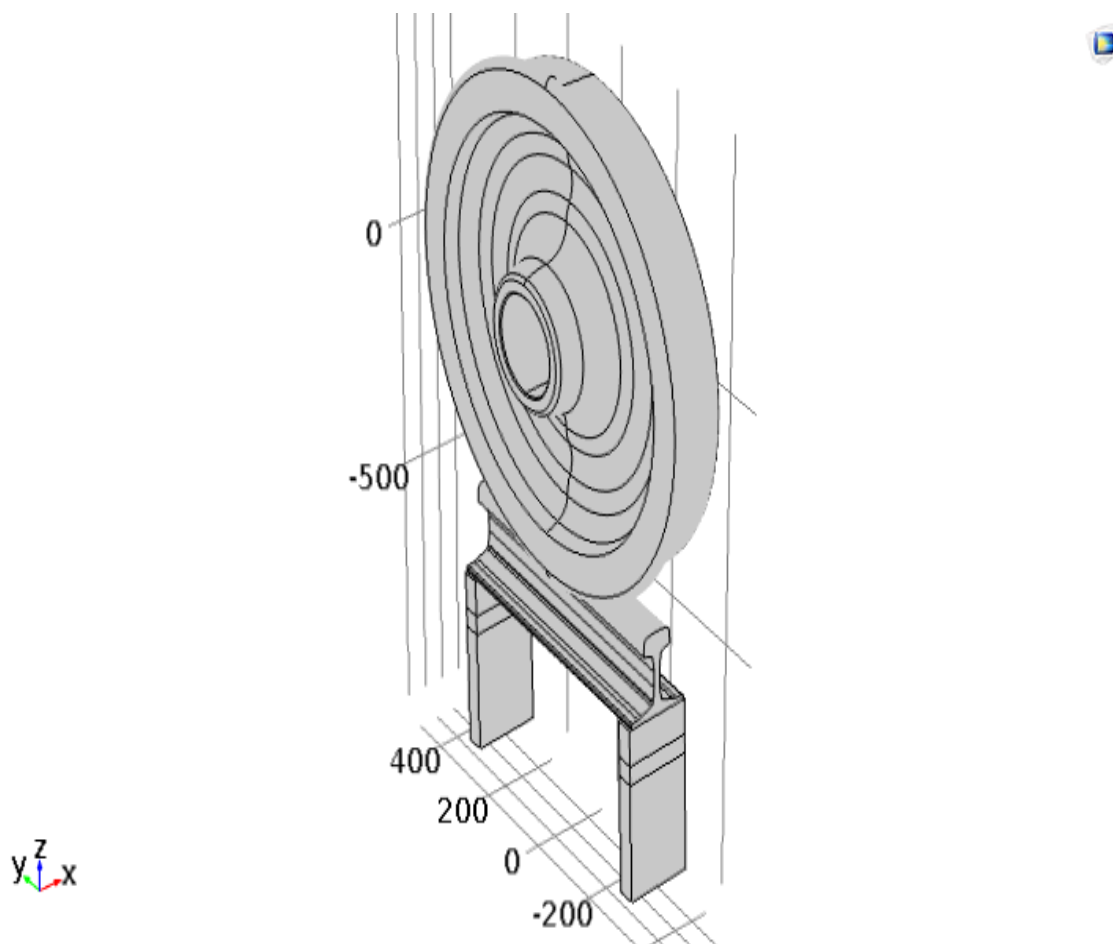
Figure 4. 2: The global track model assembly of s1002 wheel, and 60E1 rail, rail pad, sleeper, and ballast substructures.

The influence of including the rail substructures with technical details of spring stiffness as summarised in Table 4.3, is observed to lower the contact pressure by only few MPa (approximately 3.173 MPa).

Parameter	Value
Rail pad stiffness	200E6 [N/m]
Sleeper mass	314 [Kg]
Ballast stiffness	125E6 [N/m]

Table 4. 3:Details of rail supporting structure. Vasic, G. (2013).

Quasi-static wheel-rail rolling contact conditions, best confined in the assumptions of steady state Hertzian contact theory elaborated in Ahmad, A., et al (2008) and Vasic, G. (2013) combined with theory of elastic deformation as discussed in Popovici, R. (2010) result in

equation 4.1 below. The maximum contact pressure assumed to be at the origin $P(0,0)$ in equation 4.2, for 80 kN axle load typical of passenger train is evaluated to be about 1027MPa.

$$P(x,y) = \frac{3}{2}\frac{N}{\pi ab}\sqrt{1 - \left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2} \tag{4.1}$$

$$P(0,0) = \frac{3}{2} \times \frac{80,000}{\pi \times 6 \times 5} = 1027\frac{N}{mm^2} \tag{4.2}$$

Where N is the normal force in contact patch, and the major and minor semi axis of contact ellipse are denoted by a and b.

Figure 4.3 below highlights the difference in contact pressure distribution between inelastic deformations in MATLAB based on Hertz model and measured normal pressure distribution for new and worn profile shapes respectively. The effect of improved conformity between wheel and rail worn shapes is observed to case a 400MPa reduction in contact pressure.
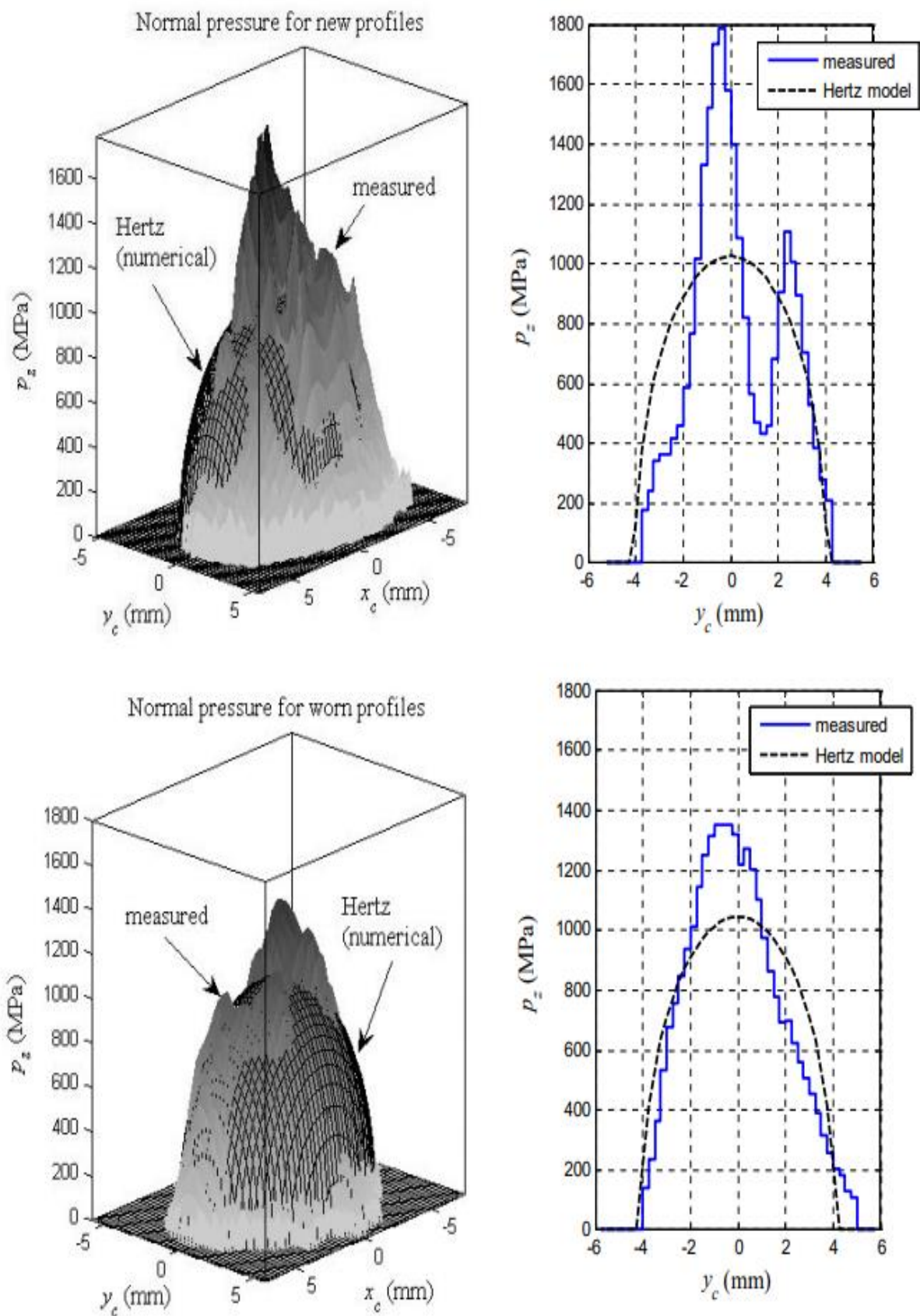
Figure 4. 3: Shows the comparison Hertz model and measured contact pressure of a new and worn profile (Rovira, A., et al 2011).

Bezin. Y, (2008) extensively investigated integrated flexible track models and discusses the force exerted on the rail support structure. Generally, vertical force on the rail pad and ballast varies along the track and can be observed to be influenced by the cant deficiency. Also a difference of about -20% to +13% is observed between the forces predicted using constant and variable support stiffness. Bezin. Y, et al (2009) emphasises the significance of using sufficiently long enough section of a track to investigate influence of support structure on contact forces, which confirms the existence of potential trend between support stiffness variation and changes in forces on track. In the global contact model presented in section 4.3.1 decreasing rail pad stiffness (for tangent track) by 50% is observed to decrease displacement of rail proportionally with a negligible change in support interaction forces by as little as 2.58N, 0.541N, 3.7N in the vertical, lateral and longitudinal directions. Further decrease of rail support parameters to about 25% of original value in Table 4.3 is observed to decrease rail displacement proportionally and increase support reaction forces by 5.15N, 1.06N, 7.5N in vertical, lateral, and longitudinal directions respectively. Also the total vertical displacement of the rail is observed to increase by about 100% to approximately 0.098mm compared to a non-supported rail section. This is attributed to appreciation in bending moment of the rail.

Even though contact stresses have been extracted from FE analysis of the wheel rail contact model in this thesis, Srivastava, J.P., Sakar, P.K., & Ranjan, V. (2014) elaborated as presented in equation 4.3a to 4.3c, the analytical estimates of principal stresses as a function maximum contact pressure.

$$\sigma_x = -2\mu P_{max} - (1 - 2\mu)P_{max}\frac{b}{a+b} \tag{4.3a}$$

$$\sigma_y = -2\mu P_{max} - (1 - 2\mu)P_{max}\frac{a}{a+b} \tag{4.3b}$$

$$\sigma_z = -P_{max} \tag{4.3c}$$

Where σ_x, σ_y, σ_z are principal stresses. μ is the friction coefficient. P_{max} is the maximum contact pressure. a and b are the major and minor semi axis of the contact patch.

4.3.2 Local fracture mechanics model

The local fracture mechanics model adopted for predicting the behaviour of cracks on rail is based on linear elastic fracture mechanics formulations of plane stress assumptions (i.e. stress distribution on thin plane containing crack). This PhD investigation focuses on tension (mode-I) and shear (mode-II) fracture modes mainly due to their dominance is surface RCF damage propagation as elaborated in Lewis, R., & Olofsson, U. (2009). Pure, mixed, and biaxial loading case studies have been demonstrated and presented in Appendix D1, Appendix D2 and Appendix D3 respectively. The method depicted in Figure 4.5, involves exporting the wheel rail interface stress distribution generated from the global track model. The stress distribution is processed to estimate KI and KII as functions of nominal stress acting on a finite plane containing the RCF crack. The equivalent effect of KI and KII represented as Keqv is estimated. By relating Keqv to material fracture toughness of the rail steel, the growth rate of the defect is estimated as function of ΔKeqv. The branching direction θ, is assumed to be same as the angle of the maximum tangential stress component. Residual life estimates have also been determined as functions of number of loading cycles until failure. For elliptical equivalent representation of RCF damage on rails, mathematical formulations of fracture mechanics parameters related to a variety of defect type/configuration (see section 4.4.1) have been presented in this section. All equations detailed in this section are functions of nominal stress and exclude Irwin's modification of crack tip yielding (existence of a small plastic region at crack tip). Equation 4.4a and 4.4b describe purely tensile crack fracture and its independence on the influence of crack inclination angle, while equations 4.5 - 4.21 are also uniaxial formulations but are related to generic crack inclination angle of cracks on rails. For more realistic prediction of RCF damage behaviour in rails, bi-axial models of KI and KII as described in equation 4.22 are adopted in this PhD research.
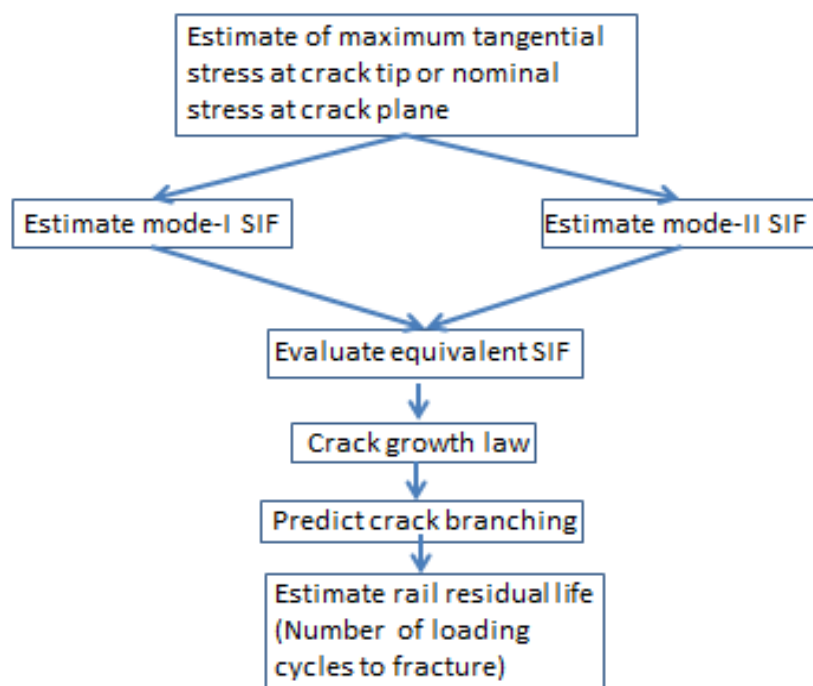
Figure 4. 4: Shows the block diagram of the proposed local fracture mechanics model.

It is well established in the book of Dalberg, T., & Ekberg, A. (2002) and Anderson, T.L. (2011) that the classical J-integral formulation can be related to SIF by a function of remote nominal stress, crack length and Young's modulus as prescribed in equation 4.15a and 4.15b for pure Mode-I and pure Mode-II SIF respectively.

$$J_I = \frac{K_I{}^2}{E'} = \frac{\sigma_{z\infty}{}^2 \pi a}{E'} \times f \qquad (4.4a)$$

$$J_{II} = \frac{K_{II}{}^2}{E'} = \frac{\sigma_{zy\infty}{}^2 \pi a}{E'} \times g \qquad (4.4b)$$

Where f and g are the configuration correction factors introduced by Dalberg, T., & Ekberg, A. (2002) that depend upon the material and geometry, which have been concluded to be negligible for simple geometries. The symbol ∞ denotes remote stress far away from crack.

Taking into account the case of an inclined ellipse in a wide plate under the influence of uni-axial load represented by the nominal stress σ_{nom} in Figure 4.6 below, the tangential stress

σ_θ along the free border of the elliptical crack geometry as reported by Livieri, P., & Seagala, F. (2007) can be expressed as in equation 4.5.
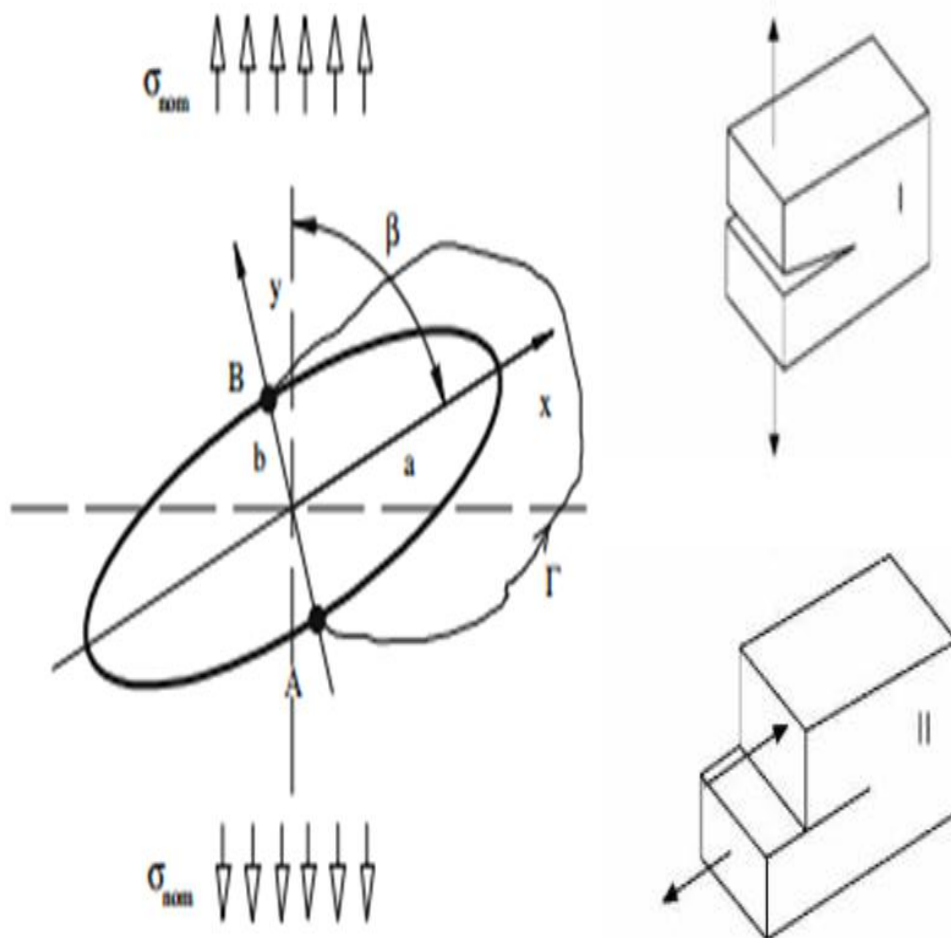
Figure 4. 5: Inclined elliptical crack under tensile load (right), and relevant mode-I and II fracture modes (left) (Livieri, P., & Seagala, F. 2007)**.**

$$\sigma_\theta(\beta, \eta, \xi_o) = \sigma_{nom} \left[\frac{sinh2\xi_o + cos2\beta - e^{2\xi_o} cos\, 2(\beta - \eta)}{cosh2\xi_o - cos2\eta} \right]$$

(4.5)

Where η and ξ_o are the elliptical coordinates. β is the angle between major axis (a) and the load direction as depicted in Figure 4.6.

In order to obtain analytical expression of the J integral for generic inclination angle β, it is useful to define symmetric and skew-symmetric components elaborated in equations 4.6a - 4.6c below.

$$\sigma_\theta = \sigma_{\theta,E} + \sigma_{\theta,O}$$

(4.6a)

$$\sigma_{\theta,E} = \frac{1}{2}[\sigma_\theta(\beta, \eta) + \sigma_\theta(-\beta, \eta)]$$

(4.6b)

$$\sigma_{\theta,O} = \frac{1}{2}[\sigma_\theta(\beta,\eta) - \sigma_\theta(-\beta,\eta)] \tag{4.6c}$$

Utilising trigonometric and hyperbolic functions in equations 4.6 – 4.11 below.

$$x = a\,cos\eta \tag{4.7a}$$

$$y = b\,sin\eta \tag{4.7b}$$

$$a = c\,cosh\xi_o \tag{4.8a}$$

$$b = c\,sinh\xi_o \tag{4.8b}$$

$$c = \sqrt{a^2 - b^2} \tag{4.8c}$$

$$\sinh(2\xi_o) = \frac{2ab}{c^2} \tag{4.9a}$$

$$\cosh(2\xi_o) = 1 + \frac{2b^2}{c^2} \tag{4.9b}$$

$$\tanh(\xi_o) = \frac{b}{a} \tag{4.9c}$$

$$e^{2\xi_o} = \frac{c^2}{(a-b)^2} \tag{4.10}$$

$$\sin(2\eta) = \frac{2xy}{ab} \tag{4.11a}$$

$$\cos(2\eta) = 1 - \frac{2y^2}{b^2} \tag{4.11b}$$

Where x and y are the equivalent Cartesian coordinate representation of η and ξ_o. While a is the major semi axis and b is the minor semi axis of the elliptical crack. And c is the ellipse eccentricity.

The even and odd hoop stresses in equation 4.6b and 4.6c are re-arranged as a functions of y co-ordinate as expressed in equation 4.12a and 4.12b.

$$\sigma_{\theta,E}(a,b,c,y) = \sigma_{nom}\frac{c^2}{(a-b)^2}\left[\frac{ab}{c^2}\frac{1+\frac{(a-b)^2}{c^2}}{\frac{b^2}{c^2}+\frac{y^2}{b^2}} - 1\right] \tag{4.12a}$$

$$\sigma_{\theta,O}(a,b,c,y) = \sigma_{nom}\frac{2c^2}{(a-b)^2}\left[\frac{y}{b^2}\frac{\sqrt{b^2-y^2}}{\frac{b^2}{c^2}+\frac{y^2}{b^2}}\right] \tag{4.12b}$$

The x-y co-ordinate dependent expression in equation 4.12 is more suitably evaluated analytically in accordance with equation 4.13. This integral is performed along the minor axis (integral from point $-b$ to b) of the V rounded shape notch (i.e. an ellipse with b that tends towards 0).

$$J_v = \int_{-b}^{b} \frac{\sigma_\theta{}^2}{2E'} dy \qquad (4.13)$$

Where J_v is the J-integral of a V-shaped notch (ellipse with b that tends towards 0). b is the minor semi axis length of the elliptical notch. E' is the Young's modulus in the case of plane stress assumption.

Taking into account equations 4.7 - 4.13, the mode-I SIF at the tip of the crack for a generic inclination angle β is determined as follows.

$$K_I = \sigma_{nom}\sqrt{\pi a} \times \sin^2(\beta) \qquad (4.14a)$$

While the mode-II SIF at the tip of the crack is estimated for generic loading angle β by:

$$K_{II} = \sigma_{nom}\sqrt{\pi a} \times \sin(\beta)\cos(\beta) \qquad (4.14b)$$

where: σ_{nom} is nominal stress component, , K_I mode-I SIF, K_{II} mode-II SIF, E' is the shear modulus, inclination angle β, a is crack major semi axis, b is minor semi axis of crack, c is eccentricity of crack.

Livieri, P., & Seagala, F. (2012) modified equations 4.15 to account for mixed mode I and II loading case. This modification is based on the re-definition of the hoop stress in equation 4.5 as the sum of three stress components defined as;

$$\sigma_\theta = \lambda_1\sigma_{\theta 1} + \lambda_2\sigma_{\theta 2} + \lambda_3\sigma_{\theta 3} \qquad (4.15a)$$

$$\sigma_{\theta 1} = \frac{e^{2\xi_o}}{1+2\frac{a}{b}}\left[\frac{sinh2\xi_o(1+e^{2\xi_o})}{cosh2\xi_o-cos2\eta} - 1\right] \qquad (4.15b)$$

$$\sigma_{\theta 2} = \frac{2abe^{2\xi_o}}{(a=b)^2}\left[\frac{sin2\eta}{cosh2\xi_o-cos2\eta}\right] \qquad (4.15c)$$

Where the coefficients $\lambda_1, \lambda_2, \lambda_3$ (see equation 4.16) are obtained by forcing the summation of the right hand side of equation 4.18a to the left hand side using FE analysis.

$$\lambda_1 = \sigma_{nom}(1 + 2\frac{a}{b})sin^2\beta \qquad (4.16a)$$

$$\lambda_2 = \sigma_{nom}\left(\frac{(a+b)^2}{ab}\right) - sin\beta cos\beta \qquad (4.16b)$$

The analytical formulation of J-integral for mixed mode V rounded shape notch case is achieved by considering only the two components of equation 4.15a (i.e. $\sigma_{\theta1}$, $\sigma_{\theta2}$) that agree with the classic J-integral when notch tip radius of the ellipse is almost zero. The analytical solution for mixed mode I and II J-integral of a V rounded shape notch J_{vmI} and J_{vmII} are obtained by substituting equation 4.15b and 4.15c in equation 4.13 and the results are presented in equation 4.27a and 4.27b.

$$J_{vmI} = \frac{\lambda_1^2}{(1+2\frac{a}{b})^2 E'}\left\{\frac{2a^4}{c(a-b)^2}\left[\frac{cb}{a^2} + \tan^{-1}\left(\frac{c}{b}\right)\right] - \frac{4a^2bc}{(a-b)^3}\tan^{-1}\left(\frac{c}{b}\right) + \frac{bc^4}{(a-b)^4}\right\} \tag{4.17a}$$

$$J_{vmII} = \frac{\lambda_2^2}{E'}\left\{\frac{4a^2b^2c}{(a+b)^2(a-b)^2}\left[\frac{1}{2}\tan^{-1}\left(\frac{c}{b}\right) - \frac{bc}{2a^2}\right] - \frac{4a^2b^4}{(a+b)^{5/2}(a-b)^{5/2}}\left[\frac{c}{b} - \frac{3}{2}\tan^{-1}\left(\frac{c}{b}\right) + \frac{bc}{2a^2}\right]\right\} \tag{4.17b}$$

Livieri, P., & Seagala, F. (2012) further related the mixed mode I and II J-integral formulations (J_{vmI} and J_{vmII}) of a V rounded shape notch (ellipse with $b \to 0$) to the classic J_I and J_{II} formulations of a crack, based on analysis of FE simulated results as elaborated below.

$$J_I = \frac{K_I^2}{E'} = \frac{J_{vmI}}{\frac{b}{a}\left(\frac{21}{5\pi}-1\right)+1} \tag{4.18a}$$

$$J_{II} = \frac{K_{II}^2}{E'} = \frac{J_{vmII}}{\frac{b}{a}\left(\frac{125}{15\pi}-1\right)+1} \tag{4.18b}$$

More realistic rail fracture predictions of surface initiated RCF damage, is ensured by adopting multi-axial formulations of SIF. Based on Figure 4.7, biaxial tension ($\aleph = 1$) and/or compression ($\aleph = -1$) descriptions of KI and KII are elaborated in equations 4.19 as elaborated in Shlyannikov, V.N., Tumanov, A.V. (2011).
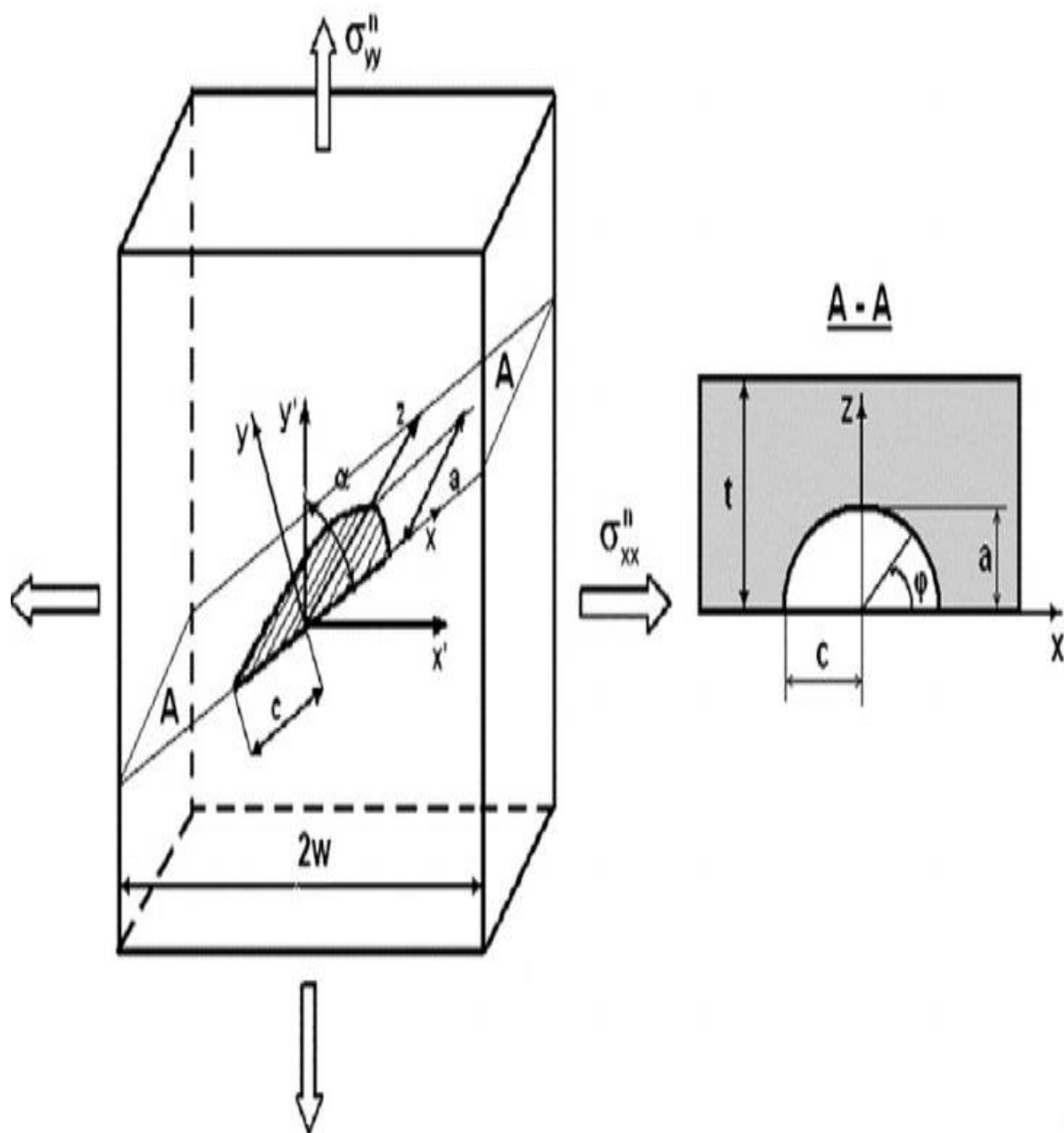
Figure 4. 6: Biaxial loading of inclined elliptical crack (Shlyannikov, V.N., Tumanov, A.V. 2011).

$$K_I = \sigma_I \sqrt{\pi L} \qquad (4.19a)$$

$$K_{II} = \sigma_{II} \sqrt{\pi L} \qquad (4.19b)$$

$$L = \frac{1}{\sqrt{\left(\frac{cos\varphi}{c}\right)^2 + \left(\frac{sin\varphi}{c}\right)^2}} \qquad (4.19c)$$

$$\sigma_I = \sigma_{yy}{}^n (\cos^2\alpha + \aleph\sin^2\alpha) \qquad (4.19d)$$

$$\sigma_{II} = \sigma_{yy}{}^n \times \frac{1-\aleph}{2} (\sin2\alpha + \cos\varphi) \qquad (4.19e)$$

$$\aleph = \frac{\sigma_{xx}{}^n}{\sigma_{yy}{}^n} \tag{4.19f}$$

Where n=1 is the adopted hardness parameter. α is the crack inclination angle, L is crack face length, $\aleph$ is the bi-axial loading factor.

For the purpose of predicting RCF surface behaviour at the wheel-rail interface, Lewis, R., & Olofsson, U. (2009) presented equivalent SIF formulations to account for mixed mode I and II fracture mechanics. This formulation detailed in equation 4.20 in collaboration with equation 4.22, is adopted in this PhD investigation due to its close correlation with extensive biaxial experimental tests using fatigue specimens of normal grade rail steel cut from the rail web.

$$\Delta K_{eq} = \sqrt{\Delta K_I{}^2 + [\frac{614}{507}\Delta K_{II}{}^{3.21}]^{2/3.74}} \tag{4.20}$$

Where ΔK_{eq} is the range of equivalent SIF. ΔK_I is the range in Mode-I SIF. ΔK_{II} is the range in Mode-II SIF

Lewis, R., & Olofsson, U. (2009) extended the application of the estimated range of equivalent SIF (ΔK_{eq}) to crack growth calculations based on same biaxial fatigue experiments as above. However, in the latter experiment the primary focus is to model the growth rate ($\frac{da}{dN}$) which has been observed to be similar to the second phase (linear growth stage) of the well-known Paris growth law. This model is best described as a function of mixed mode I and II SIF including the influence of mixed mode SIF threshold value (ΔK_{th}) as elaborated in equation 4.21 below.

$$\frac{da}{dN} = 0.000507(\Delta K_{eq}{}^{3.74} - \Delta K_{th}{}^{3.74}) \tag{4.21a}$$

$$\Delta K_{th} = \frac{4\sqrt{2}K_{II}{}^3(K_I + 3\sqrt{K_I{}^2 + 8K_{II}{}^2})}{(K_I{}^2 + 12K_{II}{}^2 - K_I\sqrt{K_I{}^2 + 8K_{II}{}^2})^{\frac{3}{2}}} \tag{4.21b}$$

The Maximum tangential stress (MTS) criterion has been adopted based on equation 4.22, assuming crack kink angle in the direction (θ) of maximum tangential stress within the vicinity of crack tip.

$$tan\frac{\theta}{2} = \frac{-2K_{II}}{K_I + \sqrt{K_I{}^2 + 8(K_{II}{}^2)}} \tag{4.22}$$

Ravaee, R & Hassani, A. (2007) explains that the number loading cycles until fracture for rail section containing a crack of known initial length is a useful tool in predictive rail maintenance. This application of equation 4.23 will however require accurate understanding of the critical crack length at which catastrophic fracture will occur. Furthermore, the amplitude of the cyclic stress and material constants are also significant in modelling the residual life calculation of a fractured rail.

$$N_f = \frac{\left|a_f^{\,1-m/2} - a_i^{\,1-m/2}\right|}{\frac{2-m}{2m}A(\Delta\sigma)^m \pi^{m/2}}$$ (4.23)

Where $\Delta\sigma$: amplitude of cyclic stress, a_i is the initial crack length and a_f is crack final length. M and A are the Paris equation material constants as presented in Table 4.4.

Material	m	A
Steel	3	10^{-11}
Aluminium	3	10^{-12}
Nickel	3.3	4×10^{-12}
Titanium	5	10^{-11}

Table 4. 4: Shows the numerical parameters related to residual life equation (Ravaee, R., & Hassani, A. 2007).

The consistency of different crack propagation methods has long been established in earlier works of Richard, H.A., & Benitez, K. (1983). An inclination sensitive experimental investigation of crack propagation has been presented in Zafosnik, B., et al (2000). This experiment utilises a compact tension specimen (see Figure 4.37) alloy of aluminium, magnesium, and manganese (ALMgMn4.5-W32) with a Young's modulus of 72,400MPa, 314.5MPA tensile strength, and Poisson ratio of 1297N/mm^3/2.
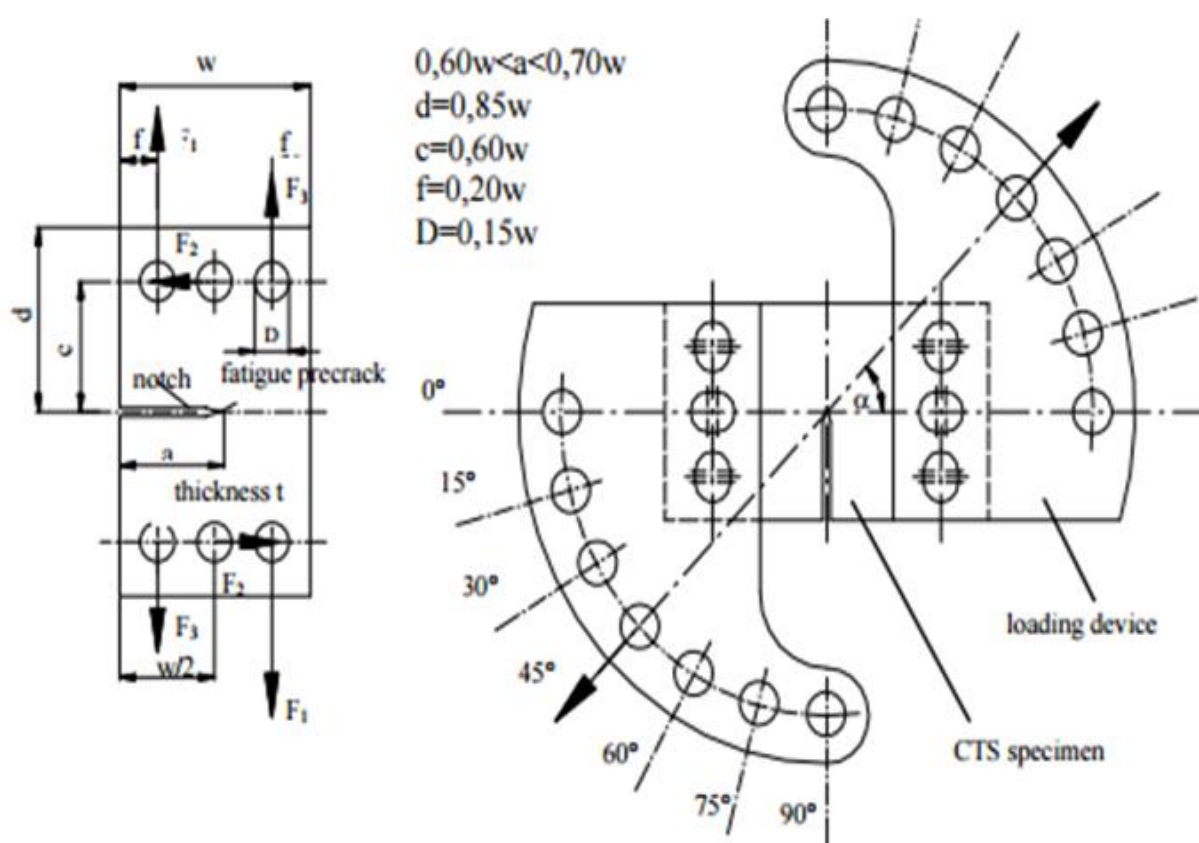
Figure 4. 7: CTS for mixed mode crack propagation analysis (Zafosnik, B., et al 2000).

The mixed mode propagation parameters of a 2.5mm crack under the influence of a static load of 15KN is adopted to replicate the experimental set up in the figure above. A modified version (by including an inclination angle of 0^0-90^0) of the block model containing defect type b (elliptical edge through crack as analysed in Appendix D2) is adopted for the validation of the fracture mechanics parameters (KI and KII). The calculated SIF on the right hand side of Figure 4.9 is compared to the experimental results on the left side of Figure 4.9. It can be observed that a variation of no more 9% is recorded, which is attributed to the fact that the experimental set up is not perfectly replicated in the FE model. The plot on the left in Figure, 4.9 at $\beta = 68^0$ both tensile and shear stresses are observed to be equal, up to this value, KI dominates KII and above it KII dominates KI. This is also the case in plot depicted on the right side of Figure 4.9, based on the fact that in the model of section 4.4.1, initial has an actual value of $\beta = 90^0$ as opposed to being $\beta = 90^0 0^0$, therefore, $\beta = 38^0$ for KI=KII can be calibrated to about 58^0.
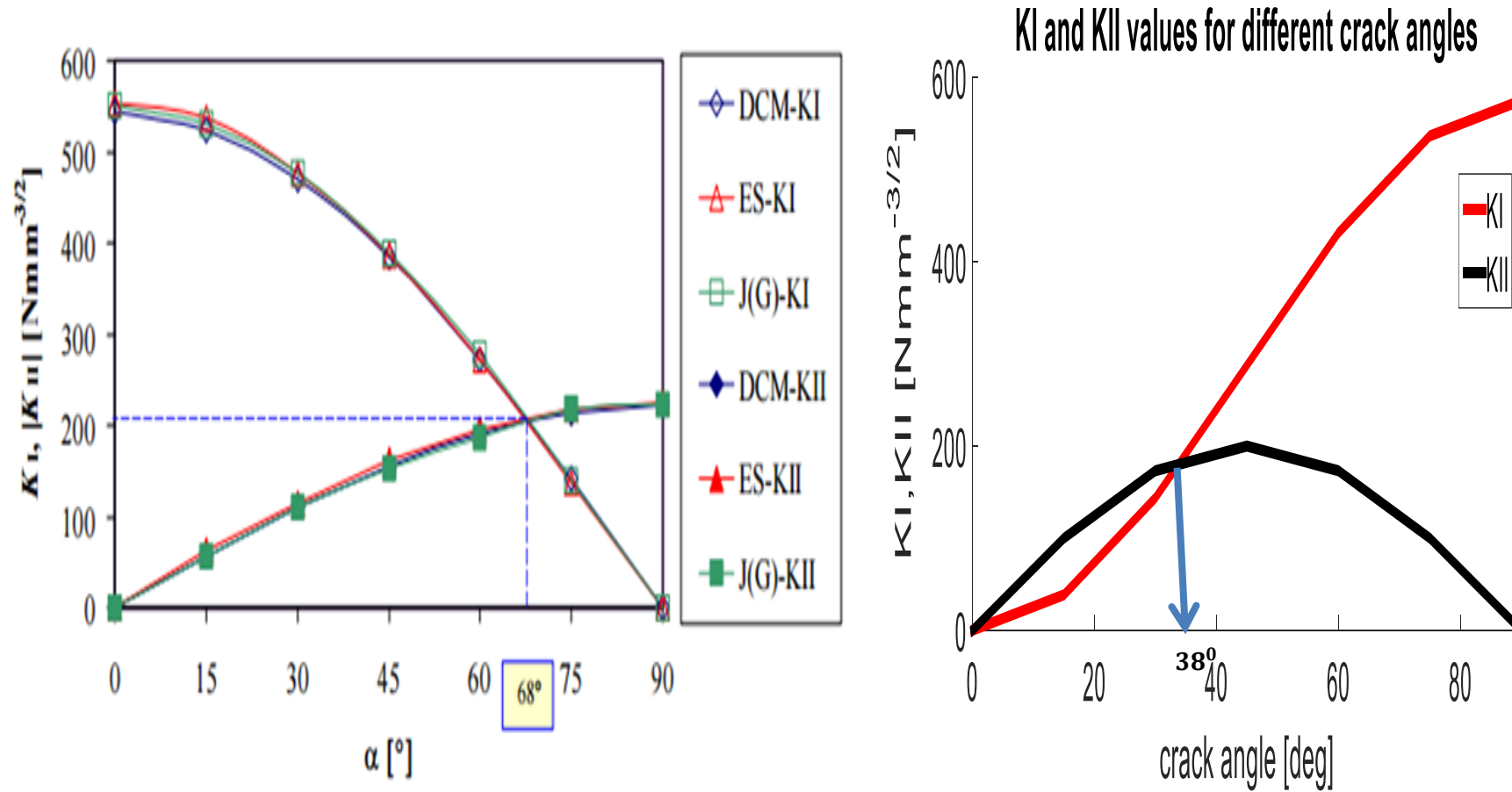
Figure 4.9: Comparison between experimentally and FE estimates of mode-I and mode-II SIF

4.4 Simulated FE (COMSOL) results and discussion

4.4.1 Fracture mechanics for tangent track data

The quasi-static simulation of wheel-rail interface is undertaken using the COMSOL Multiphysics software. The geometrical properties of the rail section (600m sleeper spacing) adopted in this research is from widely accepted publications but also considers the computational requirement and speed of the model. Solid mechanics boundary condition has been set according to practical understanding of track model. Also a user defined extra fine free tetrahedral mesh is used with maximum element size of 0.9 mm at the contact patch. Simulated results of the global contact model (for both tangent and curved track case studies) is benchmarked using VAMPIRE data (see data set 1 and 2). This section of the thesis compares the estimates of contact pressure (including contact area for tangent track) as a function of lateral displacement and curve radius respectively.

In the case of tangent track, the wheel is initially positioned at the centre line of railhead and an axle load of 80kN is applied on the boundaries of the wheel axle rod interface. It is assumed that the lateral displacement of the right wheel is changing from central position in steps of 1mm until 6.5mm (flange contact). Both VAMPIRE and COMSOL models theoretical understanding dictates that the normal force at the contact patch appreciates as lateral displacement increases (as the wheel moves towards flange contact), due to the influence of contact angle which corresponds to an increase in contact forces in both models.
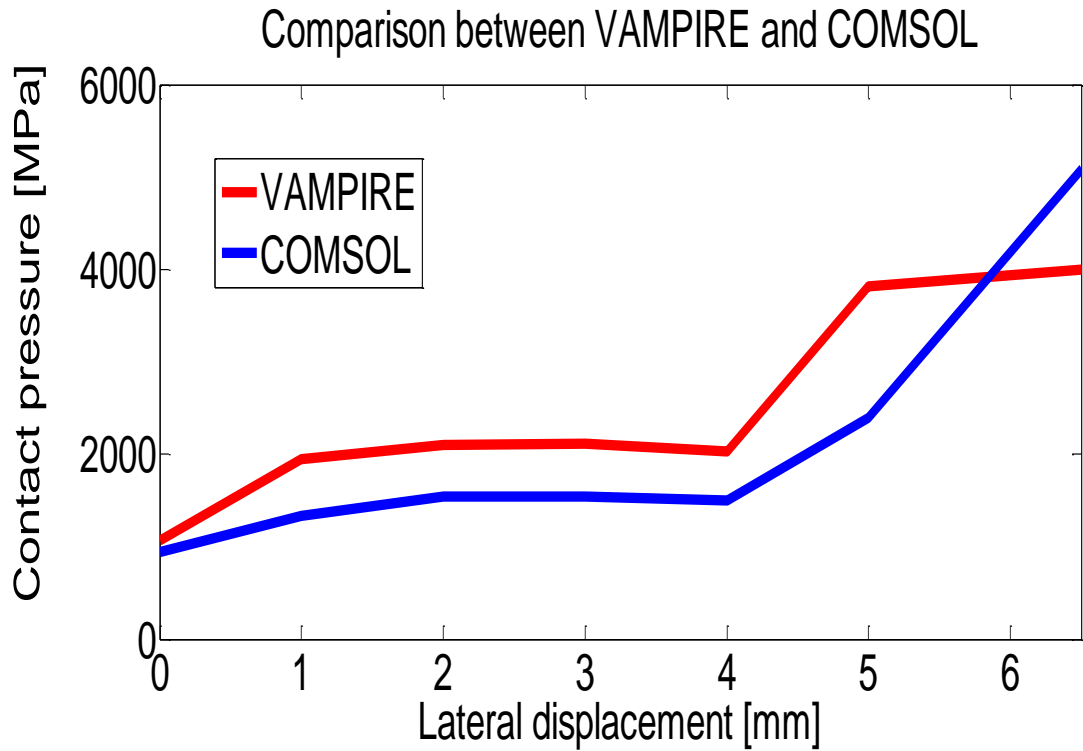
(a)
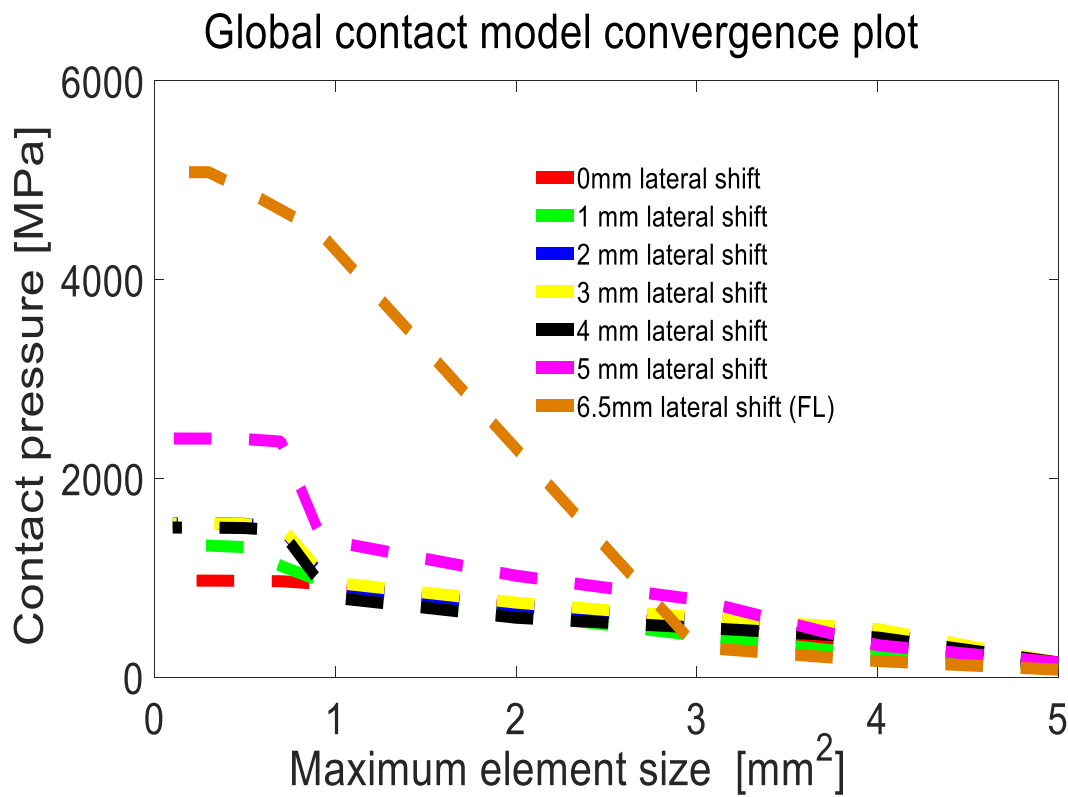
(b)

Figure 4. 10: Comparison between FE simulated contact pressure and VAMPIRE equivalent (a) and COMSOL convergence plot in (b).

The difference between simulated results in both models depicted in Figure 4.10a, is concluded to be a direct consequence of the assumptions made by these two software packages in the sense FE model assumes that the wheel-rail interface is of linear elastic material property, and also includes deformation in the contact patch. This increases the contact area and reduces the stresses compared to the VAMPIRE software, which assumes the wheel and rail material as being rigid and no deformation is included in the contact patch. As a result, the VAMPIRE contact pressure tends to be higher than the COMSOL data by about 12%. The simulated results show the influence of non-linear effects as the wheel approaches flange contact for lateral displacements greater than 5mm. The convergence analysis of the COMSOL FE simulated results as presented in Figure 4.10b confirms that more accurate contact pressure is attained with smaller size of node, thus a maximum mesh size of $1mm^2$ is used for all models presented in this chapter.
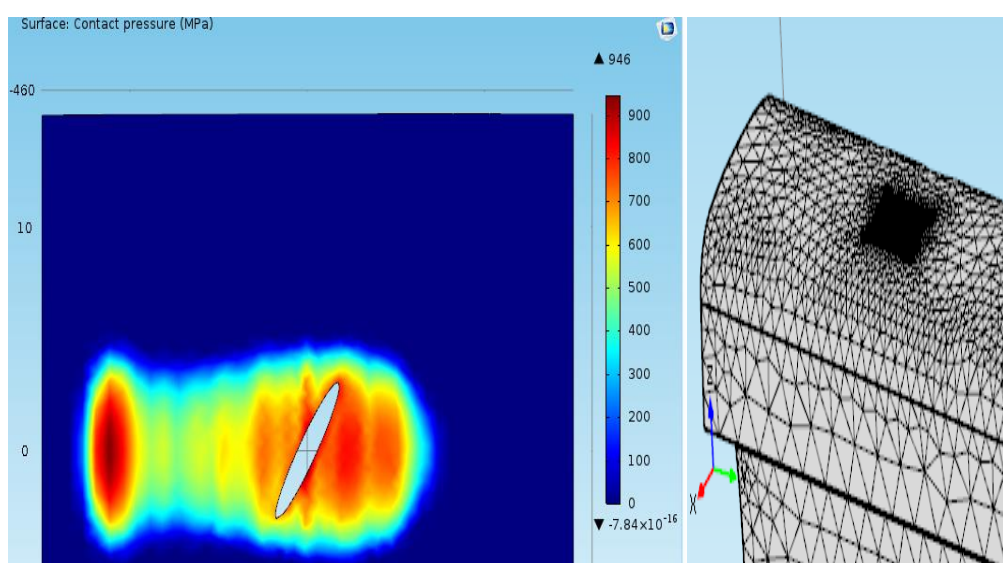
Figure 4. 11: Tangent track stress distribution for 0 lateral displacement of wheel (left), Predefined mesh maximum size of 1mm at the contact patch (right).

The simulated results of a critically grown cracks are presented in Figure 4.12, which indicate KI (see Figure 4.12a) has a maximum of value 125MPa at surface inclination angle $\beta=\frac{\pi}{3}$. An approximate peak value of 75MPa is observed for the case of $\beta=\frac{\pi}{2}$ and $\beta=\frac{\pi}{4}$. While the least amplitude of KI is observed to be at about 60MPa corresponds for $\beta=\frac{5\pi}{12}$.

In the case of KII Figure 4.12b suggests that $\beta=\frac{\pi}{4}$ and $\beta=\frac{\pi}{3}$ correspond to maximum values of 30MPa. As expected at peak axle load of 80KN, pure mode II loading case ($\beta=\frac{\pi}{4}$) is slightly higher than the latter inclination angle of $\beta=\frac{\pi}{3}$ by a few MPa. For $\beta=\frac{5\pi}{12}$, KII equal no more than 5MPa at maximum applied load of 80KN. As depicted in the graph, the minimum value of 0MPa is recorded for $\beta=\frac{\pi}{2}$ which is expected for pure mode-I fracture mechanism.

The equivalent SIF (Keqv) depicted in Figure 4.12c is observed to behave in a similar manner as the estimated KI. However, the maximum values are observed to be 100MPa, 55MPa, 50MPa, and 40MPa for $\beta=\frac{\pi}{3}$, $\beta=\frac{\pi}{4}$, $\beta=\frac{\pi}{2}$, and $\beta=\frac{5\pi}{12}$ respectively.

The estimated growth rate in Figure 4.12d is observed to behave in a similar manner to the estimated equivalent SIF. Simulated results suggest a maximum value of 12,000 nm/cycle corresponding to a surface orientation of $\beta=\frac{\pi}{3}$. The pure mode I and II coincident with $\beta=\frac{\pi}{2}$ and $\beta=\frac{\pi}{4}$ are observed to have a similar growth rate of 2000nm/cycle with a variation of no more than 1%. While the least growth rate of 1000nm/cycle is observed to correspond to a crack surface orientation of $\beta=\frac{5\pi}{12}$.

In Figure 4.12e, the growth path for $\beta=\frac{\pi}{2}$ is in agreement with the theoretical understanding of no branching angle is to be experienced by the crack (in pure mode–I loading case crack growth path is expected to be perpendicular to inclination angle). While a maximum branching angle of about 35^0 is observed in the case of $\beta=\frac{\pi}{4}$ (pure mode-II loading case) at maximum axle load of 80KN. The mixed mode loading case study of $\beta=\frac{5\pi}{12}$ and $\beta=\frac{\pi}{3}$ are observed to correspond to 22^0 and 10^0 at peak axle loads of 80KN, which suggest maximum branching angle to be synonymous to pure mode-II loading case.

The estimated number of loading cycles until failure is depicted in Figure 4.12f. These results are based on the assumption that the cyclic life of the rail is reached at crack length of 7mm which is assumed to be coincident with catastrophic fracture of the rail. Simulated results indicate no direct relationship between surface orientation and rail residual life. However, the

number of cycles until failure is inversely proportional to the estimated growth rate for the respective β in consideration. At peak axle load of 80kN the lowest residual life of the rail 8.74E3 cycles is observed to correspond to a surface inclination angle of β=$\frac{\pi}{3}$. For β=$\frac{\pi}{2}$ and β=$\frac{\pi}{4}$ (pure mode I and II loading case) a residual life of about 4.80E4 and 4.01E4 cycles is recorded respectively. And the highest residual life of 9.10E4 is observed to correspond to a surface inclination angle of β=$\frac{5\pi}{12}$. This behaviour in residual life estimate is logically expected such that crack configurations with fast growth rates will have less residual life and vice versa.

Figure 4. 82: Tangent track simulated Mode-I SIF (a), Mode-II SIF (b), Equivalent SIF (c), Growth rate (d), Growth angle (e), Number of cycles to failure Nf (f).

Furthermore, the fracture behaviour of a newly initiated crack (with maximum surface length of 1mm, crack mouth opening of 0.5mm, and a depth of 2mm) is analysed for tangent track contact data. The simulated KI in Figure 4.13a, indicate a maximum of value 100MPa for $\beta=\frac{\pi}{3}$. In the case of pure mode-I fracture mechanism ($\beta = \frac{\pi}{2}$) a maximum value of 75MPa is recorded for KI. It is also observed that in the case of $\beta = \frac{5\pi}{12}$ KI corresponds to 60MPa. And a minimum KI value of about 50MPa corresponds to $\beta=\frac{\pi}{4}$. The fact that the estimated KI of a newly initiated crack is observed to differ from that of a critically grown defect by approximately 25%, it is generally attributed to the complex interaction between contact of stress distribution and the size of the defect at each value of β. In the case of KII depicted in Figure 4.13b, it is observed that $\beta=\frac{\pi}{3}$ correspond to a maximum value of KII of about 18MPa (as opposed to expecting the maximum value of KII at $\beta=\frac{\pi}{4}$). In the case of pure mode-II loading ($\beta = \frac{\pi}{4}$) a maximum value of 15MPa is recorded, which is 16% below maximum KII for all β angles. A 5MPa KII is observed for the case of $\beta = \frac{5\pi}{12}$ at maximum applied axle load of 80KN. As depicted in the graph, the minimum value of 0MPa is recorded for $\beta=\frac{\pi}{2}$ which is expected for pure mode-I fracture mechanism.

The equivalent SIF Keqv depicted in Figure 4.13c is observed to behave in a similar manner as the estimated KI. It is observed that at peak axle load, 85MPa, 55MPa, 45MPa, and 35MPa is recorded for $\beta = \frac{\pi}{3}, \beta = \frac{\pi}{2}, \beta = \frac{5\pi}{12}, \beta = \frac{\pi}{4}$ respectively.

The estimated growth rate in Figure 4.13d suggests that the maximum value of 7900 nm/cycle corresponds to a surface orientation of $\beta = \frac{\pi}{3}$. The pure mode-I is observed to have a growth rate of 2000nm/cycle for $\beta = \frac{\pi}{2}$. While $\beta = \frac{5\pi}{12}$ and $\beta = \frac{\pi}{4}$ propagate at no more than 1000nm/cycle and 500nm/cycle respectively. In Figure 4.13e, the growth path for a newly initiated crack is observed to behave in a similar manner in relation to β as with the case of a critically grown crack except for the difference in amplitude. At maximum applied axle load of

80KN, a branching angle of 25^0, 15^0, 10^0, and 0^0 correspond to β values of $\frac{\pi}{4}, \frac{\pi}{3}, \frac{5\pi}{12}$, and $\frac{\pi}{2}$ respectively.

The rail residual life is depicted in Figure 4.12f, shows that at peak axle load of 80kN, for β= $\frac{\pi}{3}, \frac{\pi}{2}, \frac{\pi}{4}$, and $\frac{5\pi}{12}$ correspond to 5.9E1, 3.2E2, 2.7E2, and 6.1E2.The simulated results indicate similar relationship with the case of a critically grown crack, such that highest and least residual life is for 75 and 60 degree inclination angle.

.

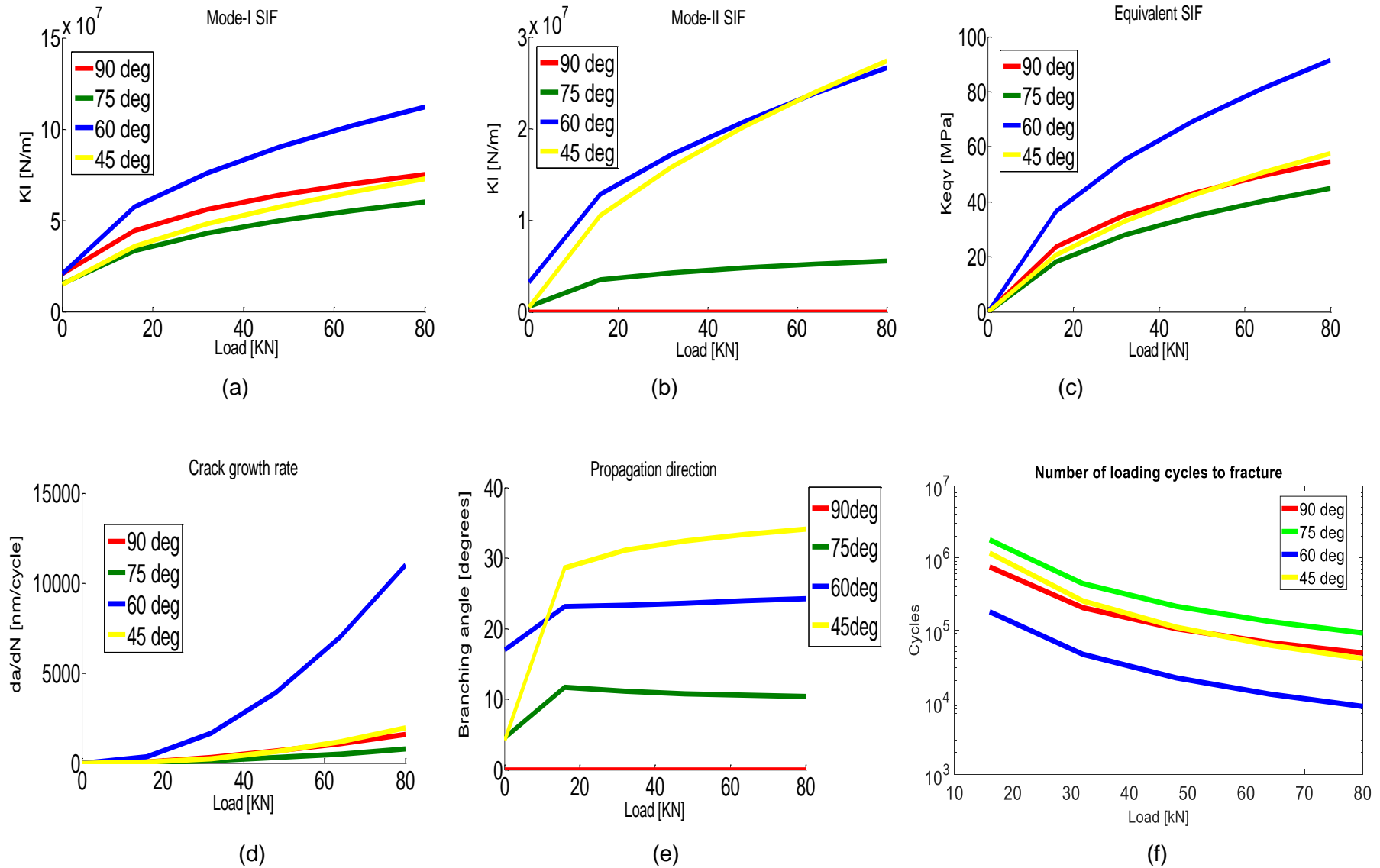$\frac{\pi}{4}, \frac{\pi}{3}, \frac{5\pi}{12}$, and $\frac{\pi}{2}$

Figure 4. 93: Tangent track simulated Mode-I SIF (a), Mode-II SIF (b), Equivalent SIF (c), Growth rate (d), Growth angle (e), Number of cycles to failure Nf (f).

4.4.2 Fracture mechanics for curved track data

In the case of curved track section, theoretical understanding dictates that for a right hand curve, the low rail of a canted track is bound to experience less pressure compared to high rail counterpart. In this thesis, a 40mm canted right hand curve with radius ranging from 200m – 1600m is adopted for the validation contact pressure exerted on the low rail of the global contact model (under an axle load of 80kN). The FE model is designed to have the same solid mechanics setting as described above in section 4.4.1, and a good agreement is recorded compared to VAMPIRE benchmark as depicted in Figure 4.14 below.
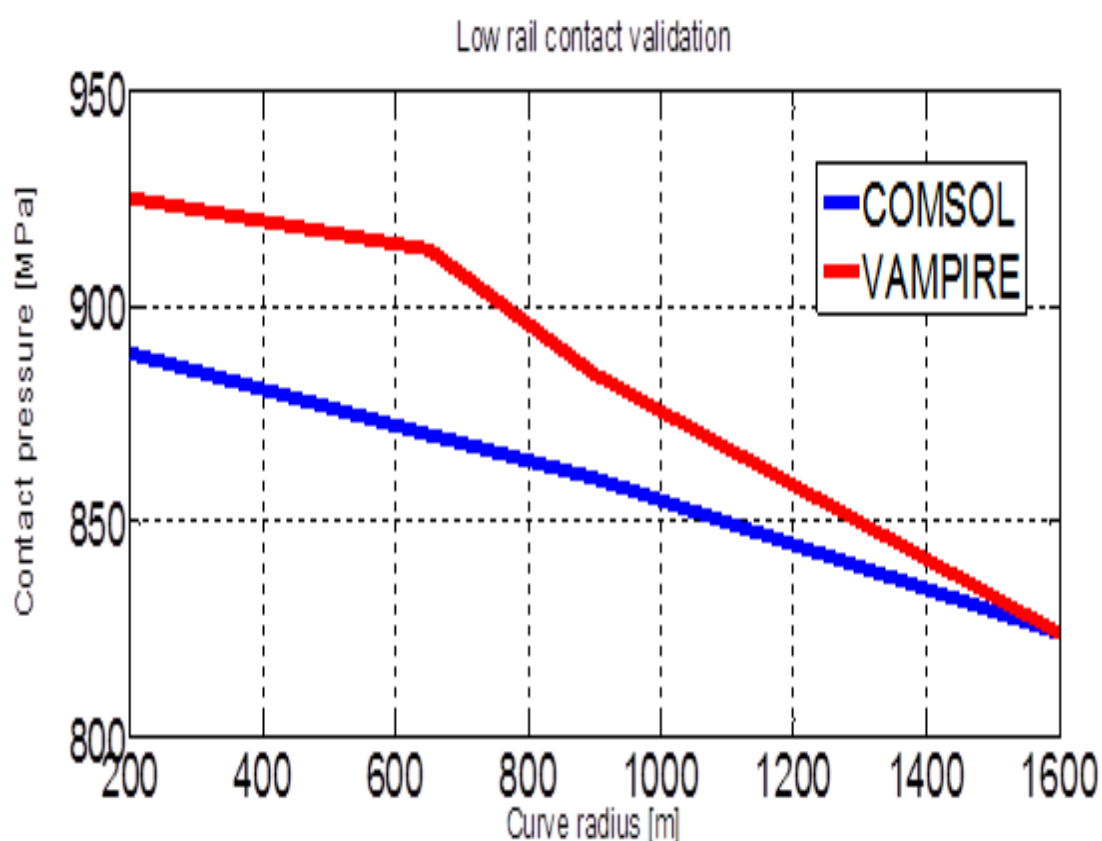
Figure 4. 110: Validation of contact pressure for low rail component of global contact model.

The difference in assumption between the two contacting bodies is a major source of variation in simulated results as depicted in Figure 4.14. This is however noticed to cause a variation of only about 3% between FE and VAMPIRE results because of the expanded contact calibration parameters adopted in this case (yaw angle, and creep forces included in curved track analysis). A similar agreement has also been established even in the case of high rail

simulated contact pressure, which includes the non-linear effect of flange contact specifically in the 200m curve radius case study as shown in Figure 4.15 below. The simulated high rail contact pressure also confirms higher amplitudes of stress existing on the high rail compared to that of the low rail section. It is worth noting that in the case of flange contact (200m curve radius) Figure 4.15 confirms the non-linearity observed in the results, such that FE model predicts higher stress amplitude compared to VAMPIRE model.
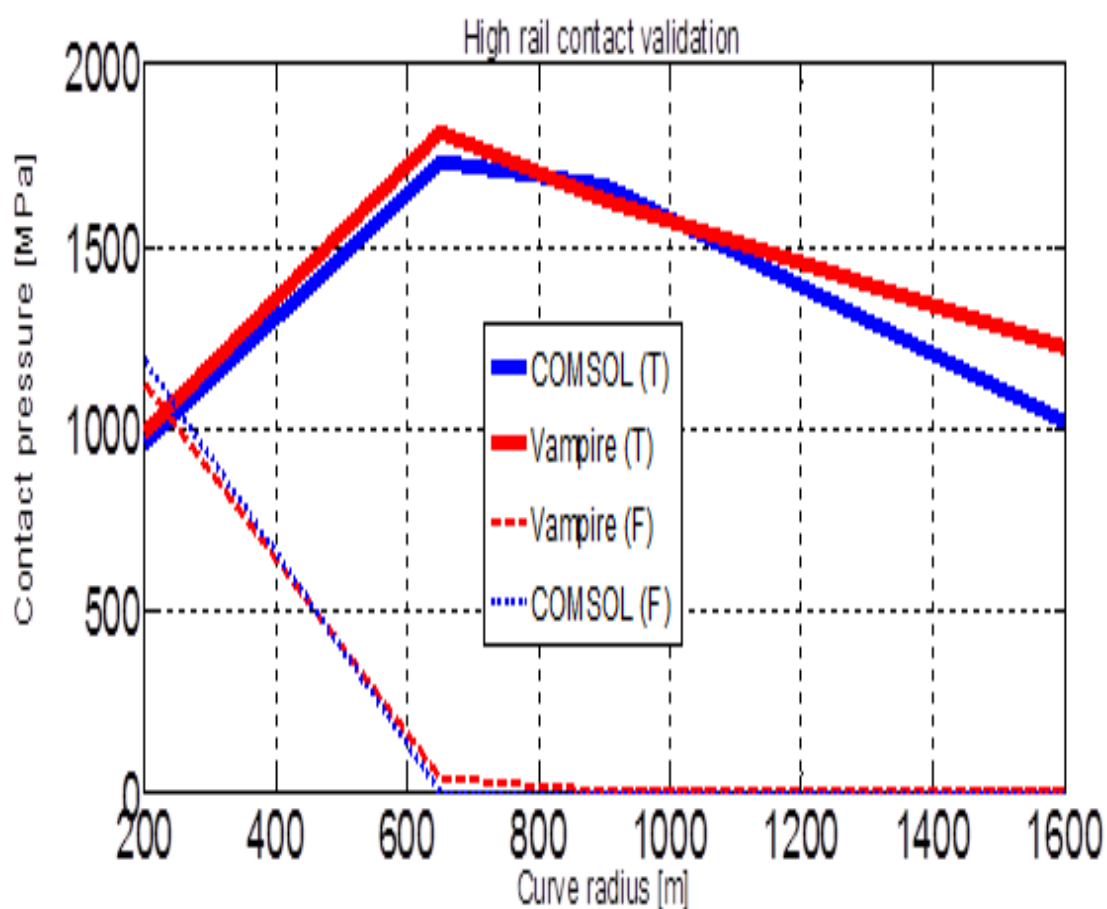
Figure 4. 15: Validation of contact pressure for high rail component of global contact model.

This section provides details of fracture behaviour of an elliptical equivalent representation of rail crack under bi-axial stresses generated on a curved track section. The contact data of the global contact model have been validated. In a similar manner to section 4.4.1, the simulated results in this section are aimed at supporting predictive rail maintenance decision making. As such these results are case studies of typical passenger axle load (80KN), surface orientation of $\beta = \frac{\pi}{3}$, with a sensitivity analysis of track curve radius. Moreover, the effect of defect size

(ranging from newly initiated to critically grown defect) and location is also highlighted in order to increase flexibility of the model towards incorporating NDT with fracture mechanics models. The results presented in Figure in 4.16, are for a critically grown surface crack with length 5mm, depth of 4mm and crack mouth opening of 1mm is adopted in this section. The simulated KI in Figure 4.16a, indicates a maximum value of 180MPa for a track with 200m curved radius. It is further observed that KI (including lateral displacement) is inversely proportional to curve radius with 140MPa, 135MPa, and 110MPa for 650m, 900m, and 1600m respectively. In Figure 4.16b KII is not related to track as observed in mode-I fracture mechanism. It is however observed that tight and wide curved sections (200m and 1600m) have higher amplitudes of about 80MPa and 50MPa respectively. While the moderate curved track (650m and 900m radius) exhibit relatively lower amplitudes of KII in the range of 10MPa. This relationship between KII and curved radius is observed to be governed by the creep forces, which are high for 200m, 1600m and considerably lower for 650m and 900m curve radius. The equivalent SIF Keqv depicted in Figure 4.16c is observed to behave in a similar manner as the estimated KI. At peak axle load, Keqv is observed to be 100MPa, 80MPa, 80MPa, and 60MPa for 200m, 650m, 900m, 1600m curve radius respectively. The estimated propagation rate in Figure 4.16d suggests a similar inverse proportionality between the growth rate and curve radius as observed for KI and Keqv. At peak axle load, a growth rate of 180000nm/cycle, 7500nm/cycle, 7200nm/cycle, and 5000nm/cycle is recorded for 200m, 650m, 900m, and 1600m respectively. The growth direction depicted in Figure 4.16e is observed to respond in a similar manner to the relationship between KII and curved radius. A maximum branching angle of 50^0, 30^0, 11^0, and 9^0 is recorded for 1600m, 200m, 900m, and 650m respectively. The estimated number of loading cycles until failure as depicted in Figure 4.16f, confirms the inversely proportional relationship between residual life and curve radius due to increased stresses at the wheel rail interface as curve radius decreases. At peak axle load, 8.1E7, 1.03E9, 1.51E8, and 1.23E8 cycles are observed for 200m, 650m, 900m, 1600m curve radius respectively.
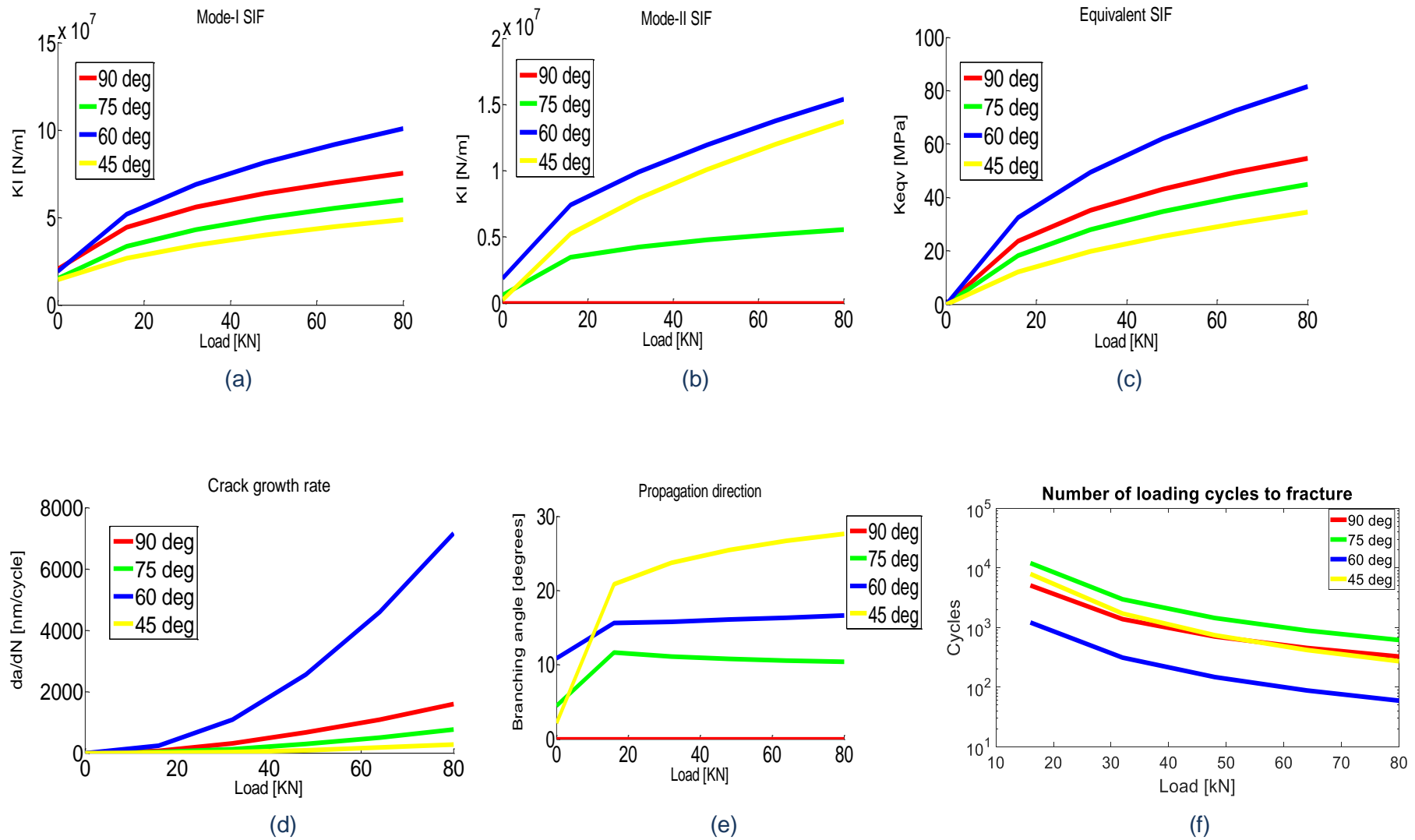
Figure 4. 116: Curved track simulated Mode-I SIF (a), Mode-II SIF (b), Equivalent SIF (c), Growth rate (d), Growth angle (e), Number of cycles until failure Nf (f).

The investigation is expanded to include the fracture behaviour of newly initiated cracks on curved track. The same loading, curve radius parameters are used for this analysis, while the crack geometry is modified to a surface length of 1mm, opening of 0.5mm, and depth of 1mm. The simulated KI in Figure 4.17a, indicates a maximum value of 80MPa for a track with 200m curved radius. It is further observed that a newly initiated crack only differs from that of a critically grown crack by 44% reduction in amplitude. In Figure 4.17b KII is not directly related to track curve radius as observed in Mode-I fracture mechanism. It is however observed that tight and wide curved track sections (200m and 1600m) have higher amplitudes of about 40MPa and 25MPa respectively. While the moderate curved track (650m and 900m radius) exhibit relatively lower amplitudes of KII in the range of about 5MPa. As with the case of KI, the relationship between KII and curved radius is observed to be governed by the creep forces, which are high for 200m, 1600m and considerably lower for 650m and 900m curve radius. The equivalent SIF Keqv depicted in Figure 4.17c is observed to behave in a similar manner as the estimated KI. At peak axle load, Keqv is observed to have an amplitude of 45MPa, 35MPa, 35MPa, and 30MPa for 200m, 650m, 900m, 1600m curve radius respectively. The estimated propagation rate in Figure 4.17d suggests a similar inverse proportionality between the growth rate and curve radius as observed for KI and Keqv. At peak axle load, a growth rate of 800nm/cycle, 350nm/cycle, 320nm/cycle, and 5000nm/cycle is recorded for 200m, 650m, 900m, and 1600m respectively. The growth path of a newly initiated crack as depicted in Figure 4.17e is observed to respond in a similar manner to the relationship between KII and curved radius (also similar in magnitude and direction to the case of a critically grown crack). A maximum branching angle of 50^0, 30^0, 11^0, and 9^0 is recorded for 1600m, 200m, 900m, and 650m respectively. The estimated number of loading cycles until failure for a newly initiated crack to reach a critical length of 118mm as depicted in Figure 4.17f, confirms a similar inverse proportionality between residual life and curve radius (as) the case of critically grown crack. This observation is due to increased stresses at the wheel rail interface as curve radius decreases. It is however logical that for newly initiated cracks, the relatively shorter lengths of the crack will require longer cycles of loading to attain critical final length. At peak axle load of

80KN, the maximum number of loading cycles is observed to correspond to amplitudes of 5.54E5, 7.0E6, 1.02E6, and 7.00E6 cycles for 200m, 650m, 900m, 1600m curve radius respectively.
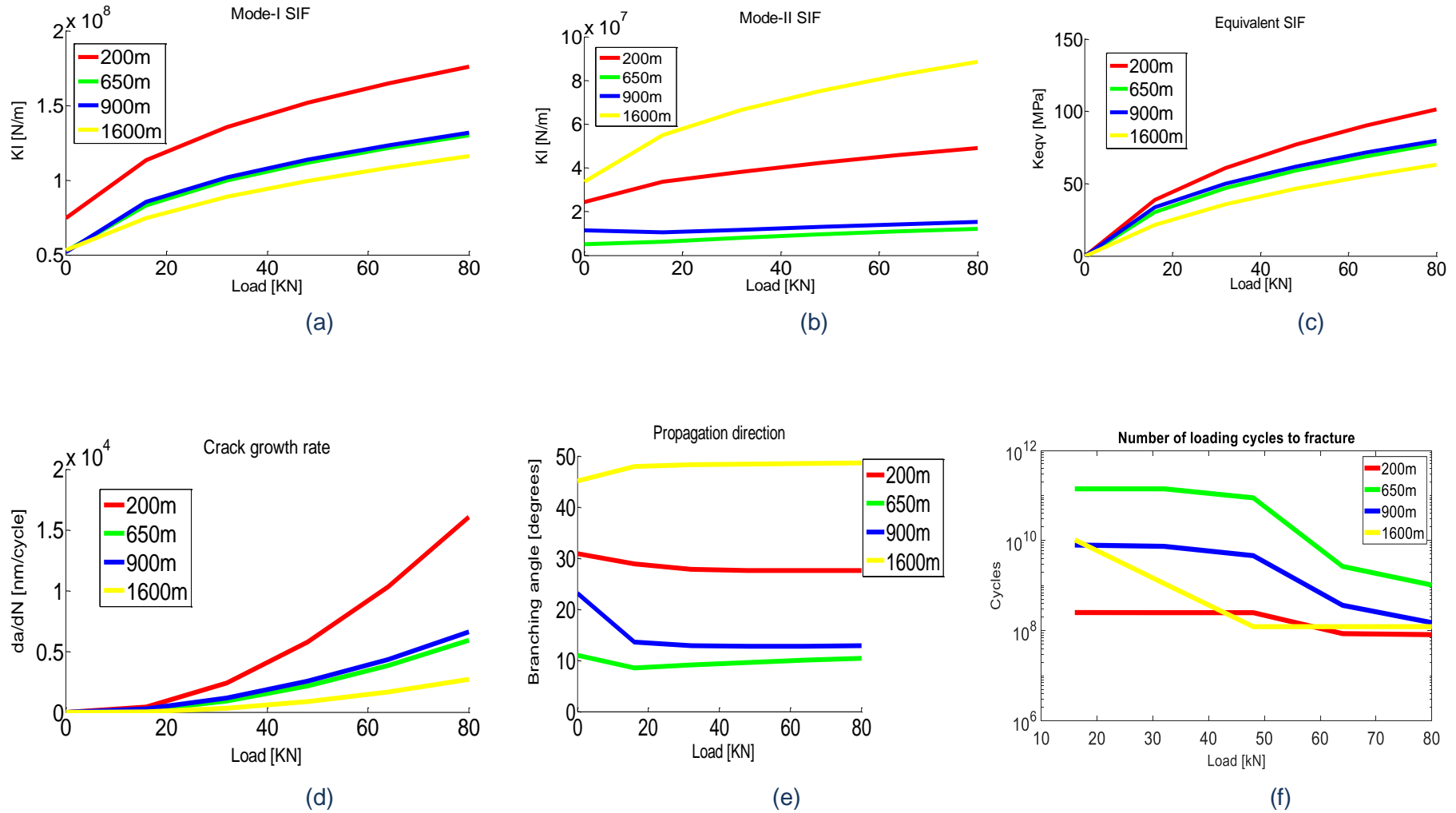
Figure 4. 127: Curved track simulated Mode-I SIF (a), Mode-II SIF (b), Equivalent SIF (c), Growth rate (d), Growth angle (e), Number of cycles until failure Nf (f).

4.4.3 Location sensitivity analysis for curved track data

The influence of changing the position of crack centroid relative to an origin that corresponds to the location of maximum contact pressure at the wheel rail interface is investigated in this section of the thesis. The simulated results of a newly initiated crack having a surface orientation of $\beta = \frac{\pi}{3}$, assumed to be under bi-axial loading, with the crack centroid further displaced 1mm in both lateral and longitudinal directions is presented in Figure 4.18. Estimated KI in Figure 4.18a is observed to have a maximum value of 80MPa at 1600m curve radius while about 40MPa is recorded for both 200m and 650m curve radius. The least value of KI is observed in the case of 900m curve radius with amplitude in the range of 10MPa. In Figure 4.18b, KII is observed to have the same trend as in Figure 4.18b. It is also noted that the tangential stresses at the crack front is highest under the influence of gravity only for 1600m curve radius (prior to the introduction of axle load). Thus resulting to an initial value of KII of about 8.5MPa, this decreases to about 2.5MPa at 10KN axle load, but proportionally increases from about 20KN to about 7MPa at peak axle load of about 80KN. In the case 900m curve radius, tangential stress at the crack tip linearly increases for 0-20KN, which yields KII of up to 1.5MPa, after which it decays to a few MPa as axle load increases. A similar behaviour in KII for 650m curve radius is observed, however the amplitude of KII at the initial 0-20KN linear phase is observed to be almost 5MPa. The value of KII decays to a few MPa for axle load ranging from 20KN-65KN, above this range, KII proportionally increases to about 2.5MPa at peak axle load. For curve radius of 200m, initial value of KII=1.2MPa is maintained until an axle load of about 20KN is attained after which KII linearly increases to a maximum value of 5MPa. The overall behaviour of KII for all the curve track data discussed is concluded to be a complex interaction of 3D contact parameters, shear stresses at crack front and crack centroid relative position within the contact patch.

The equivalent SIF (Keqv) depicted in Figure 4.18c is observed to behave in a similar manner as the estimated KI, with maximum value of 40MPa, 35MPa, 30MPa, 10MPa for 16000m, 650m, 200m, 900m curve radiuses respectively.

The estimated propagation rate in Figure 4.18d is observed to have a similar trend as KI and Keqv, such that a growth rate of 500nm/cycle, 300nm/cycle, 200nm/cycle, and 6nm/cycle is recorded for 1600m, 200m, 650m, and 900m curve radius respectively.

The growth direction depicted in Figure 4.18e is noted to have a maximum branching angle of 14^0, 10^0, 5^0, and 2^0 is recorded for 200m, 1600m, 650m, and 900m respectively.

The estimated rail residual life depicted in Figure 4.18f, is governed by growth rate estimates, such that 200m, 650m, 1600m curve radius correspond to 2.56E4, 1.1E4, 9.26E3 cycles while 900m (least growth rate) curve radius corresponds to 1.96E6 cycles.
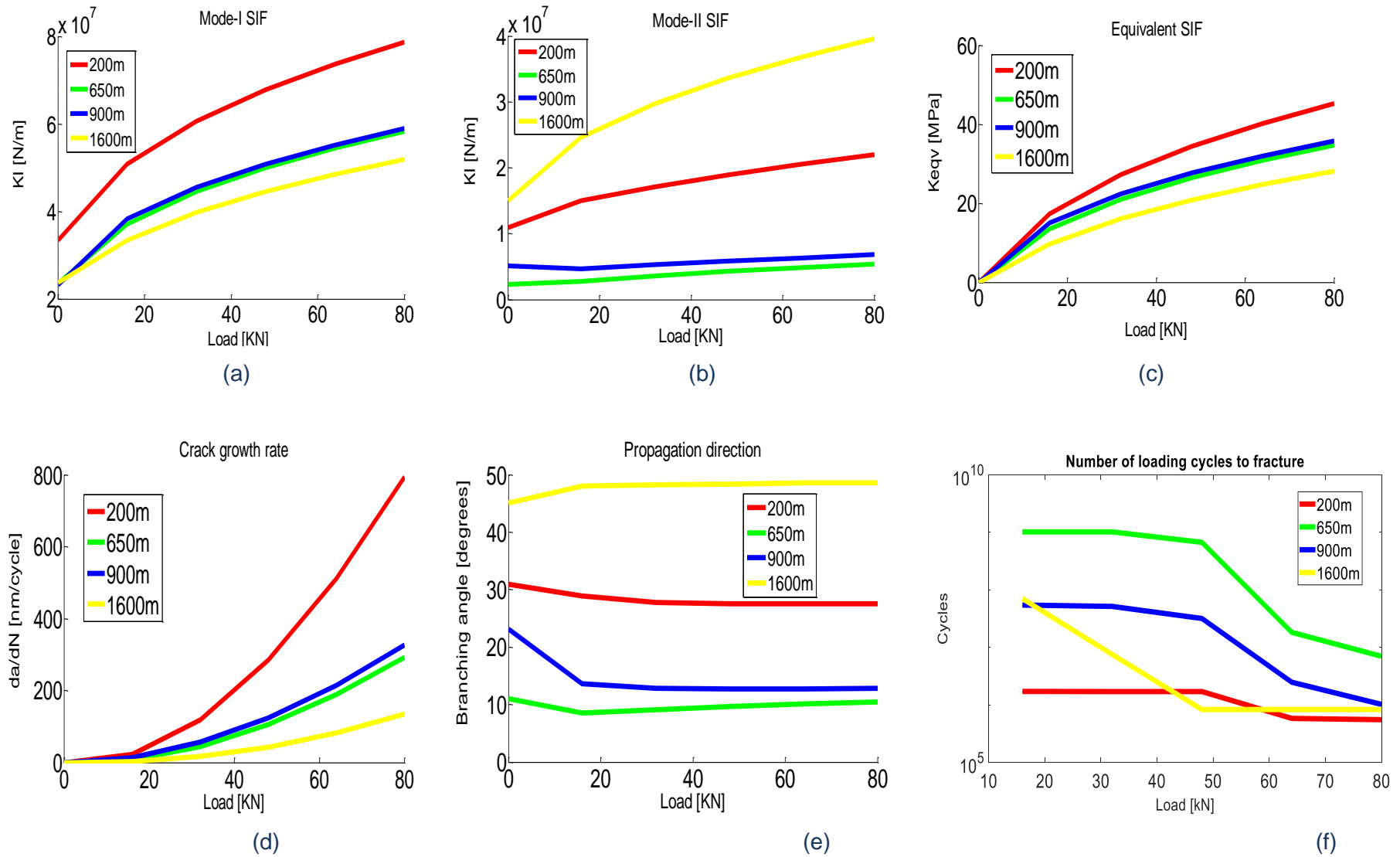
Figure 4. 138: Defect displaced (x-1,y-1) Mode-I SIF (a), Mode-II SIF (b), Equivalent SIF (c), Growth rate (d), Growth angle (e), Number of cycles until failure Nf (f).

In general it can be said that the impact of displacing crack centroid from the origin that corresponds the peak contact stress (i.e. when wheel is direction over the crack being investigated) results in a decrease in fracture mechanics variables (namely; KI, KII, Keqv, da/dN, and θ) which corresponds to an increase in Nf. This is mainly because of the reduction in tangential stresses as crack tip moves away from high stress region of the contact patch as shown in Figure 4.19 below.
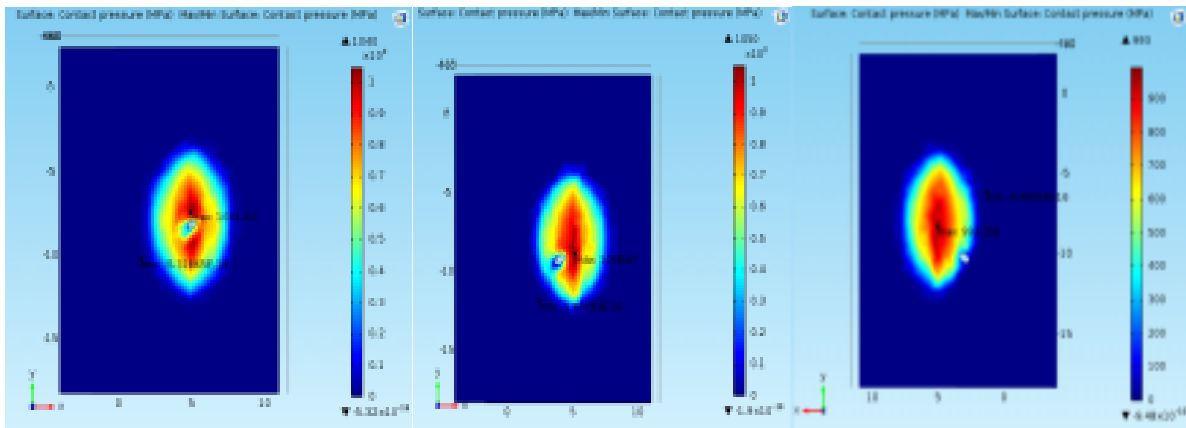
(a)
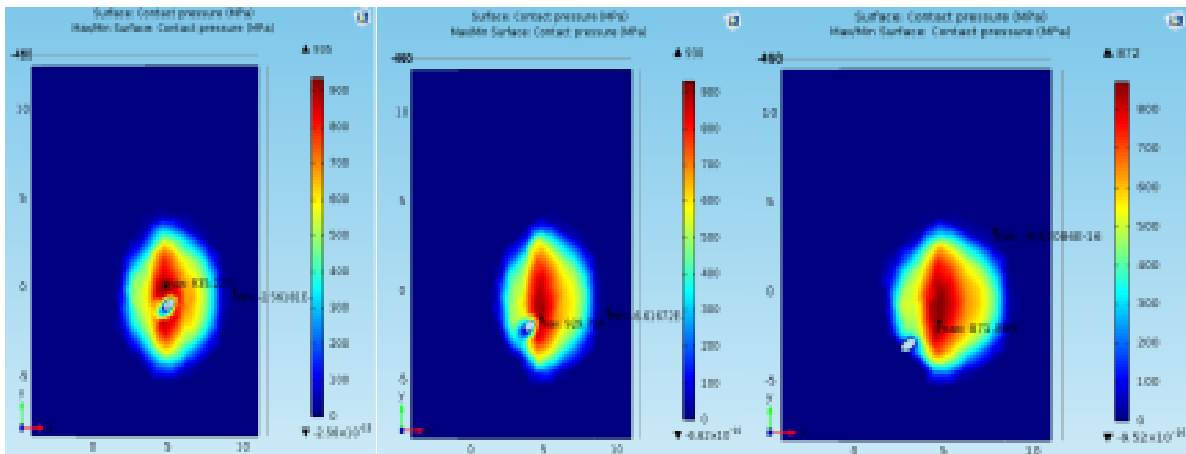
(b)
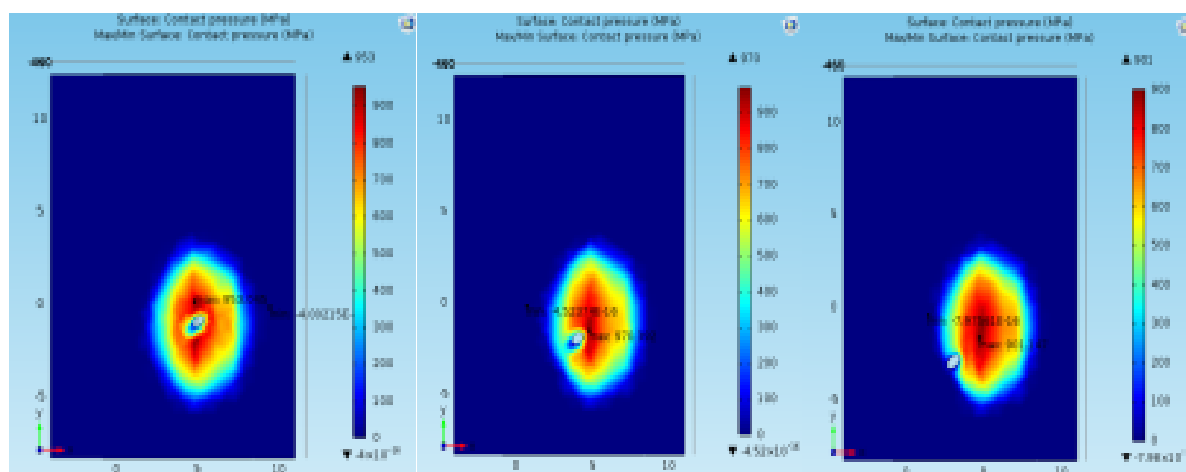
(c)
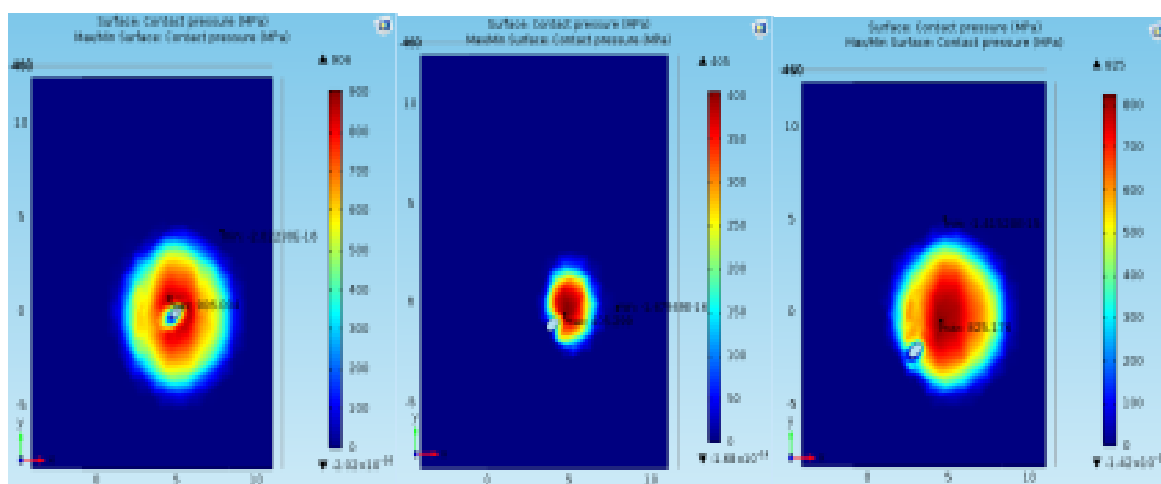
(d)

Figure 4. 149: Shows influence of displacing crack centroid (relative to peak contact stress mm-2mm for track curve radius 200m (a), 650m (b), 900m (c), 1500m (d).

The simulated results depicted in Figure 4.19 are for defect centroid displacement of 2mm in both lateral and longitudinal directions. Estimated KI in Figure 4.20a (with no obvious relationship to curve radius) is observed to have a maximum value of 25MPa, 20MPa, 19MPa, 15MPa for 650m, 200m, 900m, and 1600m curve radius.

In the case of KII, Figure 4.20b depicts an inversely proportional relationship between KII and curve radius with peak values of 15MPa, 11MPa, 10Mpa and 1MPa for 200m, 650m, 900m, and 1600m curve radius respectively.

The equivalent SIF (Keqv) depicted in Figure 4.20c is observed to behave in a similar manner as the estimated KI, with maximum value of 25MPa, 15MPa, 15MPa, 10MPa for 650m, 200m, 900m, and 1600m curve radius respectively.

The estimated propagation rate in Figure 4.20d is observed to have a similar trend as KI and Keqv. At peak axle load of 80KN, a growth rate of 50nm/cycle, 10nm/cycle, 10nm/cycle, and 1nm/cycle is recorded for 650m, 200m, 900m, and 1600m curve radius respectively.

The growth direction depicted in Figure 4.20e is noted to have a similar behaviour with KII, with maximum branching angle (at peak load of 80KN) of 50^0, 40^0, 38^0, and 8^0 recorded for 200m, 650m, 900m, 1600m respectively.

The estimated rail residual life depicted in Figure 4.20f, is governed by growth rate estimates, such that 200m, 650m, 1600m curve radius correspond to 2.49E5, 7.23E4, 1.53E6 cycles while 900m (least growth rate) curve radius corresponds to 2.50E5 cycles.
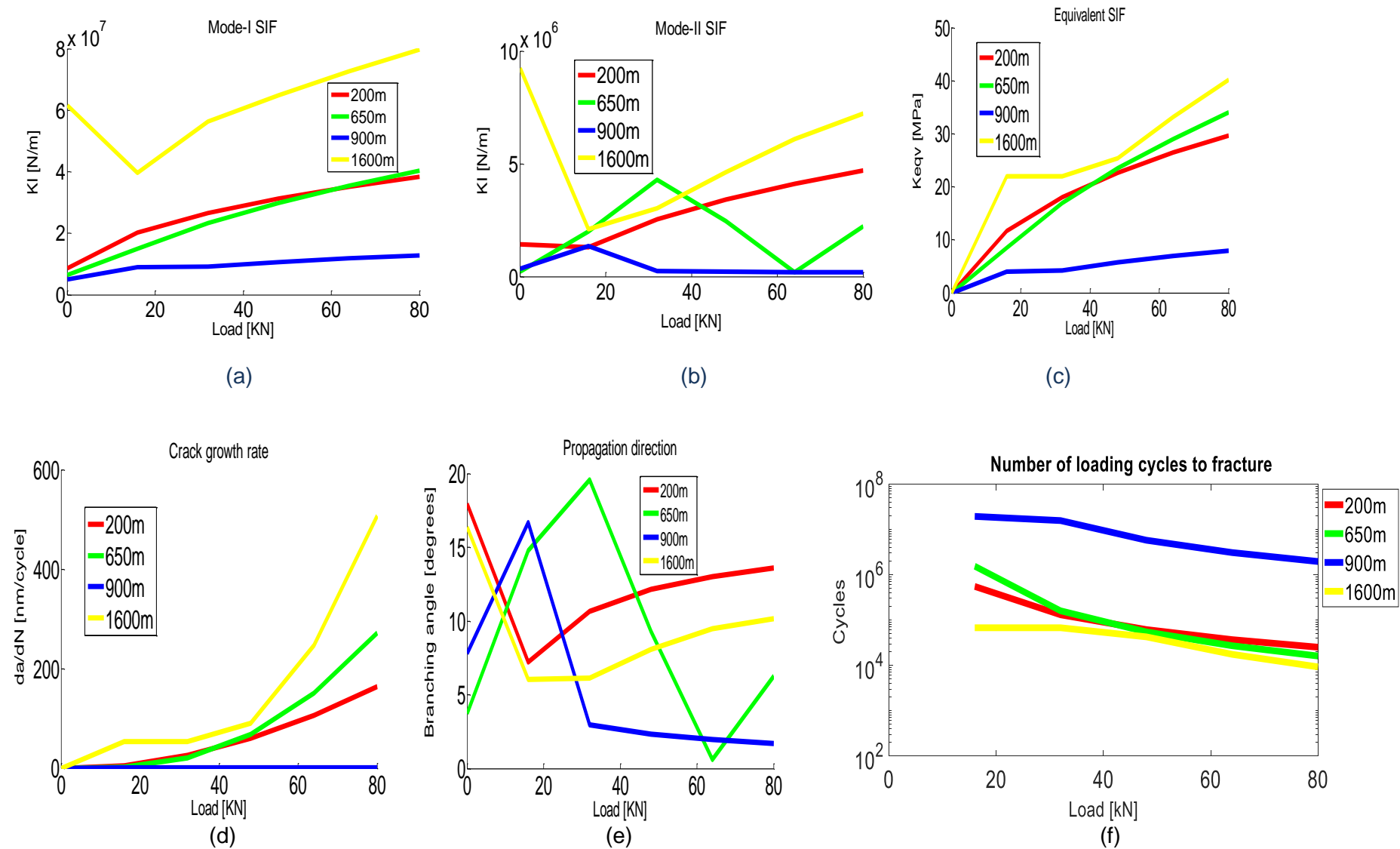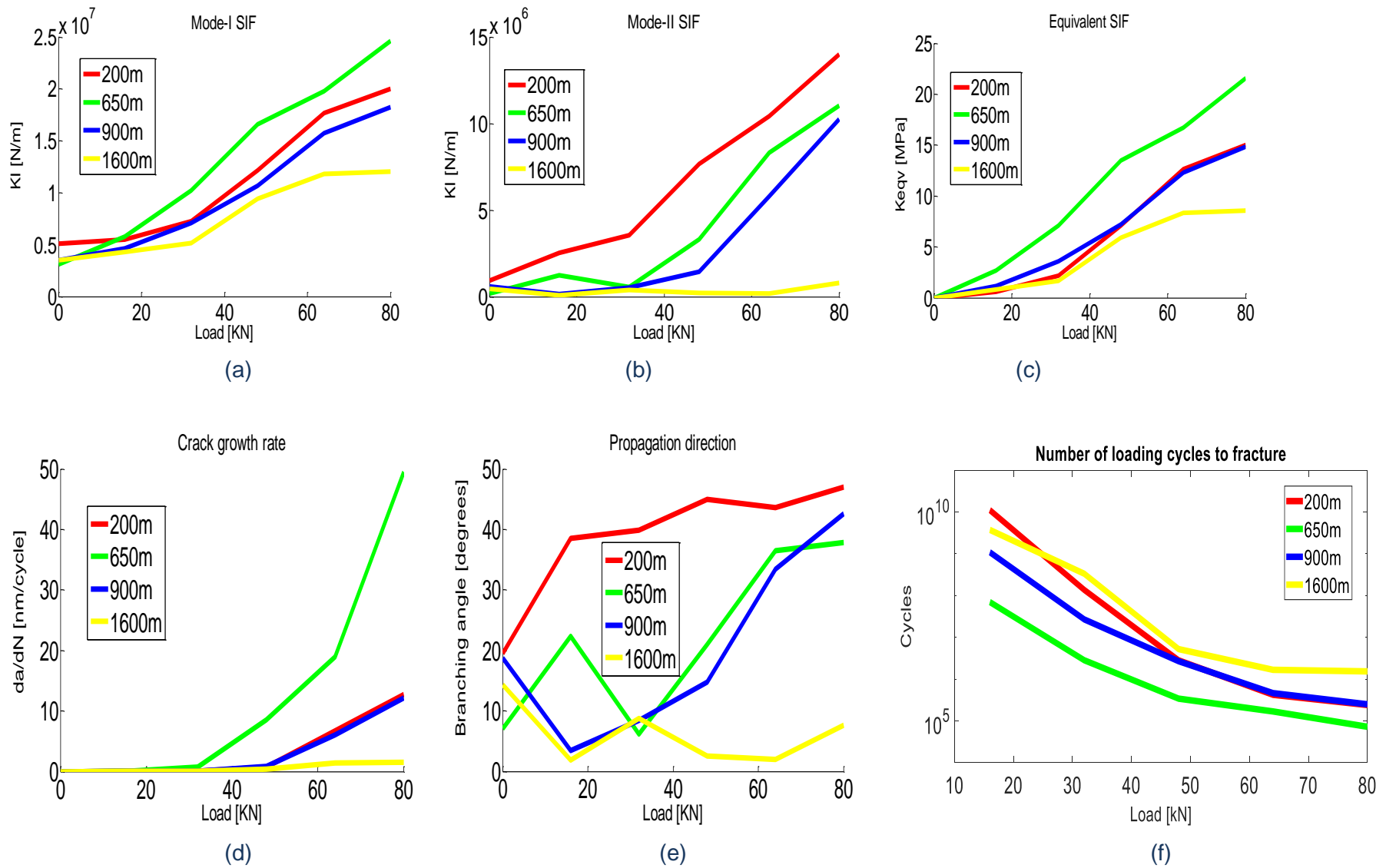
Figure 4. 2015: Defect displaced (x-2,y-2) Mode-I SIF (a), Mode-II SIF (b), Equivalent SIF (c), Growth rate (d), Growth angle (e), Number of cycles until failure Nf (f).Figure

4.5 Conclusion

In this study, the behaviour of crack growth on rail surface has been investigated using the FE analysis method for new profiles. Considerations for worn wheel and rail shapes which are likely to increase conformity of contact (thereby reducing the contact pressure and its distribution in the contact patch) should be investigated as future works. The fracture parameters are estimated from stresses generated by the global contact model, while the local fracture mechanics model is utilised for determining the growth rate, direction of propagation and residual life of the rail. For wheel-rail contact applications, Mode-I SIF are observed to be generally higher Mode-II SIF. This is in agreement with the theoretical understanding of KI being most vigour fracture mechanism compared to shear and torsional fracture mechanisms. It is also observed that KI governs the behaviour of crack propagation while KII dictates the growth path behaviour. Maximum growth rate for a defect is observed to be approximately 1/10000 of the estimated KI. The crack direction is observed to vary from 20° to 47° depending on the initial depth of the crack. The number of cycles until failure for crack is observed to be inversely proportional to growth rate, which confirms a faster propagating crack will fail faster. In general, the F.E. analysis showed that the behaviour of the crack, generated on rail surface varies by the load condition (magnitude and position) and crack length (surface and mouth opening). A significant amount literature and field tests suggest that surface crack translates to increases in depth to a certain limit where a considerable decrease is observed For both global track model and local fracture models it is believed that the variation in the prediction accuracy vehicle dynamic effects due to track irregularities, and limitation of model to only consider quasi-static vehicle behaviour over idealised track.

Chapter 5 Application of the proposed technology

5.1 Introduction

The implication of optimal maintenance strategy for rail infrastructure cannot be over emphasised in the rail industry. It is because of not only the associated costs but also health and safety risks that more intelligent and self-monitoring (optimised) maintenance systems are paramount to rail industry. For a combination of the limitations related to optimisation of maintenance strategy and increased global patronage/dependence on rail services, it is now more than ever considered increasingly important to intelligently detect damage, predict fracture behaviour, and optimise maintenance requirements in the rail industry. For example, numerical methods for accurate estimation of grinding required to reduce the propensity of RCF to an acceptable level is well understood. However, the practical tediousness associated with controlling the accuracy and also the depth of material removed to a resolution of 0.1 mm remains a challenge. While new vehicles may influence the rate of RCF initiation and growth in rails, either because they themselves are more or less damaging than those already on the route, or because they displace vehicles that are more or less damaging, a few assessments on the possibilities of improving the maintenance technology suggests including curve radii and traction matrix to remotely, in real time, incorporate multiple maintenance tasks within the same technology (such as detection, prediction, and decision making). Such intelligent, remote, and incorporated models are generally considered to be a function of big data. Thus detailed information with regards to loading, track data, vehicle characteristics, environmental conditions etc., are essential for the overall efficiency of this approach of rail maintenance. In this section of the thesis, a fully incorporated model is introduced. This automatic detection and prediction of rail crack, as discussed in section 5.2, is achieved by using the image processing based detection of RCF damage in rails, in collaboration with models for predicting the behaviour of the detected defects (governed by the mathematical model elaborated in section 3.3 and 4.3). In this section of the PhD thesis, this technology is demonstrated to be at an initial technology readiness level with the aid of MATLAB image processing toolbox in

addition to stress distribution look up matrix (as detailed in Appendix E1 and Appendix E2). Furthermore, the novelty of utilising actual defect boundary in comparison with elliptical equivalent representation of cracks is presented in section 5.3. The enablers and blockers of the proposed approach are explored in section 5.4 and the conclusion of the chapter is highlighted in section 5.5.

5.2 Application of the proposed technology

This section demonstrates the application of the proposed technology that links defect detection with prediction of RCF damage on rails. Three main indicators are adopted for these demonstrations which include growth rate, branching direction, and residual rail life. To avoid the computationally expensive requirement of using the global contact model, a simple block geometry is utilised for generating the stresses in response to 80kN bi-axial loading. The FA-H1 (heavily damaged rail sample) image of data set 2 is processed in accordance with the algorithm presented in section 3.3, and a random sample of the detected defects in Table 5.1 below are selected for this demonstration. The tabulated data summarises the maximum surface length, orientation, elliptical equivalent major axis length, elliptical equivalent minor axis length, and crack location. The sampled defect data is considered suitable for this application because it encompasses both newly initiated and critically grown defects. The data also indicate presence of clustered defects which is typical of field measurement of RCF damage on rails. The demonstration also compares the fracture behaviour of the equivalent elliptical representation of sampled detected defect to that of the actual boundary coordinates representation of each defect under investigation.

S/N	Defect number	Maximum surface Length [mm]	Defect Orientation [deg]	Major-axis Length [mm]	Minor-axis Length [mm]	Defect location (x,y) [mm]
1	2	5.22	86.11	5.46	1.43	(0.75,23.41)
2	98	11.47	38.58	11.10	5.18	(9.51,129.92)
3	97	23.47	89.33	24.83	7.81	(10.08,146.63)
4	90	18.47	65.64	20.03	3.89	(9.25,72.36)
5	88	27.07	86.66	26.96	11.19	(10.16,173.95)
6	4	6.21	84.84	7.21	2.08	(0.82,34.37)
7	41	2.67	75.40	3.06	0.65	(2.11,74.18)
8	45	4.48	84.42	4.51	1.27	(2.69,117.57)
9	50	5.21	87.00	5.76	0.87	(3.05,105.71)
10	52	22.55	87.38	27.08	3.35	(4.59,29.18)
11	59	6.33	89.30	6.88	1.49	(3.78,49.42)
12	71	6.97	77.22	6.45	2.08	(4.68,158.43)
13	73	6.88	64.92	7.32	1.68	(5.04,74.38)
14	80	11.76	68.31	11.91	1.86	(6.84,118.51)

Table 5. 1: Defect samples and their respective geometrical details.

In Figure 5.1, the highlighted rail damage is extracted for application purposes as shown in Figure 5.2, each detected defect is depicted as a blob (binary image with crack region indicated as foreground 1 and rail head as background 0). Furthermore, the boundary coordinates (in pixel units) of each defect is also presented. Based on visual inspection of Figure 5.1, Table 5.2 below describes each of the selected defects.

Defect ID	Description
1	Leaf on rail, surface strain, or rust.
2	Leaf on rail, surface strain, or rust.
3	Leaf on rail, surface strain, or rust.
4	Grind marks or surface scratch.
5	Indentation, e.g. gravel in the contact.
6	Grind marks or surface scratch.
7	Fatigue damage.
8	Indentation, e.g. gravel in the contact.
9	Spall damage.
10	Fatigue damage.
11	Spall or indentation defect, combined with

	rolling contact fatigue damage.
12	Spall defect or indentation defect, combined with rolling contact fatigue damage.
13	Spall or indentation defect, combined with rolling contact fatigue damage.
14	Rolling contact fatigue damage.

Table 5. 2: Defect samples and their respective description.

It is observed from Table 5.2 and Figures 5.1-5.2, that defects 1-6 described as presence of leaves, lubricant stain, rust marks, grind marks, and scratches; are of no interest to fracture mechanics prediction aspects of this research. Furthermore, these defects have intersecting boundary points and therefore not easily incorporable within the finite element model (due to meshing errors at the interesting boundary or edge). The simulated results of defect 7-14 described as spall, fatigue or rolling contact fatigue damage are therefore discussed in detail in the next section of the chapter.
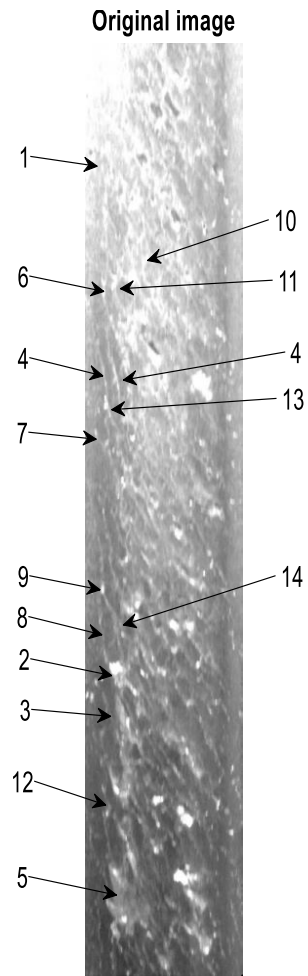
Figure 5. 1: Shows the randomly selected damage(s) of FA-H1 image applied to the proposed technology.
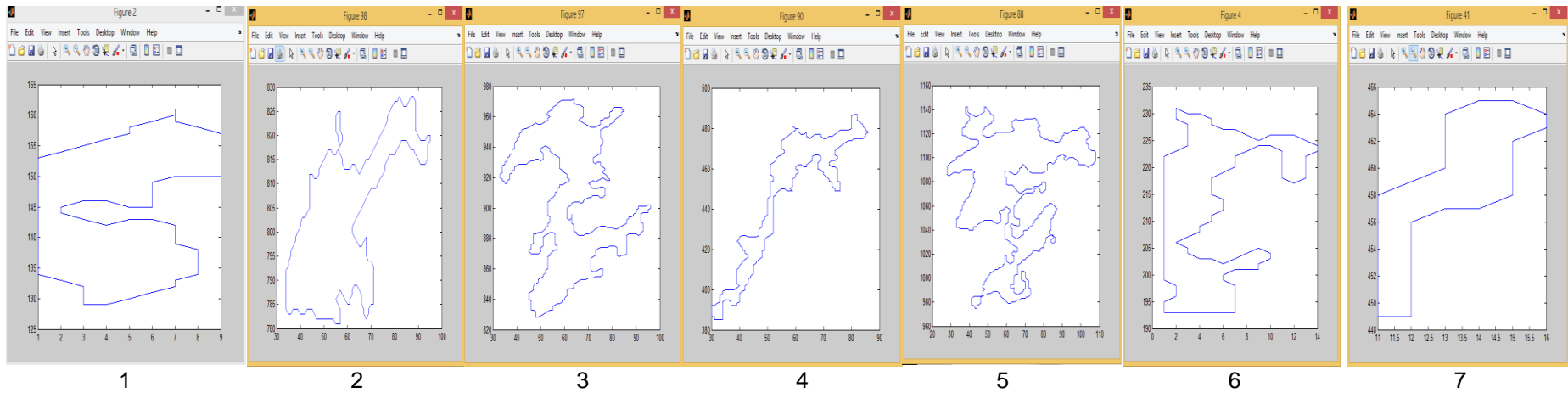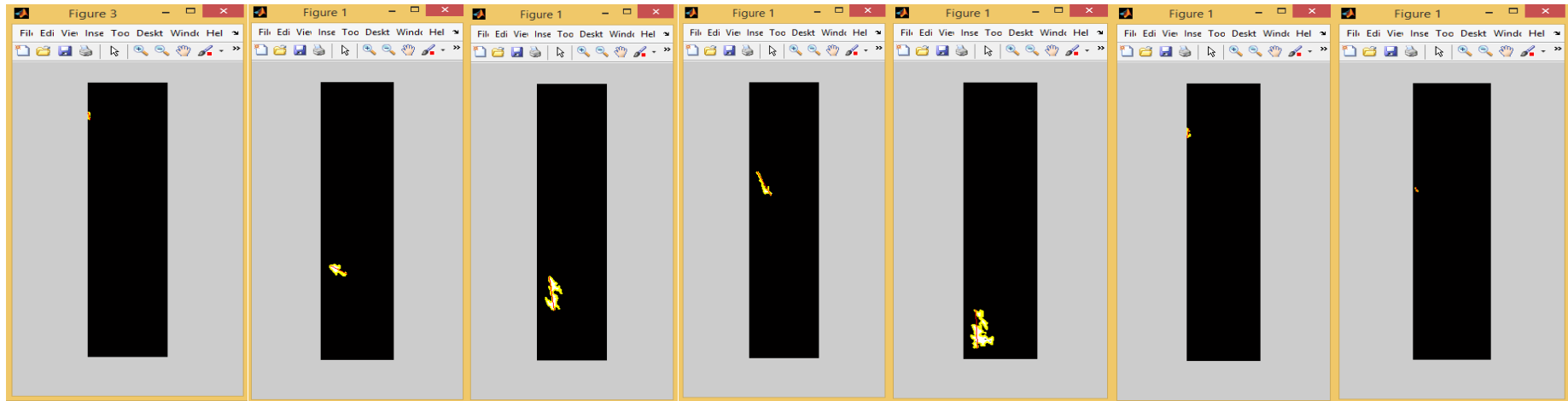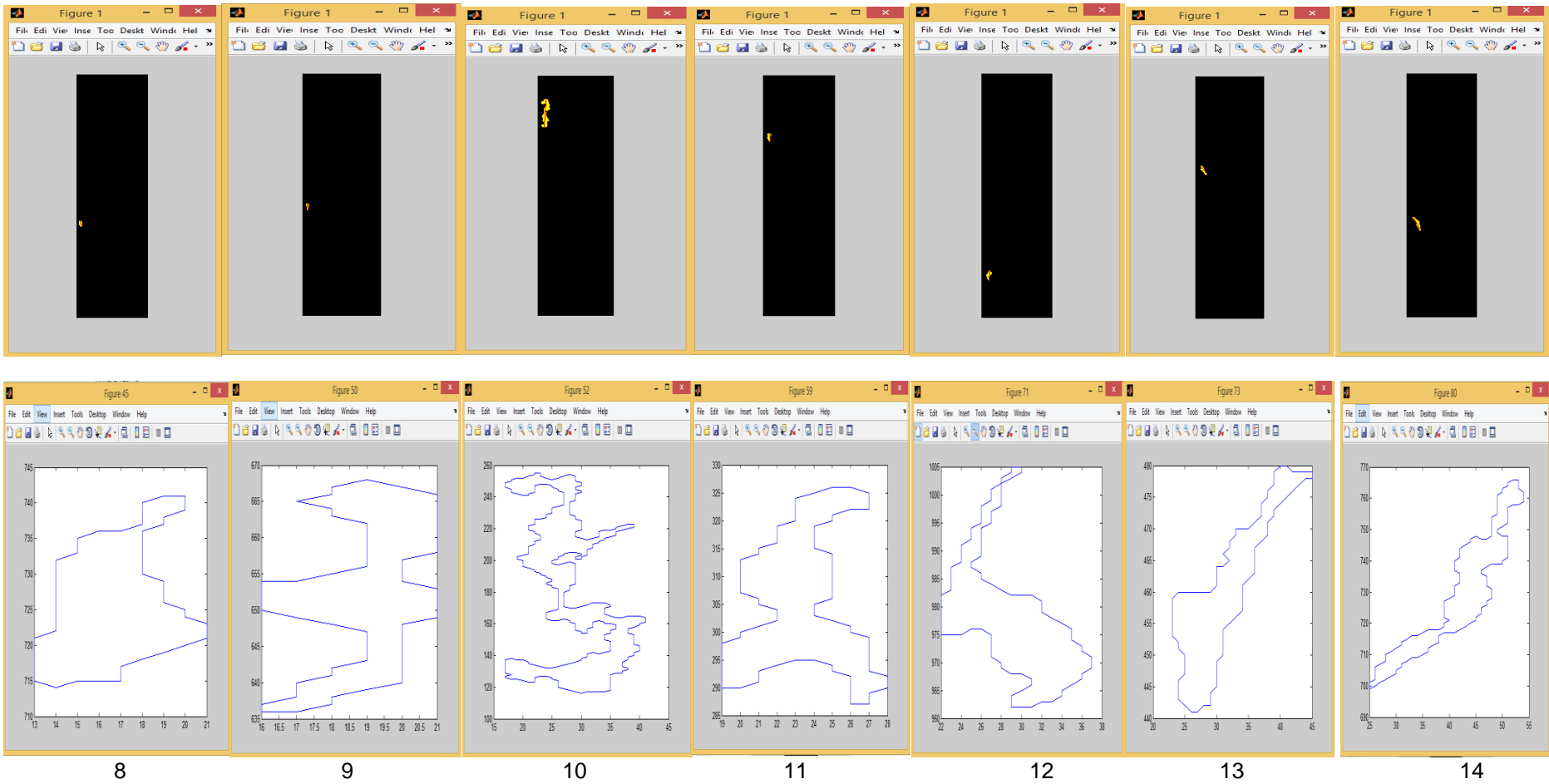
Figure 5. 2: Shows the random samples of detected defect from image H1.

5.3 Simulated results and discussion

In Figure 5.3, the FE response of the actual geometry of defect number 6 (as shown in Figure 5.2 above) to wheel rail contact stress for an axle load of 80kN is presented. The defect is of initial length of 6.2mm and an orientation of 84^0
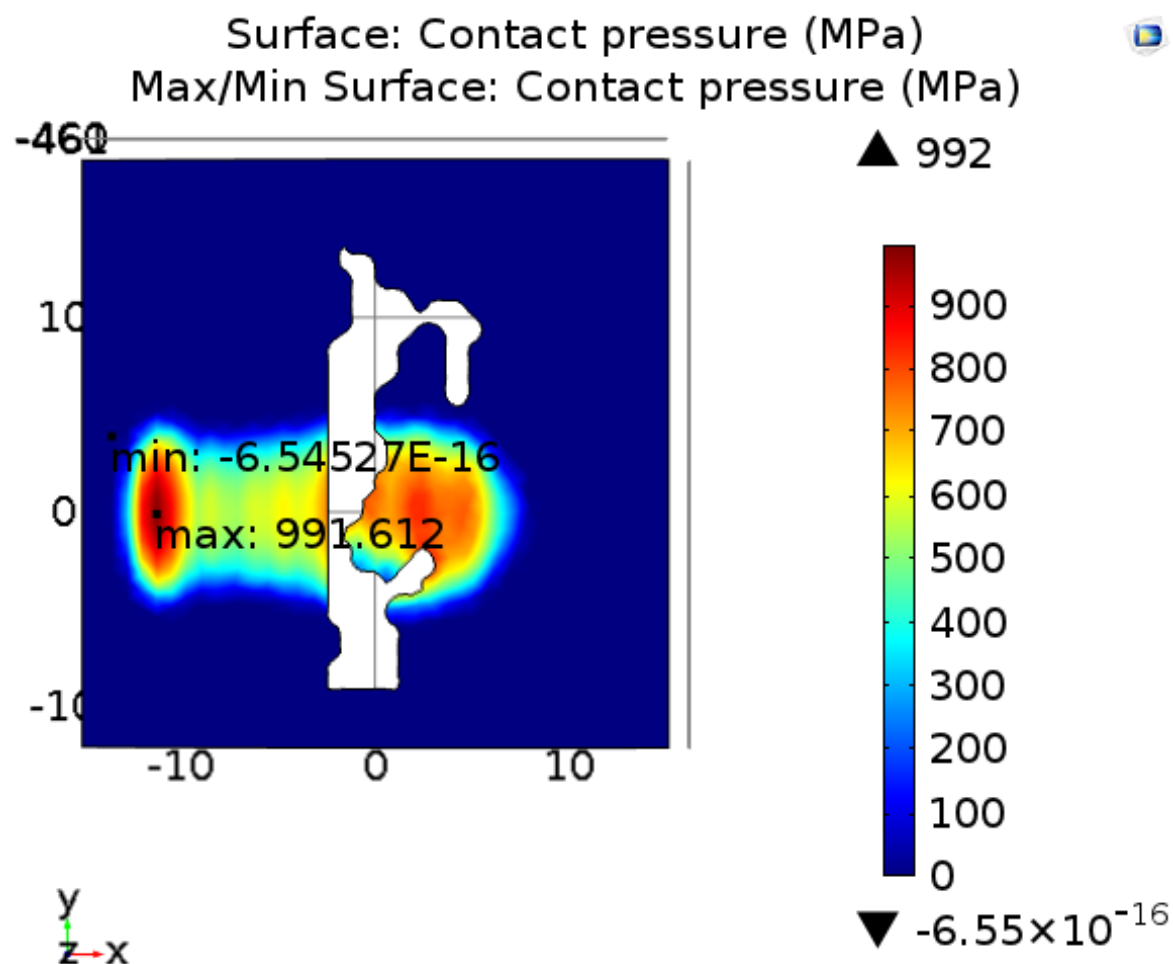
Figure 5. 3: Shows the actual geometry for defect number 6 and the resulting stress distribution.

The actual crack geometry is observed to develop stress concentration at multiple crack tips which depends on the shape of the crack, as opposed to the elliptical equivalent where stress concentration is mostly developed at both ends of the ellipse. The estimated propagation rate in Figure 5.4 is observed at peak axle load of 80KN, to be approximately 620nm/cycle.
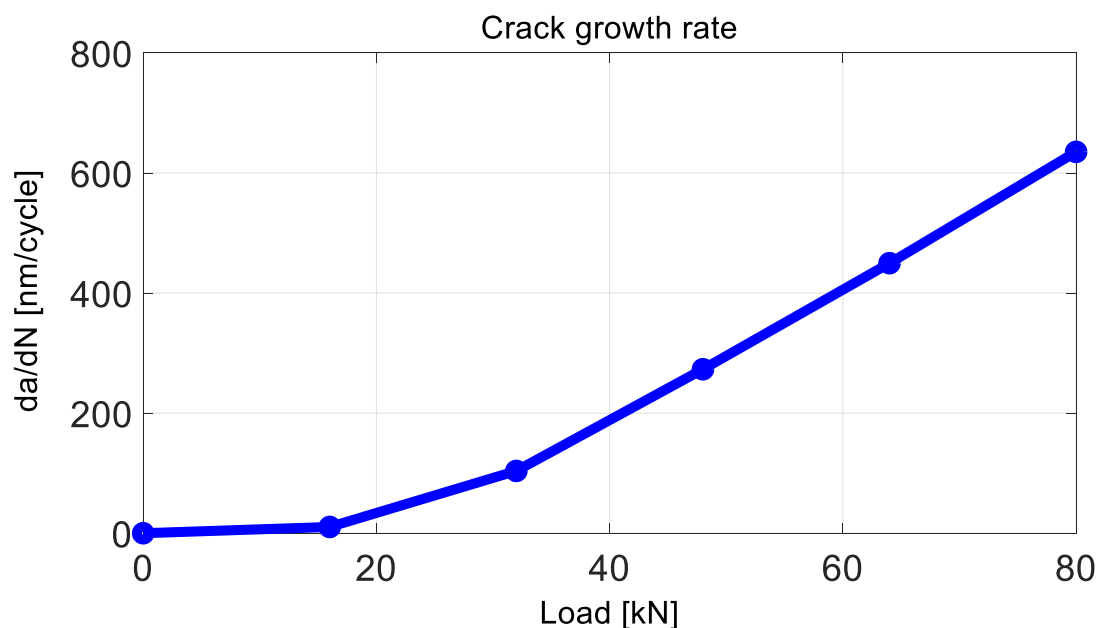
Figure 5. 4: Shows the growth rate of actual defect geometry for defect number 6.

The growth path depicted in Figure 5.5 is observed to dip from 3^0 to almost zero before it increases at above 25% of applied axle load and saturate to about 5^0 at peak axle load of 80 kN.
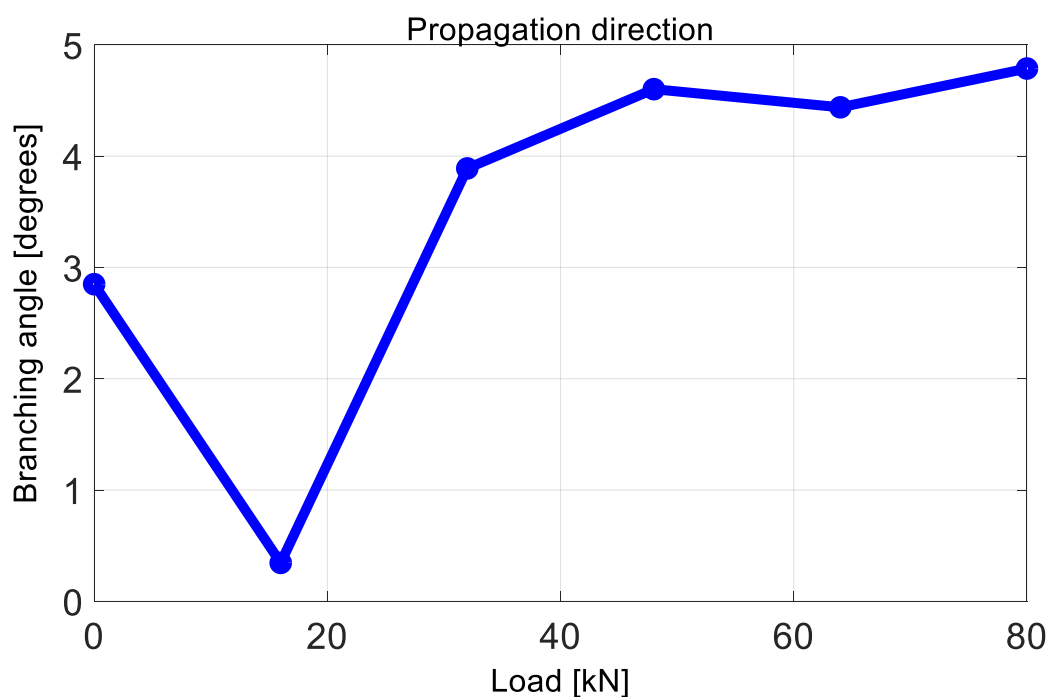
Figure 5. 5: Compares the simulated propagation direction for elliptical and actual geometry for defect number 6.And in Figure 5.6 the residual life of the rail is observed to exponentially decay from

about 7E3to 3.39E2 at maximum axle load of 80kN.
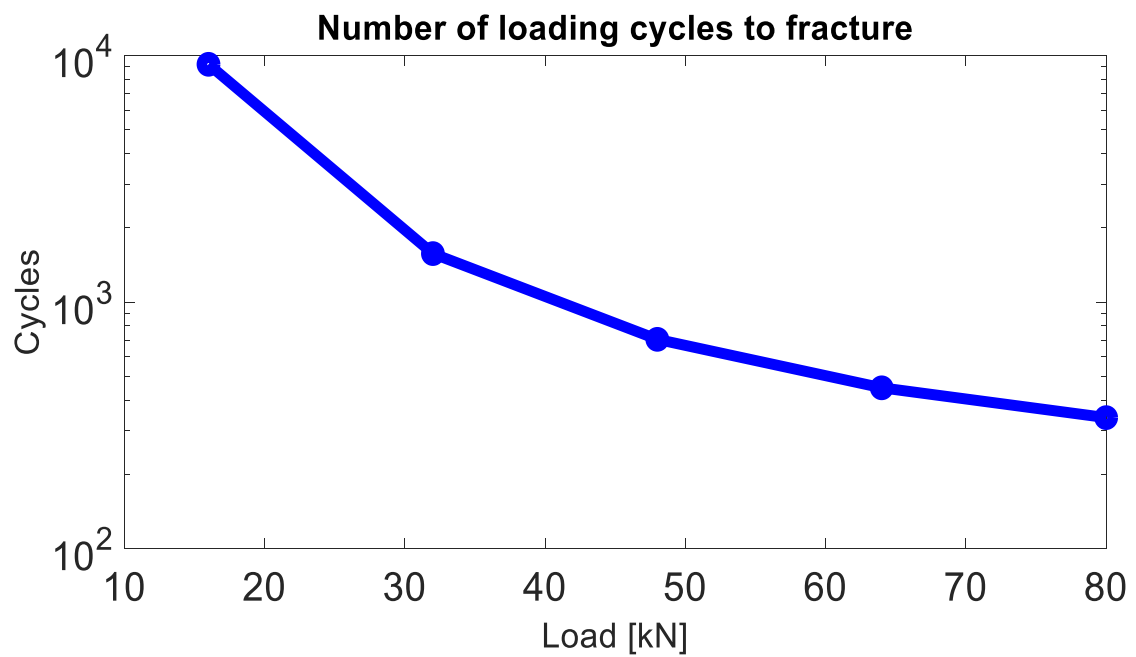
Figure 5. 6: Compares the simulated residual life for elliptical and actual geometry for defect number 6.

In Figure 5.7 the FE response of the actual geometry of defect number 7 (as shown in Figure 5.2 above) to wheel rail contact stress for an axle load of 80kN is presented. The defect is of initial length of 2.6mm and an orientation of 75^0.

Figure 5. 7: Shows the simulated stress distribution actual geometry for defect number 7.

The actual crack geometry is observed to develop stress concentration at multiple crack tips which depends on the shape of the crack, as opposed to the elliptical equivalent where stress concentration is mostly developed at both ends of the ellipse. The estimated propagation rate in Figure 5.8 is observed to at peak axle load of 80KN, to be approximately 120nm/cycle.

Figure 5. 8: Shows the simulated growth rate for actual geometry for defect number 7.

The growth path depicted in Figure 5.9 is observed to dip from 3^0 to almost zero before it increases at above 25% of applied axle load, no saturation is observed even though 4.8^0 branching angle is observed at peak axle load of 80 kN.
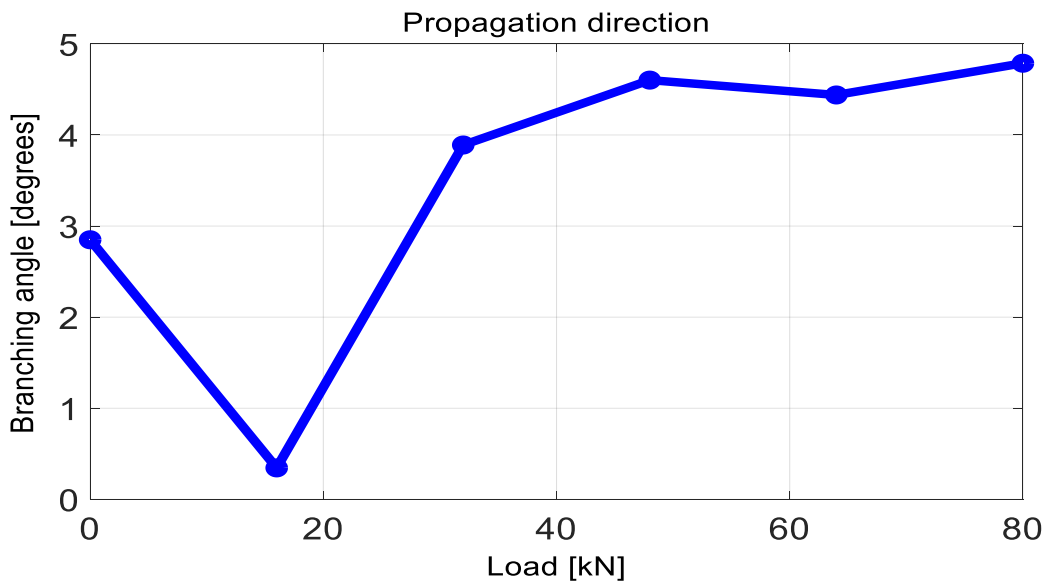
Figure 5. 9: Compares the simulated propagation direction for elliptical and actual geometry for defect number 7.

And in Figure 5.10 the residual life of the rail is observed to exponentially decay from about 3.97E5 to 1.64E4 at maximum axle load of 80kN.

Figure 5. 10: Shows the simulated fracture mechanics behaviour actual geometry for defect number 7.

In Figure 5.11 the FE response of the actual geometry of defect number 7 (as shown in Figure 5.2 above) to wheel rail contact stress for an axle load of 80kN is presented. The defect is of initial length of 5.2mm and an orientation of 87^0. The crack is considered critically grown and is expected to propagate in the second phase i.e. with three distinct linear growth rates as described in the literature review discussed in section 2.3.2.

Figure 5. 11: Shows the simulated stress distribution for actual geometry for defect number 9

The actual crack geometry is observed to develop stress concentration at multiple crack tips which depends on the shape of the crack, as opposed to the elliptical equivalent where stress concentration is mostly developed at both ends of the ellipse. The estimated propagation rate in Figure 5.12 is observed at peak axle load of 80KN, to be approximately 125nm/cycle.

Crack growth rate

Figure 5. 12: Shows the simulated growth rate for actual geometry for defect number 9

The growth path depicted in Figure 5.13 is observed to oscillate with a peak branching angle of 3.8^0 and a trough of 0.5^0 at peak axle a branching angle of about 2^0 is recorded.

Propagation direction

Figure 5. 13: Shows the simulated growth path for actual geometry for defect number 9

And in Figure 5.14 the residual life of the rail is observed to exponentially decay from about

1.4E4 to 2.94E3 at maximum axle load of 80kN.
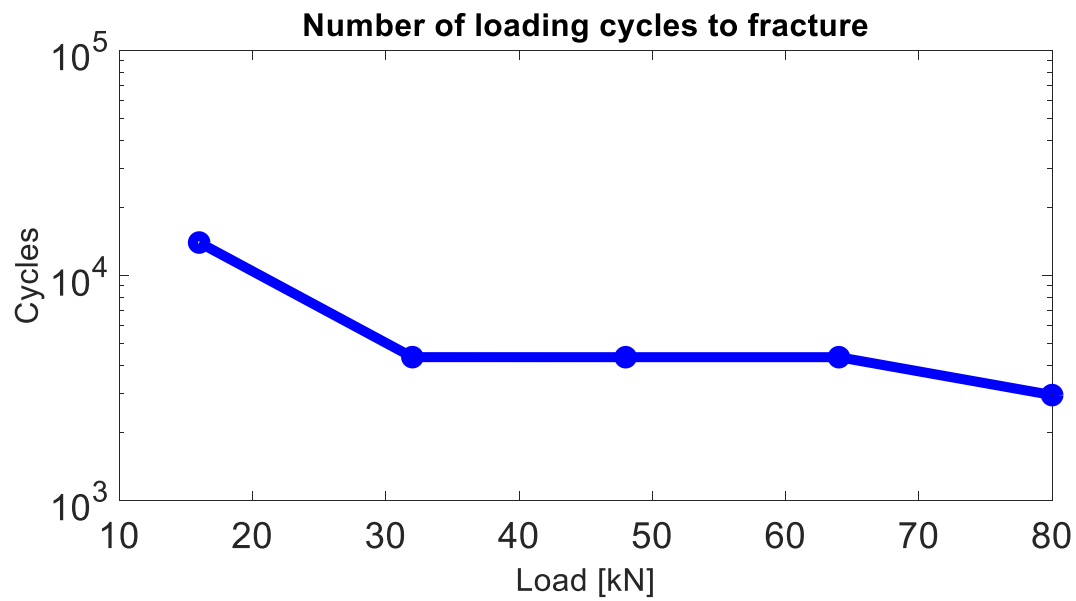
Figure 5. 14: Shows the simulated fracture mechanics behaviour for actual geometry for defect number 9

In Figure 5.15 the FE response of the actual geometry of defect number 7 (as shown in Figure 5.2 above) to wheel rail contact stress for an axle load of 80kN is presented. The defect is of initial length of 6.8mm and an orientation of 64^0.
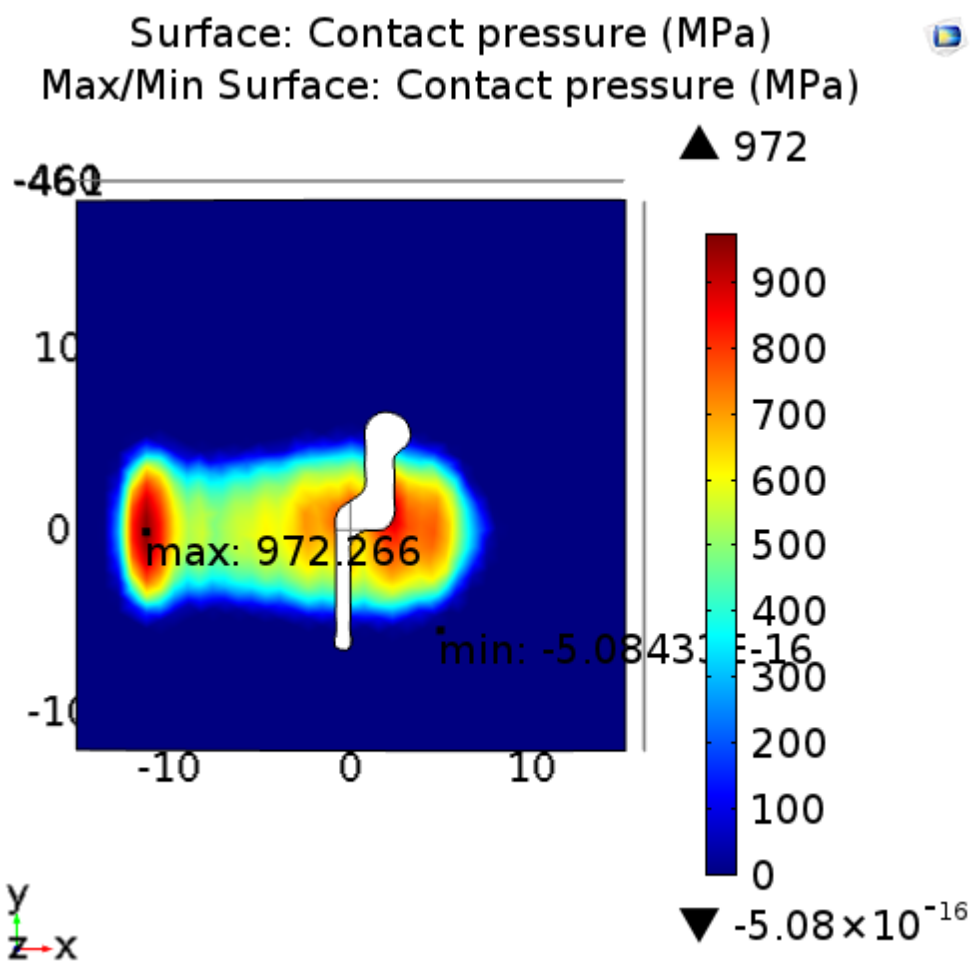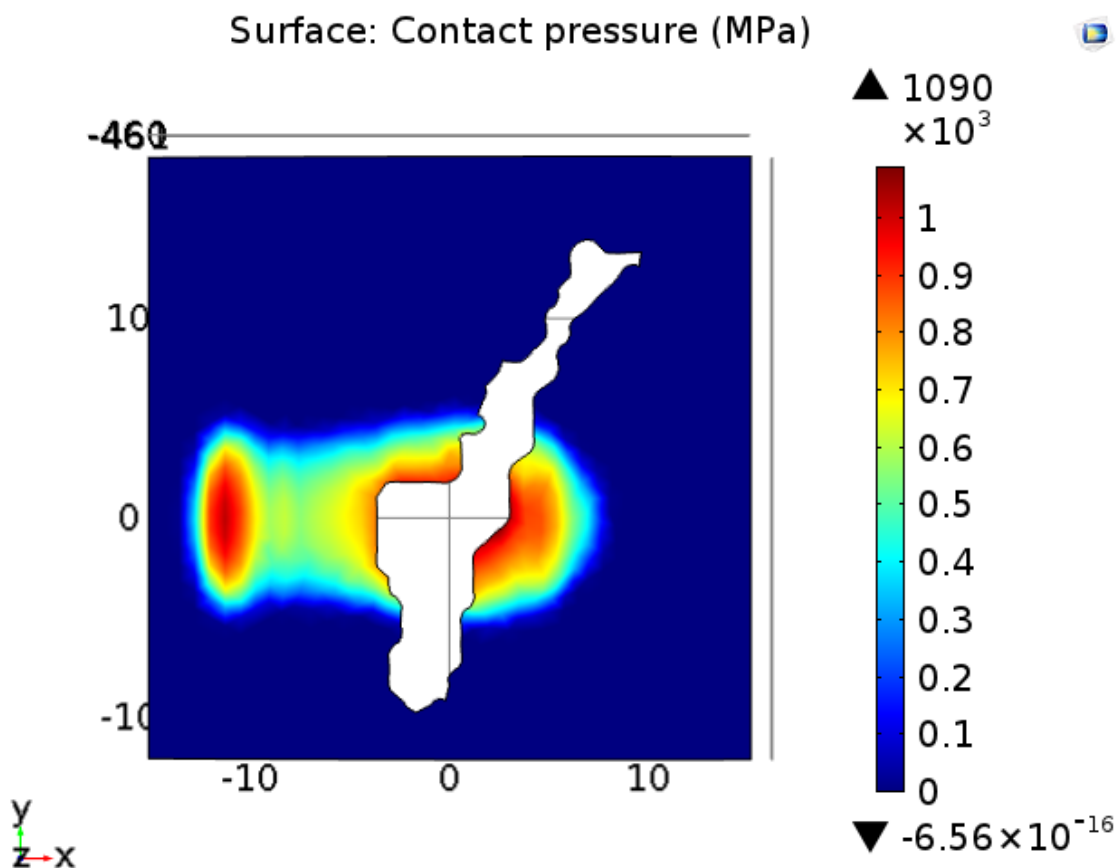
Figure 5. 15: Shows the simulated stress distribution for actual geometry for defect number 13

The actual crack geometry is observed to develop stress concentration at multiple crack tips which depends on the shape of the crack, as opposed to the elliptical equivalent where stress concentration is mostly developed at both ends of the ellipse. The estimated propagation rate in Figure 5.16 is observed at peak axle load of 80KN, to be approximately 410nm/cycle.
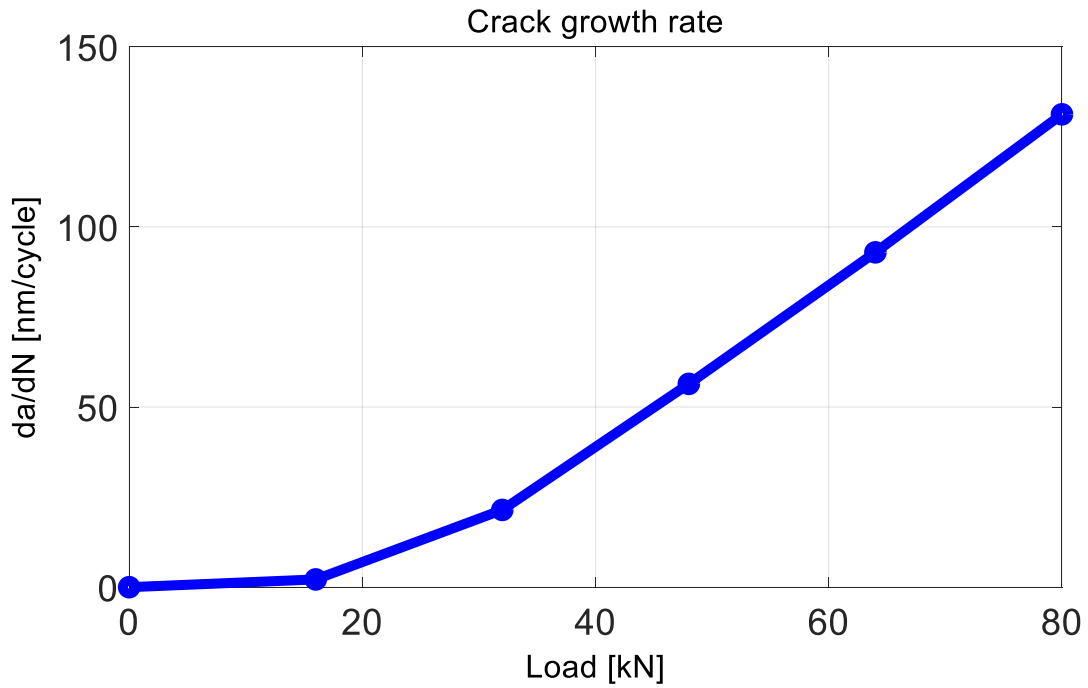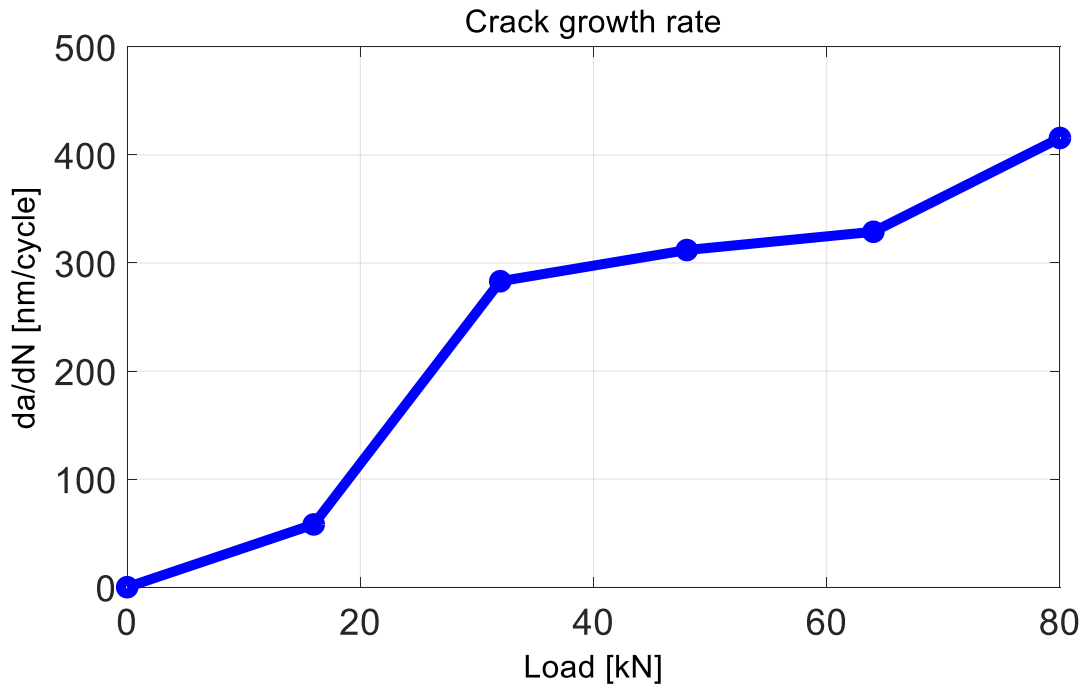
Figure 5. 16: Shows the simulated growth rate for actual geometry for defect number 13

The growth path depicted in Figure 5.17 is observed to be maximum of above 25^0 at 25%-50% of applied axle load and a steep decrease to less than 5^0 and subsequently at maximum axle load of 80 kN the branching angle is estimated to be about 15^0.
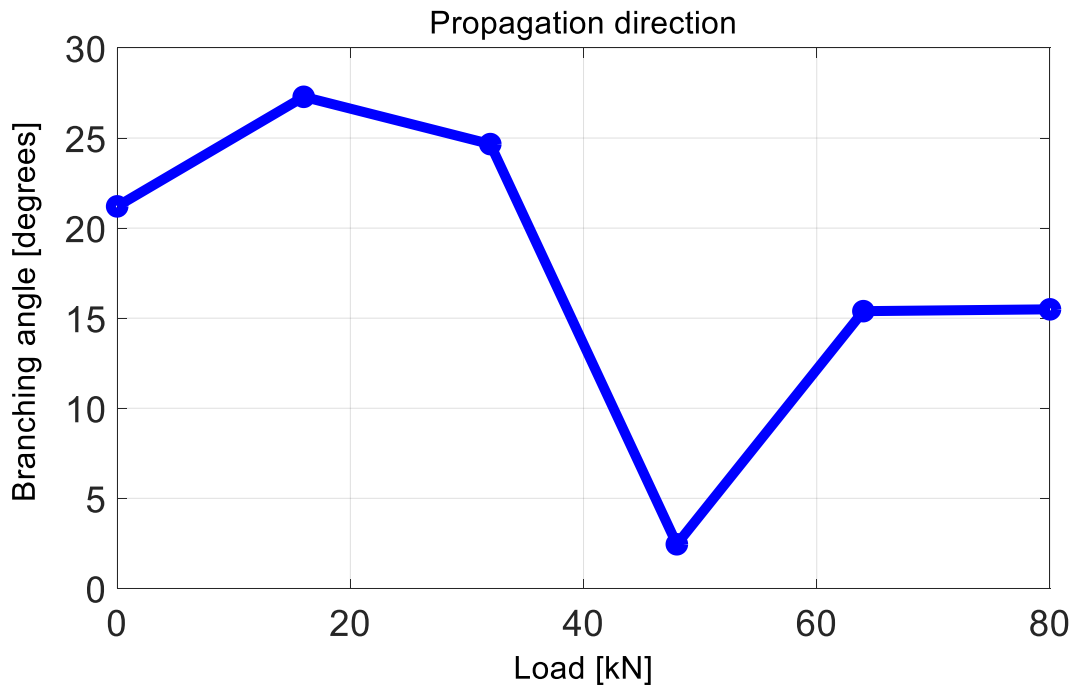
Figure 5. 17: Shows the simulated growth path for actual geometry for defect number 13

And in Figure 5.18 the residual life of the rail is observed to exponentially decay from about
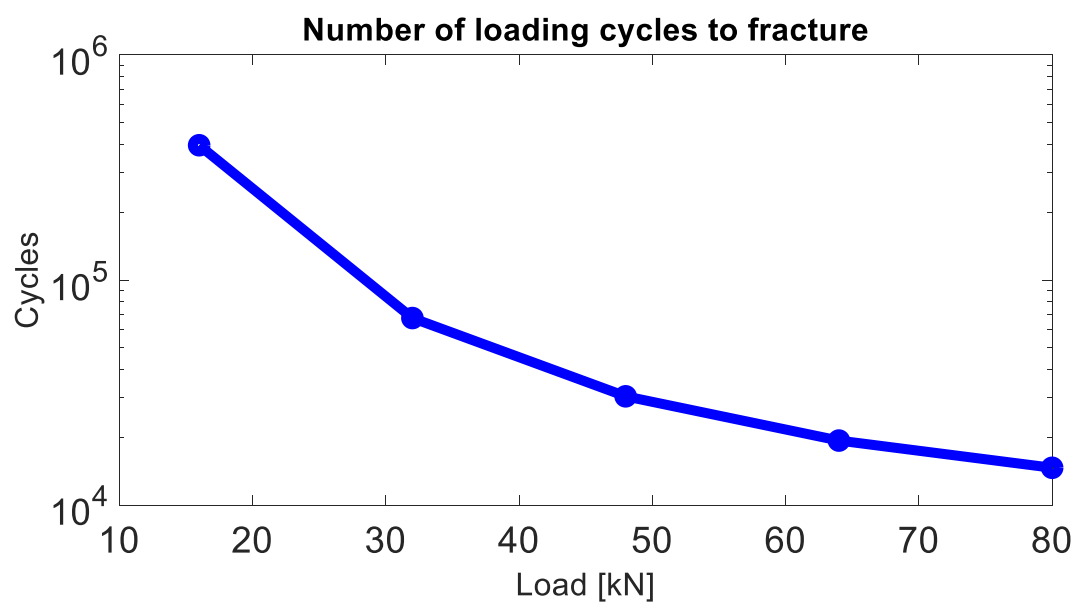
2.4E2 to 4.5E1 at maximum axle load of 80kN.
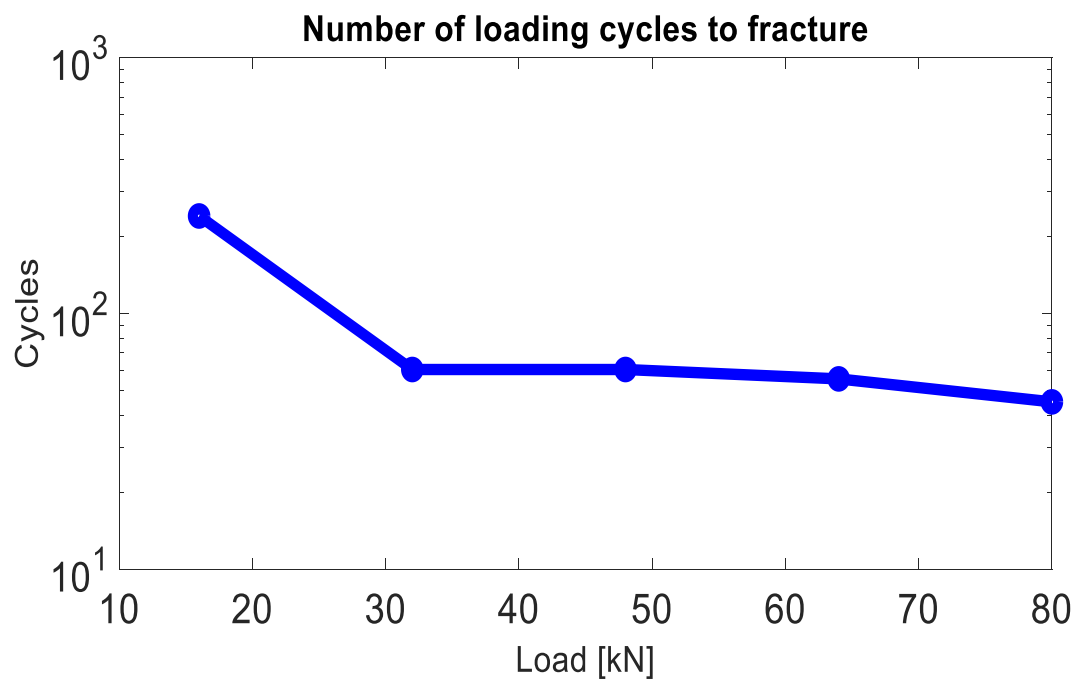
Figure 5. 18: Shows the simulated fracture mechanics behaviour for actual geometry for defect number 13

In Figure 5.19 the FE response of the actual geometry of defect number 7 (as shown in

Figure 5.2 above) to wheel rail contact stress for an axle load of 80kN is presented. The

defect is of initial length of 11.7mm and an orientation of 68^0.
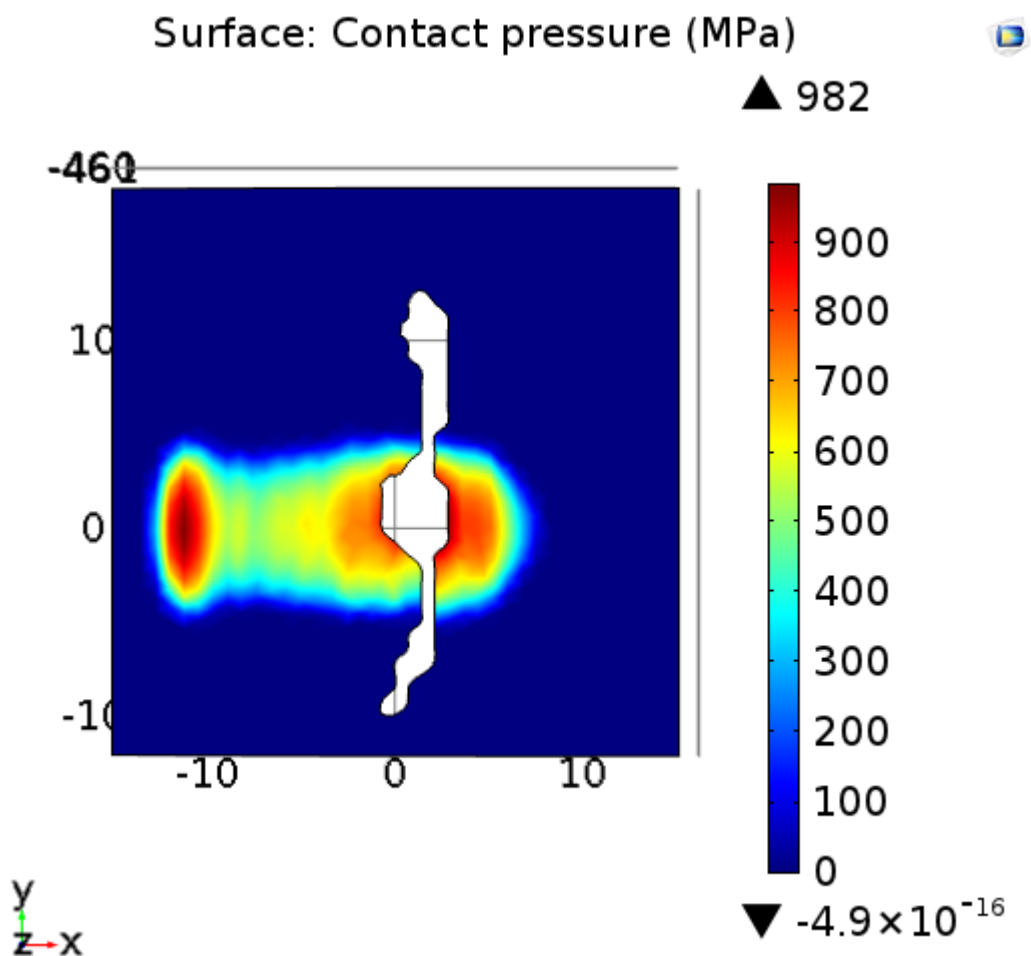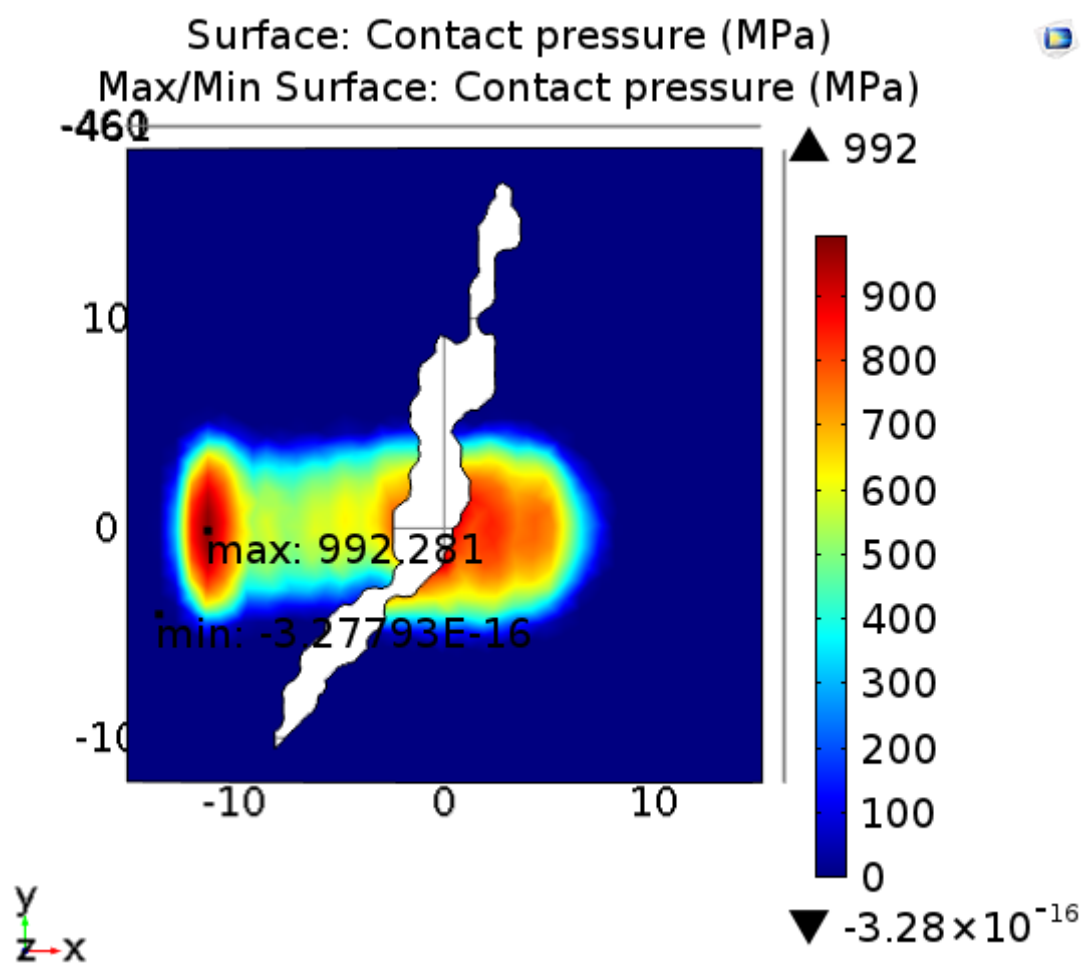
Figure 5. 19: Shows the simulated stress distribution for actual geometry for defect number 14

The actual crack geometry is observed to develop stress concentration at multiple crack tips which depends on the shape of the crack, as opposed to the elliptical equivalent where stress concentration is mostly developed at both ends of the ellipse. The estimated propagation rate in Figure 5.20 is observed at peak axle load of 80KN, to be approximately 300nm/cycle.
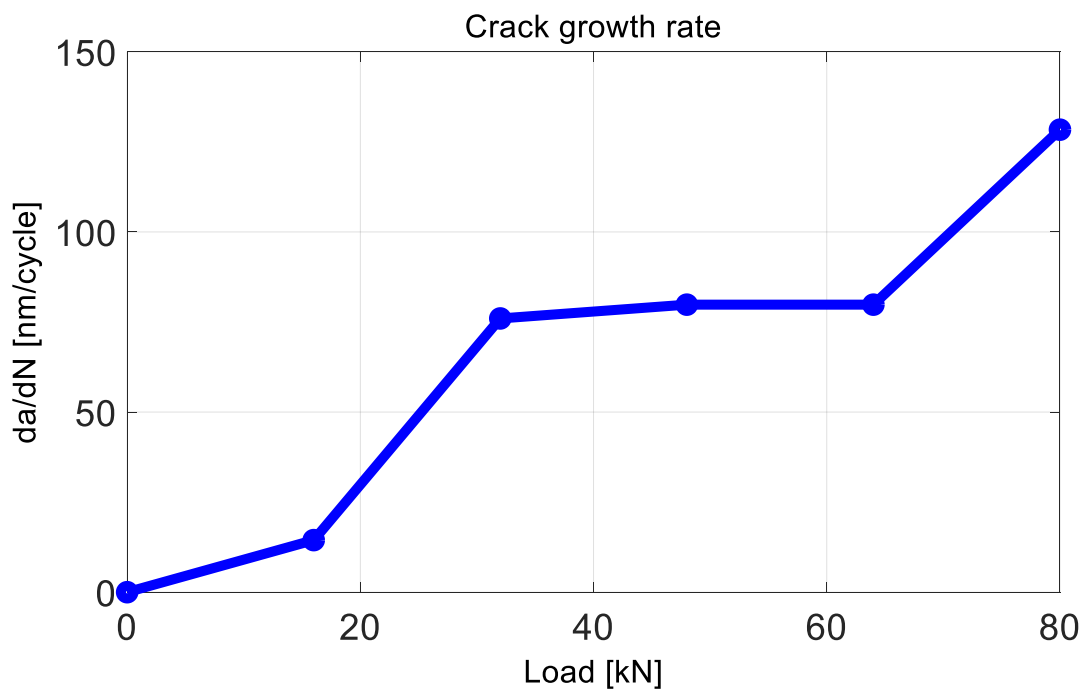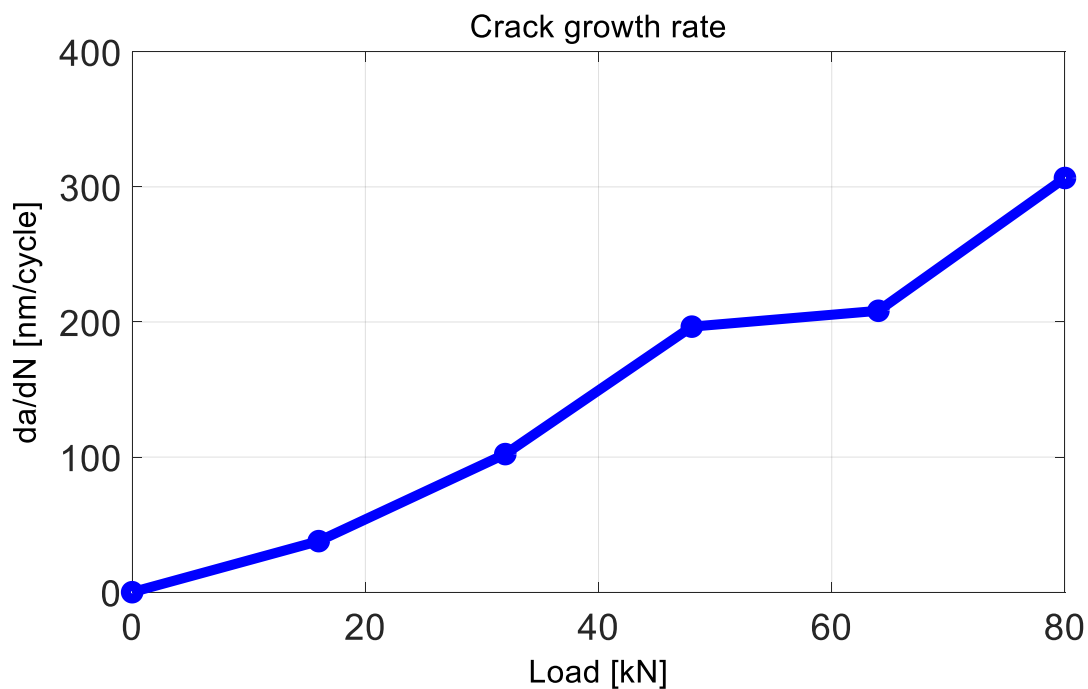
Figure 5. 20: Shows the simulated growth rate for actual geometry for defect number 14

The growth path depicted in Figure 5.121 is observed to be maximum of above 25^0 at 25%-50% of applied axle load and a steep decrease to less than 5^0 and subsequently at maximum axle load of 80 kN the branching angle is estimated to be about 15^0.
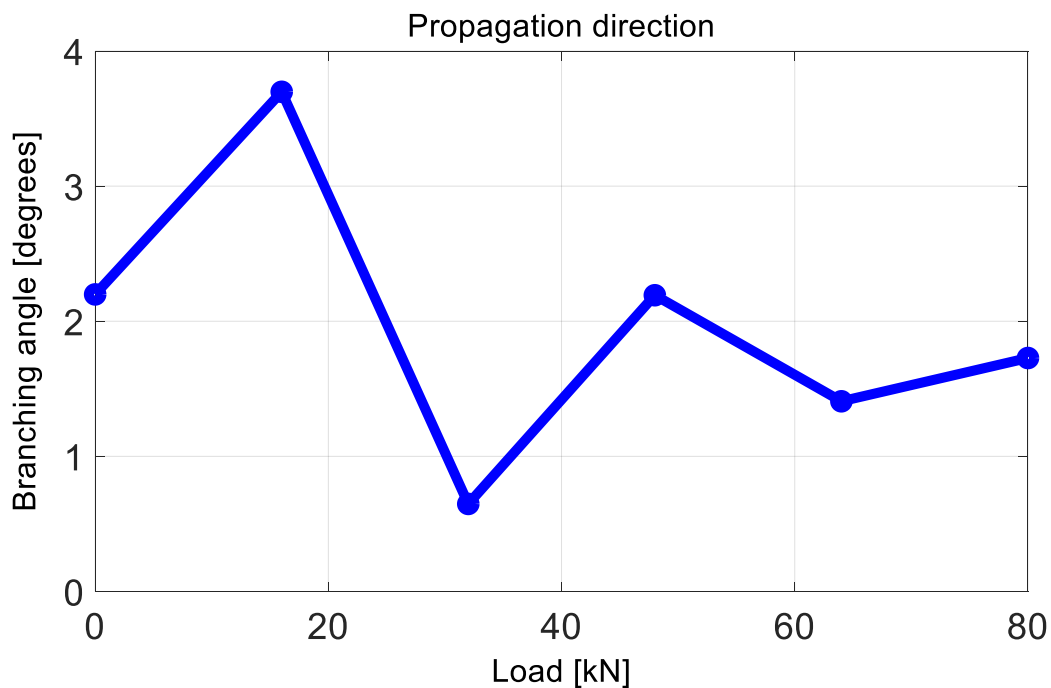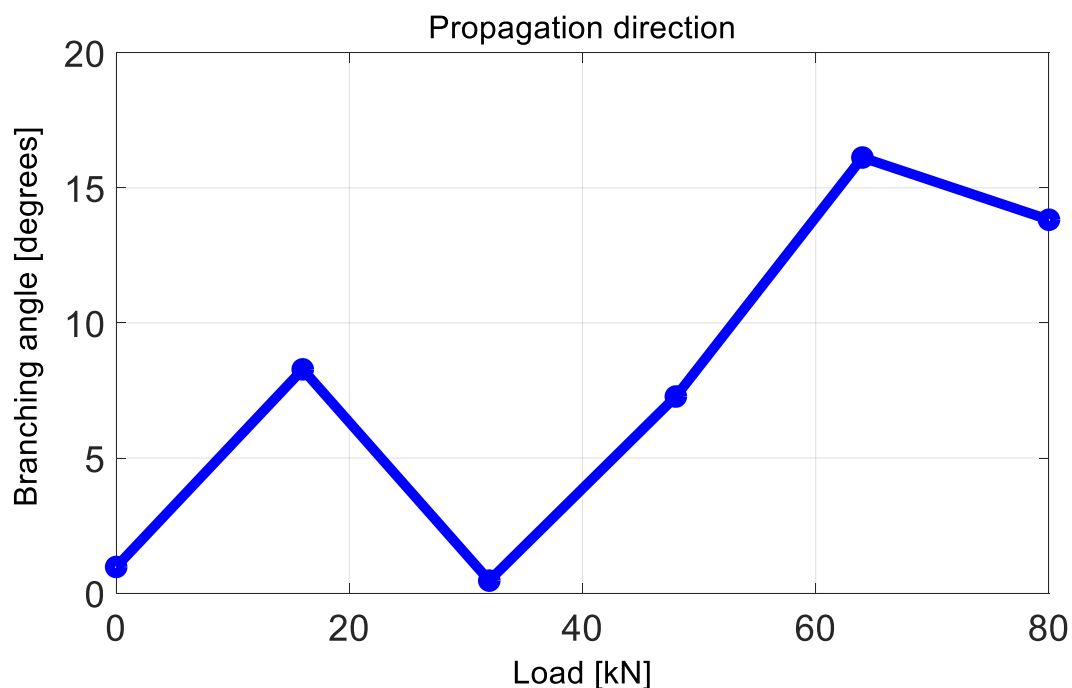

Figure 5. 21: Shows the simulated growth path for actual geometry for defect number 14

And in Figure 5.18 the residual life of the rail is observed to exponentially decay from about

5.10E3 to 7.58E2 at maximum axle load of 80kN.
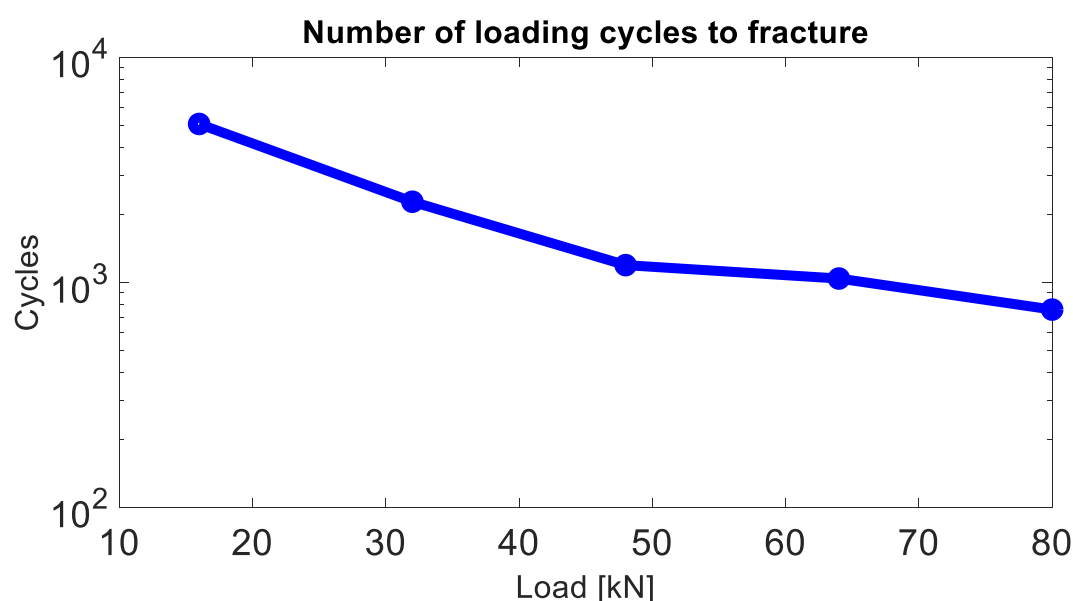
Figure 5. 22: Shows the simulated fracture mechanics for actual geometry for defect number 14

The growth path of these defects are observed to vary by, 2^0, 30^0, and 20^0 respectively. It is worth mentioning that in the case of defect number 13 the growth path for actual geometry of the defect is higher than that of the elliptical equivalent geometry, which is attributed to the crack length and overall shape of this defect. As noted in Chapter 4, the residual life of all simulated results is governed by the growth rate and the same conclusion can be established in this section.

5.5 Enablers and blockers of adopting the proposed technology in industry

It is hoped that this novel application of incorporated detection and prediction models of rail damage will be a useful tool for maintenance purposes in industry and the following factors are key to the prospect of full scale implementation of the proposed technology.

I.   On-board and in service analysis of detection and prediction results is possible.

II.   After installation of image acquisition system, all other processes are fully automated.

III.   The flexibility of the technology supports wide range of operational conditions.

IV.   Multi-tasking nature of the application will simplify maintenance operations.

V.   Availability of diverse image data collection and processing technologies will encourage adoption of the proposed method in industry.

VI.   Little or no human intervention in the method promotes safety considerations, which might encourage the application of the technology.

VII.   The robustness of the method allows for continuous expansion of the technology's scope and intermediate objectives using bigger database or new findings.

VIII.   The ease of including measurements of other NDT methods such as eddy currents or MRX will enhance the self-learning capability of the technology (i.e. bench marking or calibration of measurements).

IX.   The overall cost of the proposed methods in the long run is highly competitive in relation of other methods currently used in industry

In spite of all the above mentioned enablers, the proposed technology is observed to have the following limitations.

I.   The fracture mechanics prediction of some detected defects is not possible due to the presence of intersecting crack boundary.

II.   The most accurate method for generating stresses at the wheel-rail interface is computationally too expensive.

III.   The environmental and rail illumination properties are not easy to control and can affect the performance of detecting RCF damage by image processing.

IV.   The influence of track irregularities, curvature, and cant deficiency may affect detection and/or calibration of defect geometrical measurements.

Based on the pros and cons highlighted above, it is logical to further pursue a Reliability, Availability, Maintainability, and Safety (RAMS) test on the proposed technology. This will introduce a standardisation and clearly stated procedures for the application of the method to ensure very low occurrence of failures with the rail network.

5.4 Conclusion

This technology provides the benefit of on-line and off-line assessment of rail images. The assessed images will positively impact the prioritisation and planning by rail maintenance engineers through classification of severity of RCF damage observed on a track section and use of this information to predict propensity of the RCF damage to grow and eventually cause failure. This technology provides evidence on the need to move to a risk-based maintenance approach, where maintenance is scheduled based on the actual condition and predicted future life rather than exceeding the damage threshold limit. The method is also capable of reducing the number of manual track inspection and therefore less risk of injury to maintenance staff. Initial solutions to the major limitation of the incorporated model (high-computational time of the global contact model and inconclusive results for intersecting boundary coordinates of detected defects) have been successfully curtailed using look up tables of pre-determined contact stresses (or analytical estimates of contact stress) and the use of elliptical equivalent representation of the detected defect respectively. Further analysis and parametric studies of the technology will be required for developing the readiness level of the proposed technology in rail industry.

Chapter 6 Thesis conclusions and future work

The main aim of this PhD project was to link the detection and prediction of fracture behaviour in rails. The aim(s) and intermediate objectives outlined in   section 1.1.2 have been achieved and the following conclusions were reached.

6.1 Conclusions from the conceptualisation of the project (chapter one)

A broad understanding of the limitations associated to state of the art technologies used for rail RCF damage detection and quantification will not only encourage research and development of novel methods but also improve the overall efficiency of maintenance activities. The significance of more intelligent and self-learning condition monitoring practices in rail industry is an important criterion for examining and discussing the suitability and performance of NDE techniques of interest to this research work.

6.2 Conclusions from the review of existing work (chapter two)

It is important to understand the differences between the existing NDTs, and the survey conducted in Chapter 2 suggested the necessity to consider various requirements and more importantly the information provided by the NDE method. This supports the decision of choosing the best suited NDE to be linked with fracture mechanics model. Existing publications suggest even the most popular Ultrasonic test and Eddy current technique are not preferred over novel or even hybrid systems. The prospects of linking automated visual inspection of rails to prediction models is highly favoured as a direct consequence of fast inspection speed, flexibility, and remoteness. A critical appraisal of the existing image processing algorithms particularly those adopted in rail RCF detection must take cognisance of the effect of rail head non-uniform reflection property, rust, grind marks, and scratches. The mechanisms responsible for initiating and propagating RCF damage in rails are well understood in literature. In this review the conclusion affirms that in the U.K, majority of RCF is initiated as a result of Mode-0 and Mode-2 initiation mechanism. While the dominant loading condition and resulting SIF for surface propagation are the mode I and II. RSSB, establishes

the influence of traction, wear, entrapped fluid, and multiple crack interaction on the behaviour of crack propagation.

6.3 Conclusions from the development of image processing algorithm (chapter three)

Chapter 3 provided more insight on the effectiveness in using image processing algorithm for the automatic detection of RCF damage in rails. From all the data set available in this research, it is evident that no single method for each stage of the algorithm is equipped to solve the challenges of true positive detection. The major finding from this chapter of the thesis is that variation of intensity value between neighbouring pixels is not the best method to identify RCF cracks, even though the best indicator of rail damage. The benefits of combining one or more methods depending on a wide variety of condition cannot be over emphasised. The receiver operating characteristics of the proposed image processing algorithm has been observed to deliver up to 92% true positive detection rate for laboratory acquired images and more than 80% true positive detection for field acquired images. After thorough analysis of simulated results, the significance of illumination invariant methods cannot be neglected. Global methods that are sensitive to some of the influential parameters such as mean intensity and variance of neighbouring pixels were modified to local versions with improved performance recorded. Initial correlation between severity of damage and optimum local window size has been established. Even though further analysis and investigation of the proposed algorithm is possible, it is however in the interest of this research to explore the potential applications of the generated geometrical data related to each defect with the hope of enhancing maintenance team planning and predictive maintenance.

6.4 Conclusions from fracture mechanics prediction (chapter four)

In this study, the behaviour of crack growth on rail surface has been investigated by using the F.E. analysis and the mathematical models related to Linear Elastic Fracture Mechanics (LEFM). The fracture parameters are estimated from stresses generated by the global contact model, while the local fracture mechanics model is utilised for determining the growth rate,

direction of propagation, and residual life of the rail. For wheel rail contact applications, Mode-I SIF are observed to be generally higher Mode-II SIF. This is in agreement with the theoretical understanding of KI being most vigour fracture mechanism compared to shear and torsional fracture mechanisms. It is also observed that KI governs the behaviour of crack propagation while KII dictates the growth path behaviour. Maximum growth rate for a defect is observed to be approximately 1/10000 of the estimated KI. The crack direction is observed to vary from 20° to 47° depending on the initial depth of the crack. The number of cycles until failure for crack is observed to be inversely proportional to growth rate, which confirms that a faster propagating crack will fail faster. In general, the F.E. analysis showed that the behaviour of the crack generated on rail surface varies by the load condition (magnitude and position) and crack length (surface and mouth opening). A significant amount of literature and field tests suggest that surface crack translates to increases in depth to a certain limit from where a considerable decrease is observed. Other influential factors such as friction, braking and traction are well understood and have been reported to closely relate to crack length and load. The absence of friction is observed to increase growth rate for a crack whose length is 3 times its distance to maximum contact pressure. Furthermore, the crack length at which the crack growth rate is maximum became shorter as the friction coefficient increased. In the case of braking, the crack growth rate is maximum when crack length is 3 times the crack mouth opening regardless of the magnitude of the braking force. While in the case of traction force, growth rate decreases as load position increases. For both global track model and local fracture models it is believed that the variation in the prediction accuracy is due to track irregularities, and quasi-static vehicle behaviour over idealised track.

6.5 Conclusions from the application of the proposed technology (chapter five)

The application of the technology confirms the flexibility of performing both on-line and off-line assessment of rail images. The assessed images will positively impact the prioritisation and planning of rail maintenance. By means of identifying and classifying severity of RCF damage observed on a track section, the information can be used to predict propensity of the RCF

damage in terms of growth and residual life. This technology provides evidence on the need to move to a risk modelling and risk-based maintenance approach using big data analytics, where maintenance is scheduled based on the actual condition and predicted future life rather than the exceedance of a damage threshold limit. The method is also capable of reducing the number of manual track inspection and therefore less risk of injury to maintenance staff. Initial solutions to the major limitation of the incorporated model (high computational time of the global contact model and inconclusive results for intersecting boundary coordinates of detected defects) have been successfully curtailed using look up tables of pre-determined contact stress (or analytical estimates of contact stress) and the use of elliptical equivalent representation of the detected defect respectively. More analysis and parametric studies of the technology will be required for developing the readiness level of the proposed technology in rail industry.

Future work(s)

The research described in this thesis has identified a number of areas of further work as detailed below.

Performance of damage detection can be improved with the investigation of ways to practically modify reflection/refraction of the light source on the image acquisition system as opposed to illumination sensitive algorithms. Improvements can be made to the proposed technology with the introduction of an Artificial Intelligence (AI) module to interpret the features extracted by the image processing algorithm, and a discrimination function can be derived by clearly defining (for each neuron through axon/link weights) input number of attributes for known number of class labels.

The use of more sophisticated crack definition such as Least square support vector machine classifiers could be explored for processing feature extraction data obtained within the image-processing algorithm.

The incorporation of big data analytics on cracking information can be used to improve defect feature classification and also encourage risk modelling alongside predictive maintenance of

rail integrity. It is highly recommended that measured worn profiles, track quality effects, and more information regarding vehicle-track characteristics are included in the global track model. This will involve considering how different vehicle-track characteristics (passenger and freight vehicles) influence prediction of residual life of the rail. The inclusion of actual traffic databases for a route might be an opportunity to refine the model for more accurate prediction of the stresses responsible for initiating and propagating damage in rails. More investigation and parametric studies could be performed to expand the look-up tables of wheel-rail contact stresses, contact and site data.

Other NDE methods such as eddy current measurements could be utilised for self-learning and more intelligent capacity building of the incorporated model.

Further validation of the incorporated model (proposed technology) is required to improve the readiness level of the technology.

References

Ahmad, A. Khaled, E. Sugiyama, H. (2008). Railroad vehicle dynamics A computational approach. Boca Raton, CRC Press.

Al-Samaraie, M. & Abdulmajied Al Saiyd, N. (2007). MEDICAL COLORED IMAGE ENHANCEMENT USING WAVELET TRANSFORM FOLLOWED BY IMAGE SHARPENING. *Ubiquitous Computing and Communication Journal. 6* (5), 1-8.

Ancellotti, S., Benedetti, M., Dallago, M., & Fontanari, V. (2016). Fluid Pressurization and Entrapment Effects on the SIFs of Cracks produced under lubricated Rolling-Sliding Contact Fatigue. Procedia Structural Integrity. 2, 3098-3108.

Anderson, T.L. (2011). FRACTURE MECHANICS Fundamentals and Application. Taylor & Francis Group, Boca Raton.

Arup/TTCI (2001) Rail-track Rail Head Checking Investigation, Phase 2, Final Report, ASME. 89, 459-464.

Banerjee, U. (2013). Multidiscipline Modelling in Materials and Structures. *Modelling damage mechanics of railway tracks to evolve control strategies for derailment prevention.* 9 (3), 341 – 358.

Bar-Am, M., & Yaakov, Z. (2007). *11/679,556.* United States: U.S. Patent Application Office.

Beagles, A. E. (2002). *An Initial Assessment of the Influence of Wheel Profile Changes on RCF Formation*, AEATR-Team, 218.

Beagles, M. (2002). *Whole Life Rail Model: Progress report*, AEA Technology Rail report AEATR-T&S. 206.

Beden, S.M., Abdullah, S., & Ariffin, A.K. (2009). Review of Fatigue Crack Propagation Models for Metallic Components. European Journal of Scientific Research. 28 (3), 634-397.

Belkhade, A. & Kathale, A. (2014). AUTOMATIC VISION BASED INSPECTION OF RAILWAY TRACK: A REVIEW. *IJRET: International Journal of Research in Engineering and Technology. 3* (2), 12-15.

Bezin, Y. (2008). An Integrated Flexible Track System Model for Railway Vehicle Dynamics. (Doctoral thesis), Manchester Metropolitan University.

Bhalekar, B,D., & Patil, R,B. (2016). A Review on Stress Intensity Factor. International Research Journal of Engineering and Technology. 3 (3), 1666-1668.

Bhondwe, S.D., Sanath, P.M., Gore, M.G., Shirodkar, S.V., & Kshirsagar, P. (2014). Detection of Static Obstacles for Railways. *Int. Journal of Engineering Research and Applications. 4* (2), 601-604.

Bogdanski, S., Stupnicki, J., Brown, M.N., & Cannon, D.F. (2013). A Two Dimensional Analysis of Mixed –Mode Rolling Contact Fatigue Crack Growth in Rails. International Conference on Biaxial/Multiaxial Fatigue and Fracture. Poland, 189-206.

Brian, W. (2015). *Challenges and opportunities with the Wheel/Rail interface.* U.K: Network Rail.

Brouzoulis, J., & Ekh, M. (2012). Crack propagation under rolling contact fatigue loading conditions based on material forces. International Journal of Fatigue. 45, 98-105.

Brown, M.W., & Miller, K.J. (1973). A theory for fatigue failure under multiaxial stress–strain conditions. Proceedings Institute of Mechanical Engineers. 187(7), 45–55.

Burstow, M.C. (2006). *A model to predict and Understand Rolling Contact Fatigue in Wheels and Rails.* Derby, U.K.: AEA Technology Rail.

Burstow, M.C. (2011). Improving track geometry alignment to reduce rolling contact fatigue (RCF). 9th World Congress on Railway Research (WCRR).

Cannon D.F., & Pradier, H. (1996). Rail rolling contact fatigue Research by the European Rail Research Institute, *Wear.* 191, 1-13.

Carabias, D.M. (2012). *ANALYSIS OF IMAGE THRESHOLDING METHODS FOR THEIR APPLICATION TO AUGMENTED REALITY ENVIRONMENTS* (Master Thesis). UNIVERSIDAD COMPLUTENSE DE MADRID, Spain, Madrid.

Dalberg, T., & Ekberg, A. (2002). Failure Fracture Fatigue An Introduction. Sweden, Studentliteratur.

Dallago, M., Benedetti, M., Ancellotti, S., & Fontanari, V. (2016) The role of lubricating fluid pressurization and entrapment on the path of inclined edge cracks originated under rolling–sliding contact fatigue: Numerical analyses vs. experimental evidences, International Journal of Fatigue. 2, 3098-3108.

Dirks, B., Enblom, R., Ekberg, A., & Berg, M. (2015). The development of a crack propagation model for railway wheels and rails. Fatigue & Fracture of Engineering Materials & Structures. 38, 1478-1491.

Eadie, D.T., Oldknow, K.D., Maglalang, L., Makowsky, T., Reiff, R., Sroba, P., & Powell, W. (2006). Implementation of wayside top of rail friction control on North American heavy haul freight railways. In Proceedings of the Seventh World Congress on Railway Research, Montreal.

Eadie, T.D., Elvidage, D., Oldknow, K., Stock, R., Pointner, P., Kalousek, J., & Klauser, R. (2008). The effects of top of rail friction modifier on wear and rolling contact fatigue: Full-scale rail-wheel test rig evaluation, analysis and modelling. Wear. 265, 1222-1230.

Edwards, J.R., Hart, J.M., Sawadisavi, S., Resendiz, E., Barkan, C.P.L., Ahuja, N. (2009). *Advancements in Railroad Track Inspection Using Machine-Vision Technology*. American Railway and Maintenance of Way Association Conference Proceedings, Chicago, Illinois.

Erdogan, F., & Sih, G.C. (1963). On the crack extension *in plates under plane loading and transverse shear. Journal of Basic. Engineering, ASME Tr*ans. 85, 519-525.

Bezin, Y., Shahzad, F., De Vries, E., Cavalletti, M., Iwnicki, S. & Evans, G.,. (2009). An investigation of sleeper voids using a flexible track model integrated with railway multi-body dynamics. In Proc. of IMechE, Part F: Journal of Rail and Rapid Transit. 223 (6), 597-607.

Fatemi, A., Socie, D.F. (1988). A critical plane approach to multiaxial fatigue damage including out-of-phase loading. Fatigue Fracture Engineering Material Structure. 11(1), 49–65.

Feng, M-L., & Tan, Y-P. (2004). Adaptive Binarization Method for Document Image Analysis. International Conference on Multimedia and Expo. 339-342.

Forman, R.G., Kearney, V.E., & Engle, R.M. (1967). Numerical a*nalysis of crack Formation*, AEATR-Team, 218.

Forth, S.C. (2003). Three-*dimensional mixed-mode fatigue crack growth in a functionally graded titanium alloy. Engin*eering Fracture Mechanics. 70 (15), 2175-2185.

Garcia-Martin, J., Gomez-Gil, J., & Vasquez-Sanchez, E. (2011). Non-Destructive Techniques Based on Eddy Current Testing. *Sensors. 11* (3), 2525-2565.

Gonzalez, R,C., & Woods, R,E. (2005). Digital Image Processing. Pearson Education, 2nd Ed.

Guodong, Z., Bin, S., & Hong, W. (2005). Effect of elastic modulus on parameter of low cycle fatigue performance. Journal of Aerospace Power. 20 (5), 68–71.

Han, S., Cho, S., Jang, G., & Park, J. (2015). Non-contact inspection of rail surface and internal defects based on electromagnetic ultrasonic transducer. *Journal of Intelligent Material Systems and Structures. 25*, 1-8.

Helen, T.K., Blackburn, W.S. (1975). The calculation of stress intensity factors for combined tensile and shear loading. International Journal of Fracture. 11, 605-617.

Hesse, D. (2007). *Rail Inspection Using Ultrasonic Surface Waves.* (Doctoral Thesis), Imperial College London, London.

Illonen, J., Kamarainen, J.K., & Kalviainen, H. (2005). EFFICIENT COMPUTATION OF GABOR FEATURES. Lappeenranta University of Technology, Department of Information Technology.

INFORMATION PAPER. 1 (2), 25-29.

Innotrack. (2010). *Concluding Technical Report.* Paris: Intellecta Infolog.

Ishmael, S.M., Mmamolatelo, E.M., Fulufhelo,V.N., & Tshhilidzi, M. (2011). Fingerprint Segmentation: An Investigation of Various Techniques and Parameter Study of a Variance-Based Method. International Journal of Innovative Computing, Information and Control. 7 (9), 5313-5326.

Iwnicki, S. (2003). Simulation of wheel-rail contact forces. Fatigue Fracture Engineering Material Structure. 26, 887-900.

Iwnicki, S. (2009). The Effect of Profiles on Wheel and Rail Damage. *International Journal of Vehicle Structures & Systems. 1* (4), 99-104.

Javed, A., Ashfaq Qazi, K., Maqsood, M., & Ali Shah, K. (2012). Efficient Algorithm for Railway Tracks Detection Using Satellite Imagery. *I.J. Image, Graphics and Signal Processing. 11*, 34-40.

Jeong, D., & Orringer, O. (1989). Fatigue crack growth of surface cracks in the rail web. Theoretical and Applied Fracture Mechanics. 12, 45-58.

Jian-hua, Q., Lin-sheng, L., & Jin-gang, Z. (2008). Design of Rail Surface Crack-detecting System Based on Linear CCD Sensor. In *IEEE International Conference on Networking, Sensing and Control. Sanya, China*, Retrieved from http://ieeexplore.ieee.org/document/4525481/.

Jianxi,W., Yude, X., Songliang, L., & Liying, W. (2011). Probabilistic prediction model for initiation of RCF cracks in heavy haul railway. International Journal of Fatigue. 33, 212-216.

Johnson, C.I. (2013). *Image Processing Techniques for the Detection and Characterization of Features and Defects in Railway tracks.* (Doctoral Thesis). Manchester Metropolitan University, Manchester, U.K.

Jones, R. Molent, L. Pitt, S. (2007). *Crack growth of physically small cracks.* Journal of Fatigue. 29, 1658-1667.

Jun, H.K., Fletcher, D.I., Jung, H.S., Lee, G.H., & Lee, D.H. (2011). Calculation of minimum crack size for growth under RCF between wheel and rail. Transactions on Engineering Sciences. 71, 123-134.

Kahn, W. (2013). Image Segmentation Techniques: A Survey. Journal of Image and Graphics. 1 (4), 166-170.

Kandwal, R., Kumar, A., & Bhargava, S. (2014). Review: Existing Image Segmentation Techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 4 (4), 153-156.

Kang, D.H., Kim, J.W., Park, S.Y., & Park, S. (2014). *Non-contact local fault detection of railroad track using MFL technology*. Journal of Korean Society of. Hazard Mitigation. 14(5), 275-282.

Karttunen, K. (2015). Influence of rail, wheel and track geometries on wheel and rail degradation. (Doctoral Thesis). Chalmers University of Technology, Goteborg, Sweden.

Kenderian, S., Djordjevic, B., & Cerniglia, D. (2003). *PCT/US2003/037919*. U.S: The Johns Hopkins University.

Kim, B. & Park, D. (2002). Adaptive image normalization based on block processing for enhancement of fingerprint image. *ELECTRONICS LETTERS. 38* (14), 696-698.

Kim, J.K., & Kim, C.S. (2002). Fatigue cra*ck growth behaviour of rail steel under mode I and mixed mode loadings. Materials* Science and Engineering. 338 (12), 191-201.

Kotsikos, G., & Grasso, M. (2011). Assessment of Fatigue Cracks in Rails. In Transport Research Arena– Europe Conference, Retrieved from www.sciencedirect.com.

Kotsikos, G., & Grasso, M. (2012). *Assesment of Fatigue Cracks in Rails.* Procedia - Social and Behavioral Sciences. 48, 1395-1402.

Kuber, M., Dexit, M. (2014). A Review on Modified Image Enhancement Applications. International Journal on Signal Processing, Image Processing, and Pattern Recognition. 7 (5), 71-78.

Kumar, S. (2006). *A study of rail degradation process to predict rail breaks* (Licentiate thesis). Lulea University of Technology, Division of Operation and Maintenance Engineering (Unpublished).

Kumari, A.V. (2015). A Review of Image Enhancement Techniques. In *International Journal of Trend in Research and Development (IJTRD)*, Retrieved from www.ijtrd.com.

Kundu, T. (2008). *Fundamentals of fracture mechanics*. Boca Raton New York: CRC Pres Taylor and Francis Group.

Lei, L., Wang, C., & Lui, X. (2013). Discrete Wavelet Transform Decomposition Level Determination Exploiting Sparseness Measurement. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 7* (9), 1182-1185.

Lewis, R., & Olofsson, U. (2009). Wheel-rail interface handbook. Cambridge: Woodhead Publishing and CRC Press.

Li, C.H., & Tam, P. (1998). An iterative algorithm for minimum cross entropy thresholding. Pattern Recognition Letters. 19 (8),771–776.

Li, Q., & Ren, S. (2012). A Visual Detection System for Rail Surface Defects. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, 42* (6), 1531-1542.

Livieri, P., & Seagala, F. (2007). Analytical evaluation of J-integral for elliptical and parabolic notches under Mode-I and Mode-II loading. International Journal of Fracture. 148, 57-71.

Livieri, P., & Seagala, F. (2012). Evaluation of stress intensity factors from elliptical notches under mixed mode loadings. Engineering Fracture Mechanics. 81, 110-119.

Lugg, M., Topp, D., TSC Inspection systems, Milton Keynes. (2006). *Recent Developments and Applications of the ACFM Inspection Method and ACSM Stress Measurement Method*. U.K: ECNDT.

Lui, Y., De Dominicis, L., Martin, R.R., & Wei, B. (2013). Illumination Normalization Using Weighted Gradient Integral Images. *Lecture Notes on Software Engineering, 1* (4), 329-333.

Lui, Z., Koffman, A.D., Waltrip, B.C., & Wang, Y. (2013). Eddy Current Rail Inspection Using AC Bridge Techniques. *Journal of Research of the National Institute of Standards and Technology. 118*, 140-149.

Lui, Z., Wang, W., Zhang, X., & Jia, W. (2010). Design of Rail Surface Crack-detecting System Based on Linear CCD Sensor. In *2nd International Asia Conference on Informatics in Control, Automation and Robotics, Beijing, China*, Retrieved from ieeexplore.ieee.org/document/5456793/

Lui, L. (2008). Modelling of mixed mode fatigue crack propagation. (Doctoral Thesis). Vanderbilt University. Nashville Tennessee.

Magel, E. (2011). *Rolling contact Fatigue: A comprehensive Review*. Washington DC, Office of railroad Policy and Development.

Mahmoudi, A. & Regragui, F. (2009). A Fast Segmentation Method for Defect Detection in Radiographic Images of Welds. International Conference on Computer Systems and Applications. Rabat, Morocco. Retrieved from: ieeexplore.ieee.org/abstract/document/5069430/.

Malik, Q. (2013). *Novel Methods of Object Recognition & Fault Detection Applied to Non-Destructive Testing of Rail's Surface During Production* (Doctoral Thesis). Manchester Metropolitan University, Unpublished.

Mandriota, C., Nitti, M., Stella, E., & Distante, A. (2004). Filter-based feature selection for rail defect detection. *Machine Vision and Applications. 15*, 179-185.

Martin, A. & Tosunoglu, S. (2000). Image Processing Techniques for Machine Vision. In *Computer Vision Techniques, Florida International University, Department of Mechanical Engineering Miami Florida.*

Mathworks. (2017). Mathworks support. Retrieved from:
 https://www.mathworks.com/MATLABcentral/fileexchange/?s_tid=gn_mlc_fx

Menotti, D., Najman, L., Facon, J., & Araújo, A. (2007). Multi-Histogram Equalization Methods for Contrast Enhancement and Brightness Preserving. IEEE Transactions on Consumer Electronics. 53 (3), 25-29.

Mostafa, A., Fouad, a., Houseni, M., Allam, N., Hassaniem, A.E., Hefny, H., & Aslanishvili, I. (2016). International Conference on Advanced Intelligent Systems and Informatics. 846-855.

Naik, N. & Mishra, A. (2015). Low Contrast Image Enhancement using Wavelet Transform based Algorithms: A Literature Review. *International Journal of Engineering and Technical Research (IJETR). 3* (6), 123-128.

Network Rail Limited. (2016). *Network rail annual return*. Retrieved from www.networkrail.co.uk/annual-return.

Niblack, W. (1986). An introduction to digital image processing. Prentice-Hall International.

Nixon, S.M., & Aguardo, S.A. (2008). Feature Extraction and Image Processing. 2nd Ed. Academic Press, Oxford, U.K.

Papaelias, M., Lugg, M.C., Roberts, C., & Davis, C.L. (2009). *High-speed inspection of rails using ACFM techniques. NDT and E Internal. 42* (4), 328-335.

Plu, J., Bondeux, S., Boulanger, D., & Heyder, R. (2009). *Application of fracture mechanics methods to rail design and maintenance*. Engineering Fracture Mechanics. 76 (17), 2602-2611.

Popovici, R. (2010). Friction in Wheel-Rail Contacts. (Doctoral Thesis). University of Twente, Enscheda, The Netherlands.

Resendiz, E. (2013). Automated Visual Inspection of Railroad Tracks. *IEEE Transaction on Intelligent Transportation Systems. 14* (2), 751-760.

Resendiz, E., Molina, L., Hart, J., Edwards, R., Sawadisavi, S., Ahuja, N., & Barkan, C. (2010). *Development of a machine vision system for inspection of railway track components*. Proceedings of 12th WCTR, Lisbon, Portugal.

Richard, H.A., & Benitez, K. (1983). Loading device for the creation of mixed mode in fracture mechanics. International Journal of Fracture. 22, 55-58.

Ringsberg, J.W. (2001). Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue. 23* (7), 575-586.

Roessle, M.L., & Fatemi, A. (2000). Strain-controlled fatigue properties of steels and some simple approximations. International Journal of Fatigue. 22, 495–511.

Rovira, A., Roda, A., Marshall, M. B., Brunskill, H., & Lewis, R. (2011). *Experimental and numerical modelling of wheel-rail contact and wear*. Wear. 271 (6), 911-924.

Rowshandel, H., Papaelias, M., Roberts, C., & Davis, C. (2010). *Development of Autonomous ACFM Rail Inspection Techniques.* Poster presented at NDT Birmingham Centre for Railway Research and Education. University of Birmingham, Birmingham, U.K.

Sambo, B., Bevan, A., & Pislaru, C. (2016). A novel application of image processing for the detection of rail surface RCF damage and incorporation in a crack growth model. In: International Conference on Railway Engineering 2016 (ICRE), Brussels, Belgium.

Sangrid, M.D. (2013). The physics of fatigue crack initiation. *International Journal of Fatigue. 57*, 58-72.

Sauvola, J. & Pietikainen, M. (2000). Adaptive document image binarization. Pattern Recognition. 33 (2), 225–236.

Sawadisavi, S., J. Edwards, E. Resendiz, J.M. Hart, C.P.L Barkan, and N. Ahuja. (2009). *Machine-Vision Inspection of Railroad Track.* Proceedings of the TRB 88th Annual Meeting, Washington, DC.

Senthilkumaran, N. & Vaithegi, S. (2016). IMAGE SEGMENTATION BY USING THRESHOLDING TECHNIQUES FOR MEDICAL IMAGES. *Computer Science & Engineering: An International Journal (CSEIJ). 6* (1), 1-13.

Senthilkumaran, N., & Rajesh, R. (2009). Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches. International Journal of Recent Trends in Engineering. International Journal of Recent Trends in Engineering. INFORMATION PAPER. 1 (2), 25-29.

Seo, J., Kwon, S., Jun, H., & Lee, D. (2010). Fatigue crack growth behaviour of surface cracks in rails. Procedia Engineering. 2, 865-872.

Shlyannikov, V.N., Tumanov, A.V. (2011). An inclined surface crack subject to biaxial loading. International Journal of solid and structures. 48, 1778-1790.

Singh Negi, S. & Gupta, B. (2014). Survey of Various Image Enhancement Techniques in Spatial Domain Using MATLAB. *International Journal of Computer Applications. 9* (75), 8-19.

Singh, G.M., Kohli, M.S., & Diwakar, M. (2013). A Review of Image Enhancement Techniques in Image Processing. *HCTL Open Int. J. of Technology Innovations and Research. 5*, 1-13.

Somalraju, S. Murali, V. Saha, G, Vaidehi, V. (2012). *Robust Railway Crack Detection Scheme (RRCDS) Using LED-LDR Assembly.* In *Chennai, Tamil Nadu, India,* Retrieved from http://ieeexplore.ieee.org/document/6206824/.

Song, Z., Yamada, T., Shitara, H., & Takemura, Y. (2011). Detection of Damage and Crack in Railhead by Using Eddy Current Testing. *Journal of Electromagnetic Analysis and Applications. 3*, 546-550.

Srivastava, J.P., Sakar, P.K., & Ranjan, V. (2014). Contact stress analysis in wheel-rail by Hertzian method and Finite Element method. Journal of the Institution of Engineers Series C Mechanical, Production, Aerospace, and Marine Engineering. 40 (32), 140-145.

State Bureau of Quality and Technical Supervision (SBQTS). (2002). Statistical interpretation of data normality tests. Beijing: China Standard Press.

Tanaka, K. (1974). Fatigue crack propagat*ion from a crack inclined to the cyclic tensile axis.* Engineering Fracture Mechanics. 6, 493-507.

Teng, Z., Lui, F., & Zhang, B. (2016). Visual railway detection by superpixel based intracellular decisions. *Multimedia Tools Application. 75*, 2473-2486.

Transport Research Arena– Europe Conference, Retrieved from www.sciencedirect.com. Tyfour, W.R., Beynon, J.H., & Kapoor, A. (1996). *Deterioration of rolling contact fatigue life of pearlitic rail steel due to dry-wet rolling-sliding line contact*. Wear, 197, 255-265.

UIC 725, (2007) Treatment of rail defect, International Union of Railway.

Vasic, G. (2013). *Modelling of Wear and Crack initiation in rails* (Doctoral thesis). Newcastle University, U.K.

Viduad, M. & Zwanenburg, W. (2009). Current situation on rolling contact fatigue – a rail wear phenomenon. In *Swiss Transport Research Conference*, Retrieved from http://litep.epfl.ch/.

Vijaykumar, V.R., & Sangamithirai, S. (2015). Rail Defect Detection using Gabor filters with Texture Analysis. International conference on Signal Processing Communication and Networking. 1-6.

Wang, C., & Ye, Z. (2005). Brightness Preserving Histogram Equalisation with Maximum Entropy: A Variational Perspective. IEEE Transactions on Consumer Electronics. 51 (4), 1326-1334.

Weeks, M. (2007). *Digital Signal Processing Using MATLAB and Wavelets*. Hingham Massachusetts: Infinity Science Press.

Weng, W. & Chen, H. (2015). Detecting Pit Defects on Rail Surface Using A Fast Detection Algorithm Based on Relative Gray Value. In *First International Conference on Information Science*

and Electronic Technology, Retrieved from www.atlantis-press.com/php/paper-details.php?id=16699.

Wilson, A. & Kerr, M. (2012). *TMC 226 Rail Defects Handbook* (1.2 ed.). New South Wales: RailCorp.

Xie, X. (2008). A Review of Recent Advances in Surface Defect Detection using Texture Analysis Techniques. *Electronic Letters on Computer Vision and Image Analysis. 7* (3), 1-22.

Xie, X., & Lam, K. (2006). An efficient illumination normalization method for face recognition*, Pattern Recognition Letters*, 27, (6), 609–617.

Yan, X., Zhang, Z., & Du, S. (1992). Mixed mode fra*cture criteria for the materials with different yield strengths in tension and compression.* Engineering Fracture Mechanics. 42, 109-116.

Yang, Y., Su, Z., & Sun, L. (2010). Medical image enhancement algorithm based on wavelet transform. *ELECTRONICS LETTERS. 46* (2), 1-2.

Yuan, L., & Mi, C. (2013). Influence of crack size on the propagation trend of multiple rail surface cracks under rolling contact fatigue. Journal of Applied Sciences. 13 (9), 4046-4051.

Yuan, X., Wu, L., & Peng, Q. (2015). An improved Otsu method using the weighted object variance for defect detection. *Applied Surface Science. 349*, 472-48.

Zafosnik, B., Ublin, M., & Flasker, J. (2000). Numerical analyses of mixed mode crack propagation using virtual crack extension method. International design conference. Dubrovnik, 1279-1284.

Zaitoun, M.N., & Aqel, M.J. (2015). Survey on Image Segmentation Techniques. Procedia Computer Science. 65, 797-806.

Zakeri, A.J., & Xia, H. (2008). Sensitivity of track parameters on train track dynamic interaction. In Journal of Mechanical Science and Technology. 22, 1299-1304.

Zerbst U. (2003). Application of fracture mechanics to railway components- An overview. Engineering Fracture Mechanics.1-8.

Zhang, W., Zhang, Z., Qi, D., & Liu, Y. (2014). Automatic Crack Detection and Classification Method for Subway Tunnel Safety Monitoring. *Sensors. 14*, 19307-19328.

Zhiyuan, Q., Wiqiang, Z., Zhanmu, Z., Bing, W., Jie, R., & Baoshan, Z. (2002). A Robust adaptive image smoothing algorithm. International Society for Photogrammetry and Remote Sensing. Commission IV, Symposium. Ottawa, Canada, Retrieved from: http://www.isprs.org/proceedings/XXXIV/part4/

Appendices

Appendix A: Basic filter design

A1: Low pass filter

$D(u,v)$ is the distance between the point (u,v) and centre frequency and D_0 is the cut off frequency

$$\text{Ideal:} \qquad H_I(u,v) = \begin{cases} 1 \text{ if } D(u,v) \leq D_0 \\ 0 \text{ if } D(u,v) > D_0 \end{cases}$$

(A1.1)

$$\text{Butterworth:} \qquad H_B(u,v) = \frac{1}{1+[\,D(u,v)/D_0]^{2n}}$$

(A1.2)

$$\text{Gaussian:} \qquad H_G(u,v) = \frac{e^{-D^2(U,V)}}{2D_0{}^2}$$

(A1.3)

A2: High pass filter

$H_I(u,v)$, $H_B(u,v)$, and $H_G(u,v)$ is the Ideal, Butterworth, and Gaussian filter transfer functions respectively.

Ideal:

$$1 - H_I(u,v) \qquad\qquad \text{(A2.1)}$$

Butterworth:

$$1 - H_B(u,v) \qquad\qquad \text{(A2.2)}$$

Gaussian:

$$1 - H_G(u,v) \qquad\qquad \text{(A2.3)}$$

Appendix B: Wavelet functions

Meyer Wavelet

Yves Meyer presented a smooth orthonormal wavelet basis function depicted in Figure A.B1, by defining the Fourier transform of a scaling function described by equation B.1.

$$\varphi(\omega) = \begin{cases} 1 & if \ |\omega| \leq \frac{2}{3}\pi \\ cos\left[\frac{2}{3}v\left(\frac{3}{4\pi}|\omega| - 1\right)\right] & if \ \frac{2}{3}\pi \leq |\omega| \leq \frac{3}{4} \\ 0, & otherwise \end{cases} \tag{B.1}$$

The smooth function v satisfies the condition in equation B.2, and the additional property in B.3.

$$v(t) = \begin{cases} 1, & if \ t \leq 0 \\ 0, & if \ t \geq 1 \end{cases} \tag{B.2}$$

$$v(t) + v(1 - t) = 1 \tag{B.3}$$
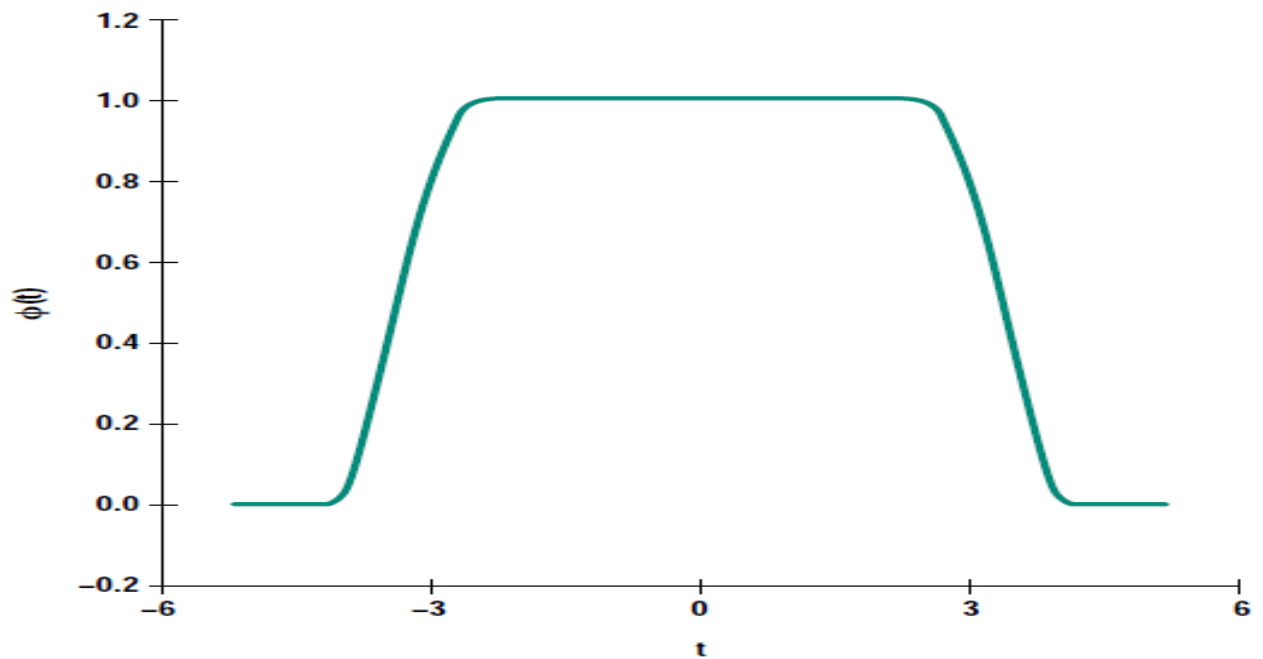
Figure A.B 1: Fourier transform of Mayer wavelet basis function.

The Mayer wavelet function can easily be derived from its Fourier transform defined in equation B.4 below.

$$\Psi(\omega) = e^{j\omega/2} \sum_{l \in Z} \varphi(\omega + 2\pi(2l + 1))\varphi(\omega/2) \tag{B.4}$$

The above expression can be simplified to equation B.5 and the graphical representation is depicted in Figure AB.2 below.

$$\Psi(\omega) = e^{\frac{j\omega}{2}}[\varphi(\omega + 2\pi) + \varphi(\omega - 2\pi)]\varphi(\omega/2) \qquad \text{(B.5)}$$
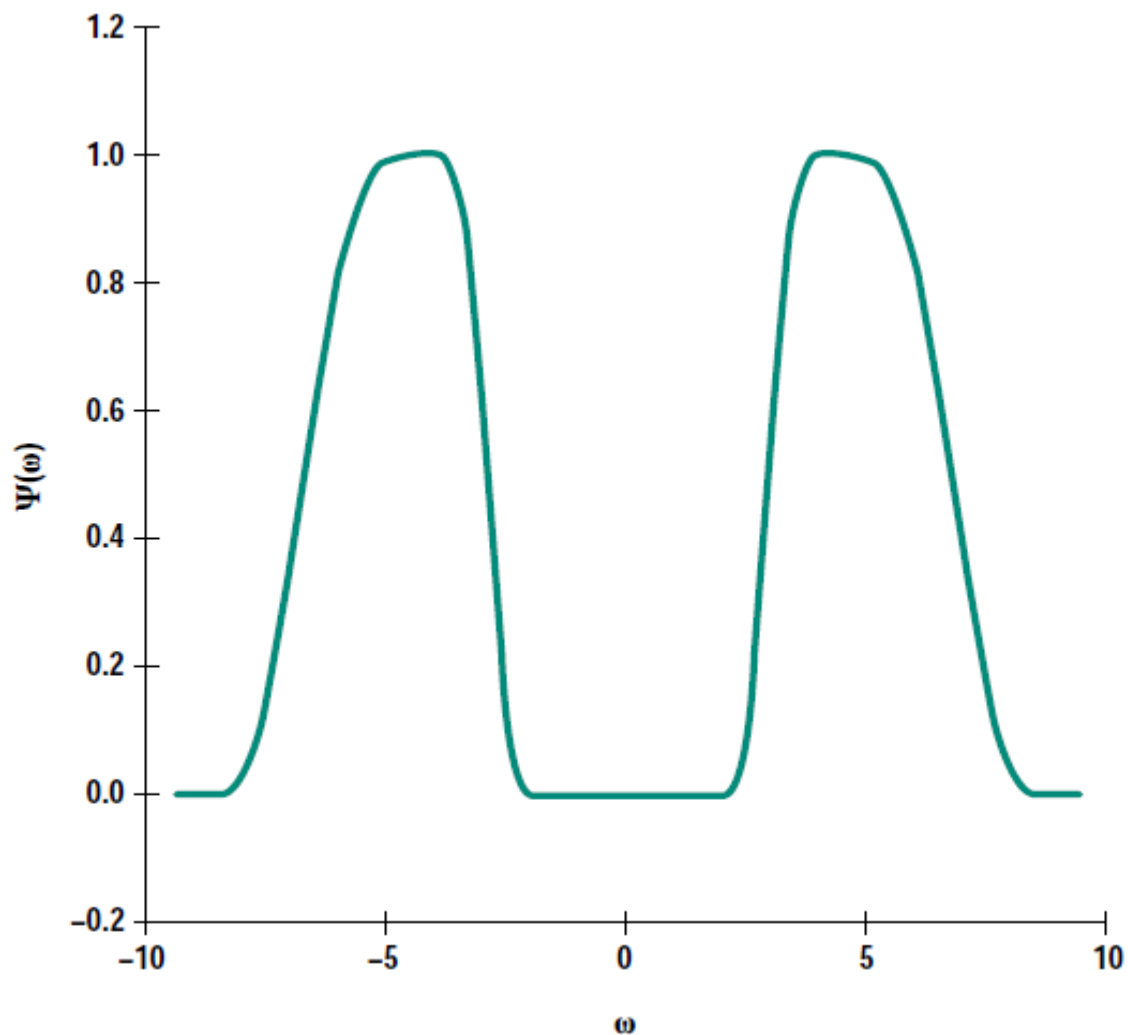
Figure A.B 2: Fourier transform of Mayer wavelet function.

From the Figure above it can be concluded the wavelet function has a finite non zero duration (compactly supported), thus can be inversely transformed in the frequency domain (see Figure AB.3) since Ψ has at least k derivatives where k is finite.
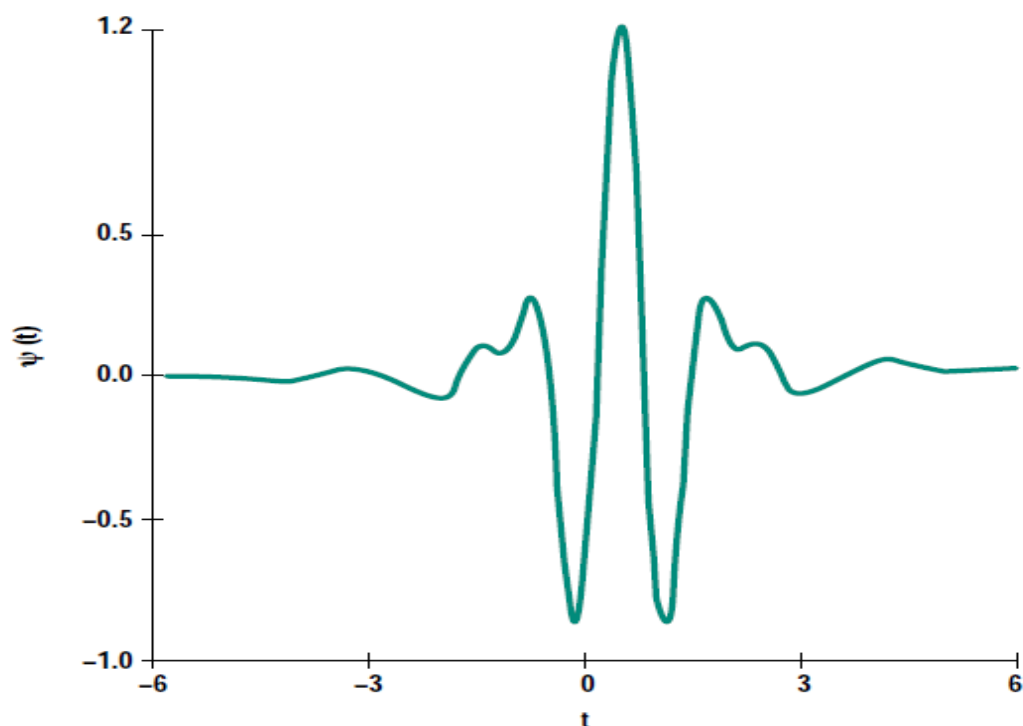
Figure A.B 3: Mayer wavelet function.

Haar Wavelet

This is often considered the simplest wavelet function and can be derived by assuming a box scaling function $\phi(t)$ satisfying equation B.6.

$$\phi(t) = \begin{cases} 1 \; if \; 0 \leq t \leq 1 \\ 0, otherwise \end{cases} \tag{B.6}$$

It is also logical to define the Haar wavelet as orthogonal to its translation and dilation as elaborated in equation B.7 and B.8 below.

$$\psi(t) = \phi(2t) - \phi(2t - 1) \tag{B.7}$$

$$\psi(t) = \begin{cases} 1, if \; 0 < t \leq 1/2 \\ -1, if \; \frac{1}{2} < t \leq 1 \\ 0, otherwise \end{cases} \tag{B.8}$$

Morlet wavelet

It is considered the most frequently applied wavelet function partly because of the flexibility associated with its Fourier transform see equation B.9; best described as a shifted Gaussian such that $\psi(0) = 0$.

$$\psi(\omega) = e^{-(\omega - \omega_o)^2/2} - e^{-\omega^2/2}e^{-\omega_o^2/2} \tag{B.9}$$

The inverse Fourier transform of the equation above is presented in equation B.10 below.

$$\psi(t) = (e^{-i\omega_o t} - e^{-\omega^2/2})e^{-t^2/2} \tag{B.10}$$

The value of ω_o is suggested to be such that the ratio of the first peak of ψ to its second peak is approximately ½, which corresponds to $\omega_o \approx 5$. The results to a negligable second term in equation B.9. The Morlet wwavelet can be considered as a modulated Gaussian waveform with real and imaginary parts depicted inFigure A.B4.

(a)   (b)
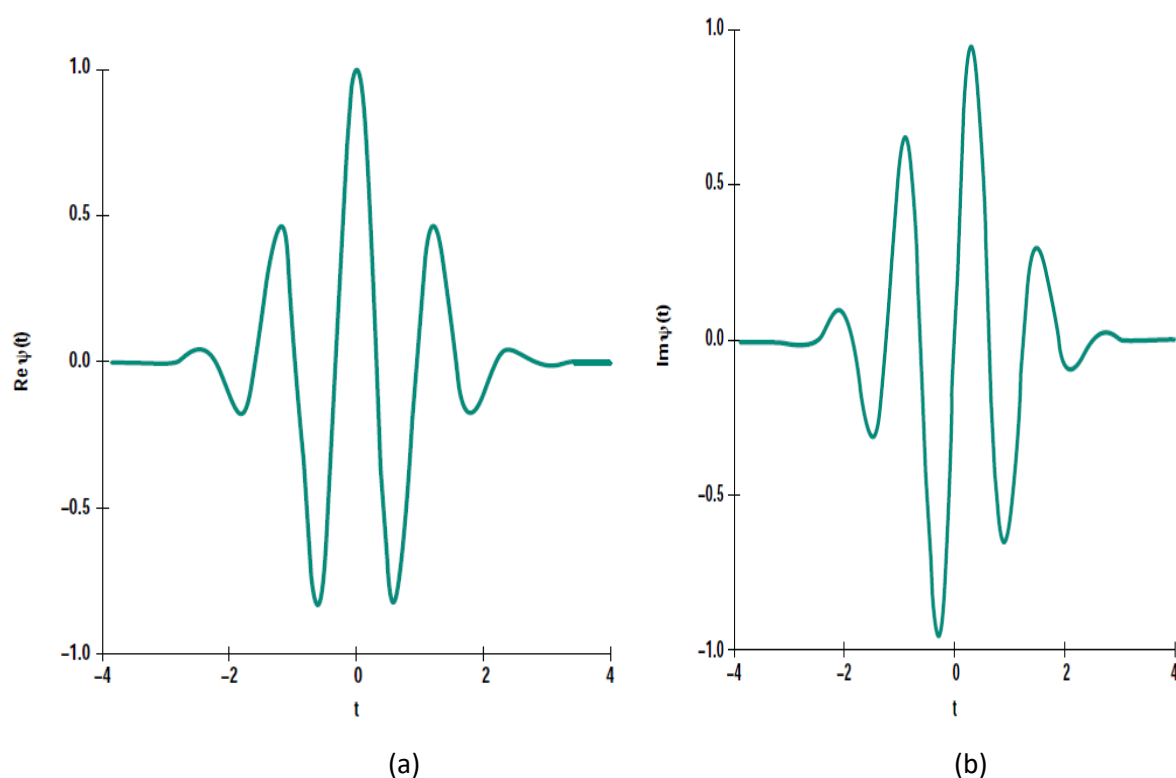
Figure A.B 4: Real (a) and imaginary (b) part of Morlet wavelet.

Daubechies wavelet

This wavelet is somehow unique given that all orthonormal wavelet basis consist of infinitely supported functions, however the Daubechies wavelet is an orthonormal function which is compactly supported by choosing a scaling function with compact support.

The mathematical model for this wavelet family is derived by defining a progression $\{\alpha_k; k \in Z\}$ satisfying the conditions in B11 - B.14 for $N \geq 2$.

$$\alpha_k = 0, if\ k < 0\ or\ K > 2N \tag{B.11}$$

$$\sum_{k=-\infty}^{\infty} \alpha_k \alpha_{k+2m} = \delta om\ , for\ all\ integer\ m \tag{B.12}$$

$$\sum_{k=-\infty}^{\infty} \alpha_k = \sqrt{2} \tag{B.13}$$

$$\sum_{k=-\infty}^{\infty} \beta_k k^m = 0\ , 0 \leq m \leq N - 1 \tag{B.14}$$

Where $\beta_k = (-1)^k \alpha_{-k+1}$, it is worth mentioning that if $N = 1$, then $\alpha_o = \alpha_1 = 1$, which corresponds to the Haar transform.

It is possible to find a scaling function $\varphi(t)$, from the progression of $\{\alpha_k\}$ as described by equation B.15 below.

$$\varphi(t) = \sum_{k=-\infty}^{\infty} \alpha_k \sqrt{2}\, \varphi(2t - k) \tag{B.15}$$

If β_k is defined for condition B.14, equation B.16 (Daubechies wavelet function) is said to be compactly supported.

$$\Psi(t) = \sum_{k=-\infty}^{\infty} \alpha_k \sqrt{2}\, \varphi(2t - k) \tag{B.16}$$

Figure A.B5 below depicts the scaling and wavelet function of equation B.15 and B.16 respectively.
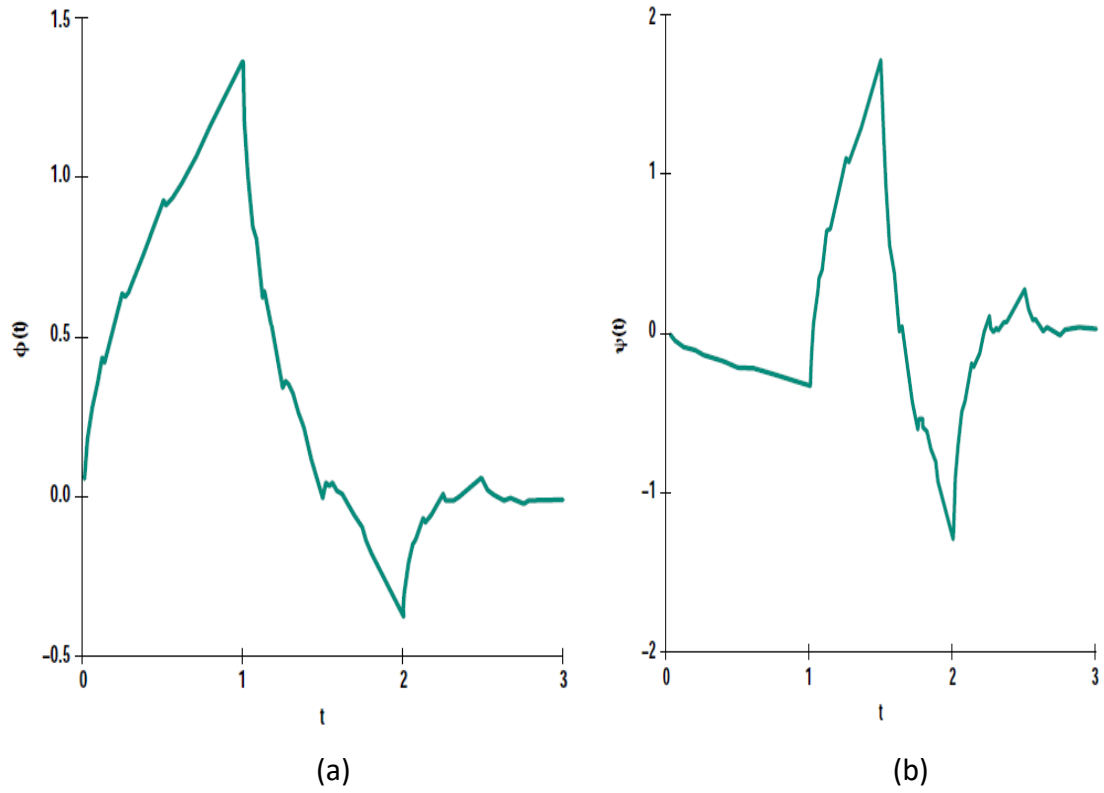
Figure A.B 5: Daubechies (a) scaling and (b) wavelet function

Appendix C: Image Processing Algorithm

Appendix C1: Image processing data set 2
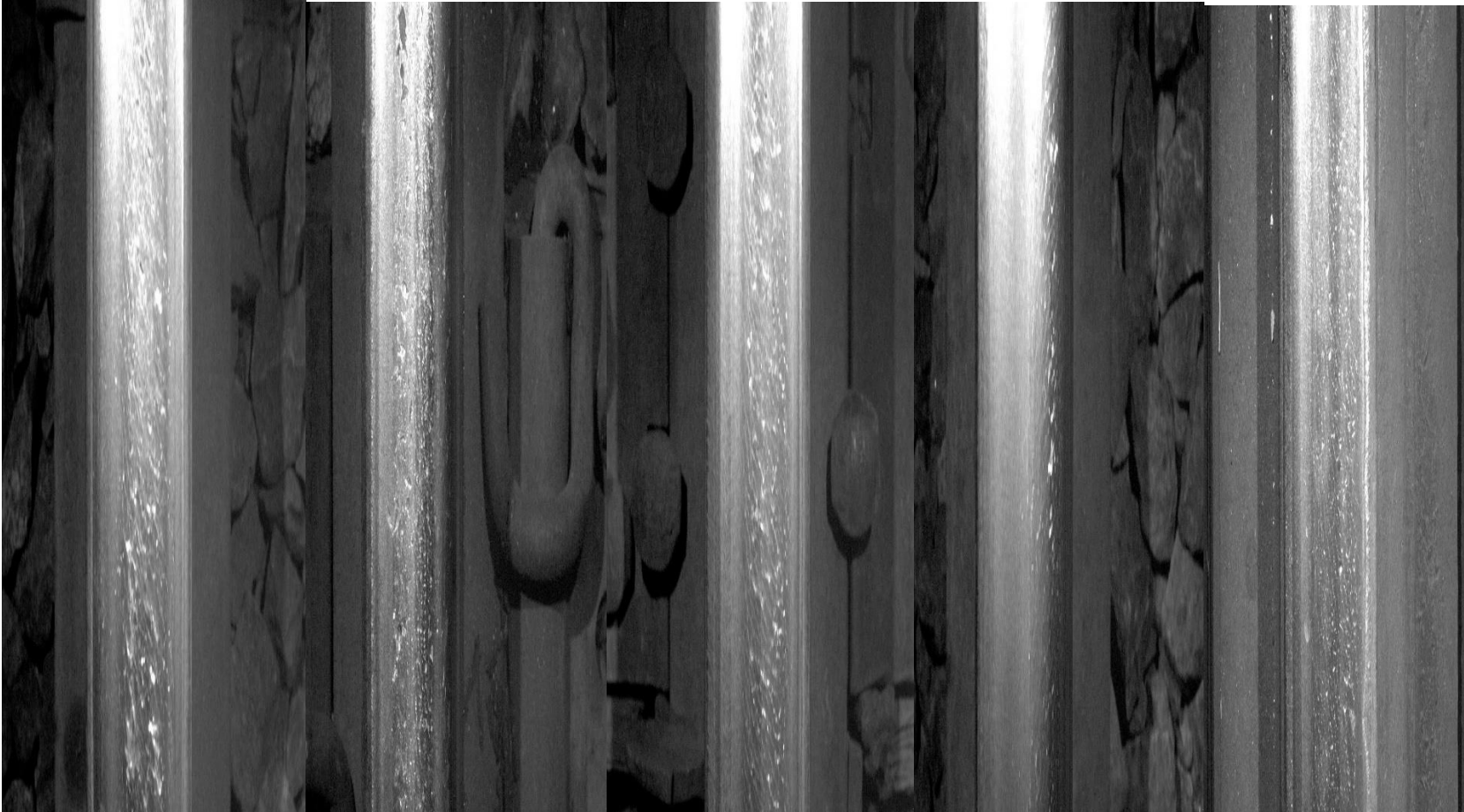
Figure A.C1. 1: Shows heavily damaged samples of data set 2 (NRC).
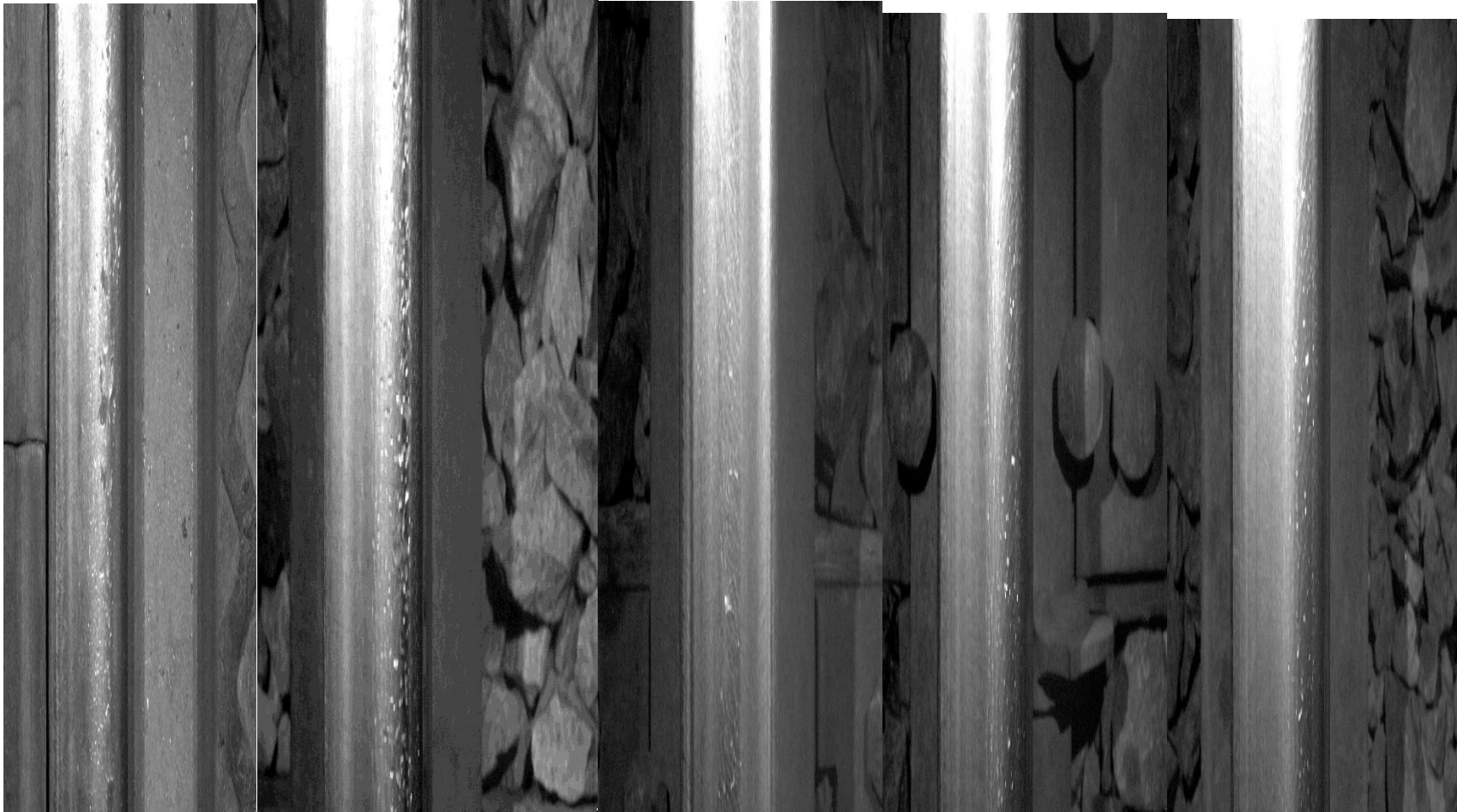
Figure A.C1. 2: Shows moderately damaged samples of data set 2 (NRC)
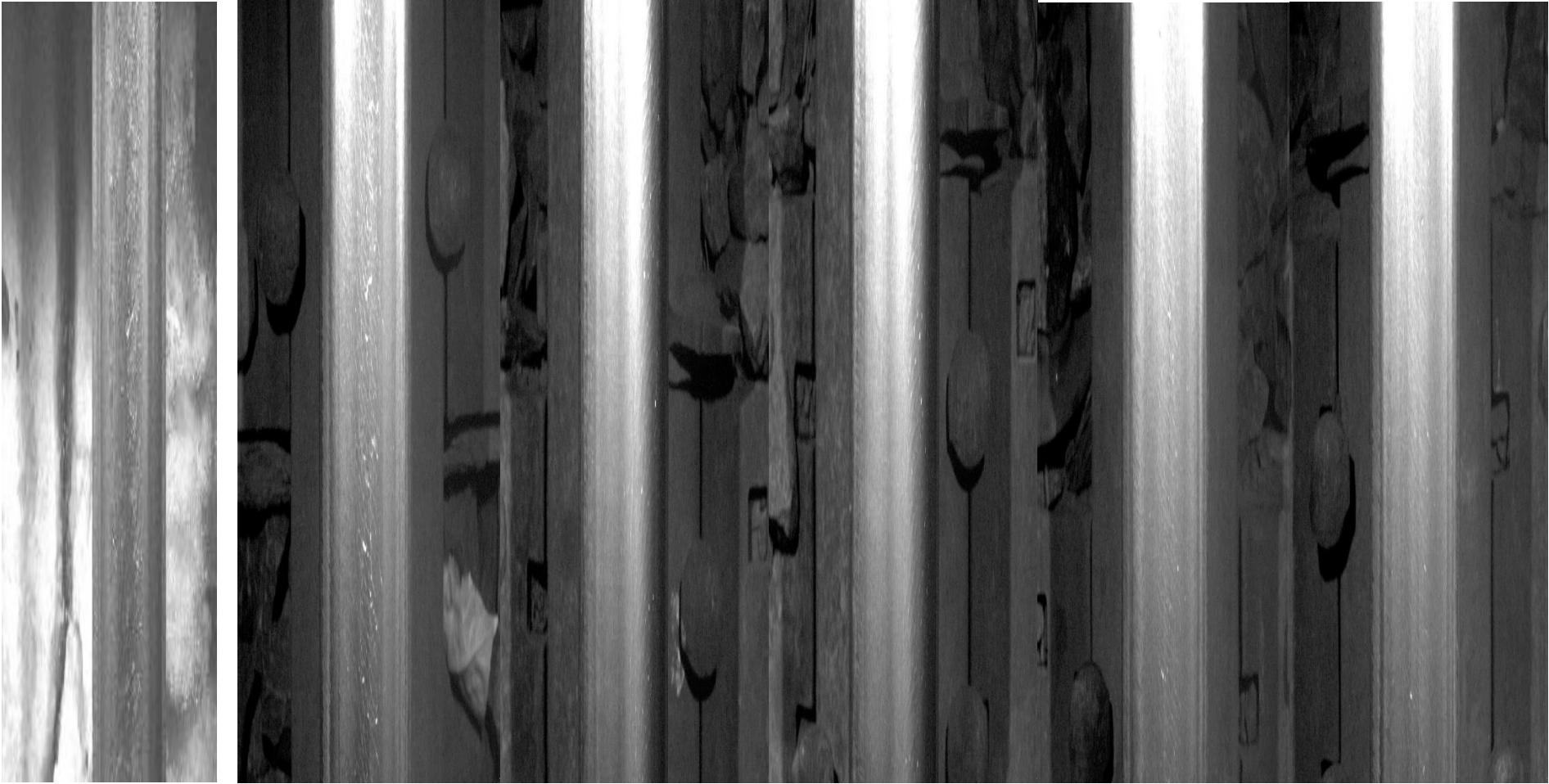
Figure A.C1. 3: Shows lightly damages samples of data set 2 (NRC)
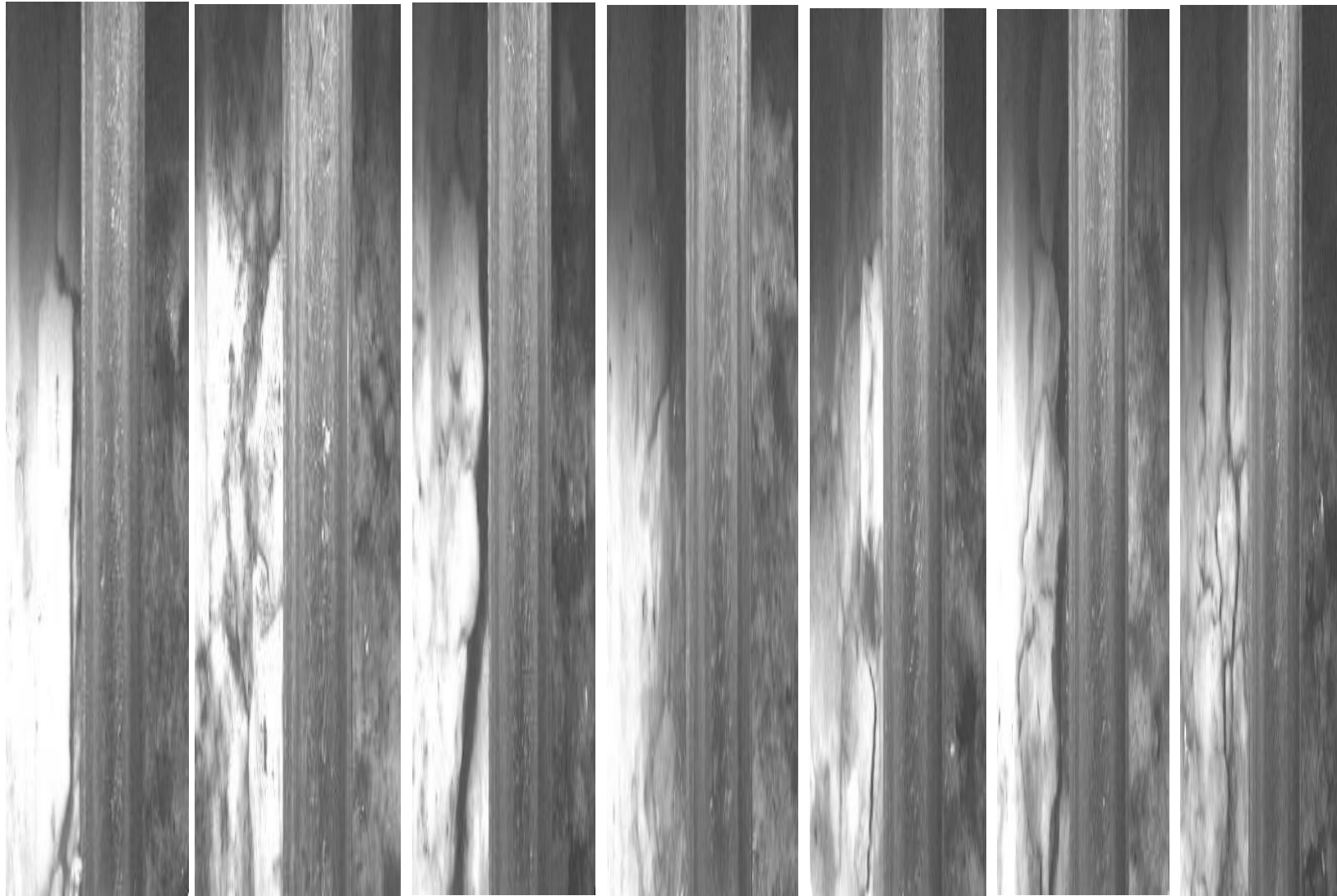
Appendix C2: Image processing data set 3
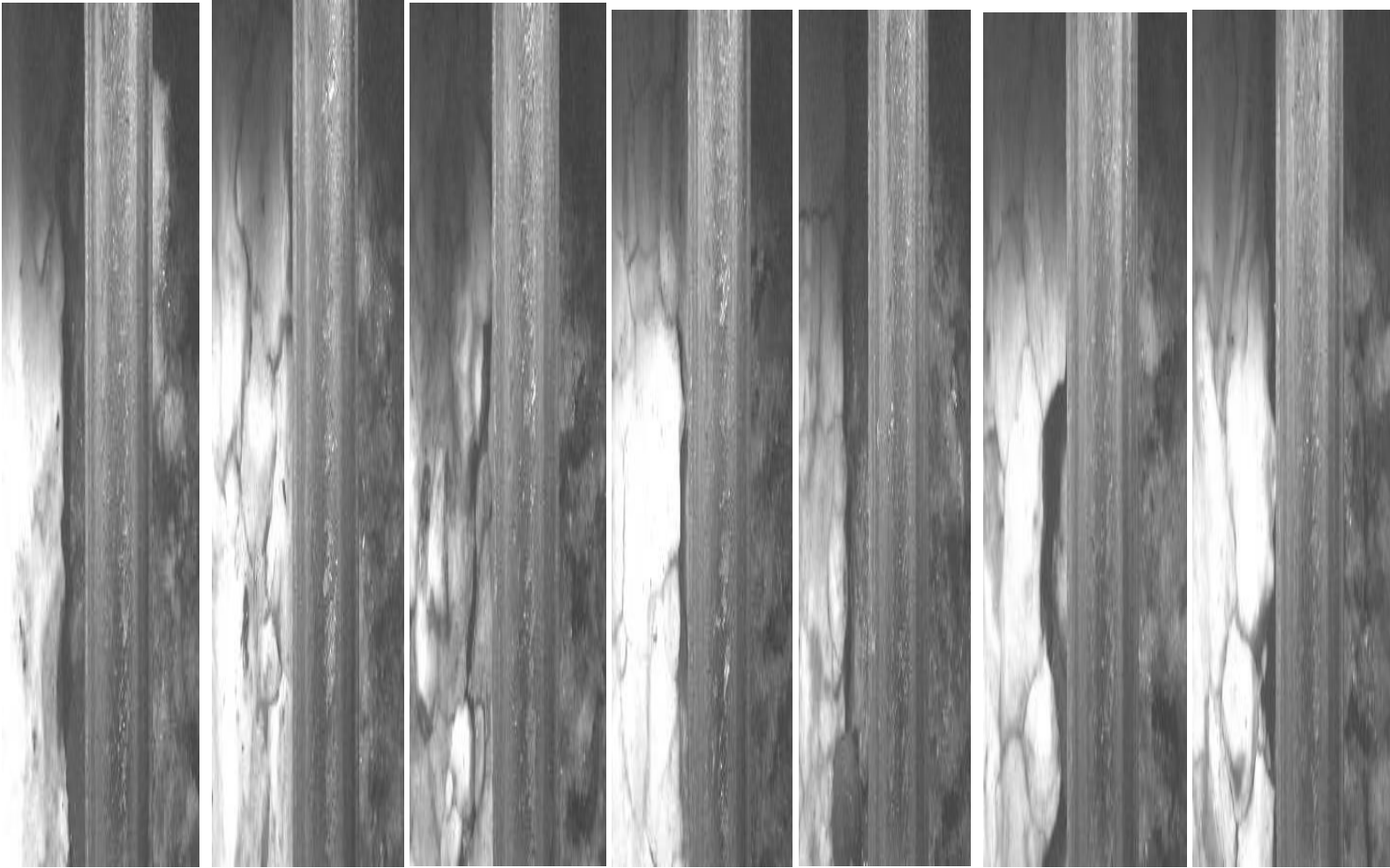
Figure A.C2. 1: Shows left rail field acquired samples of data set 3. (ICRI)
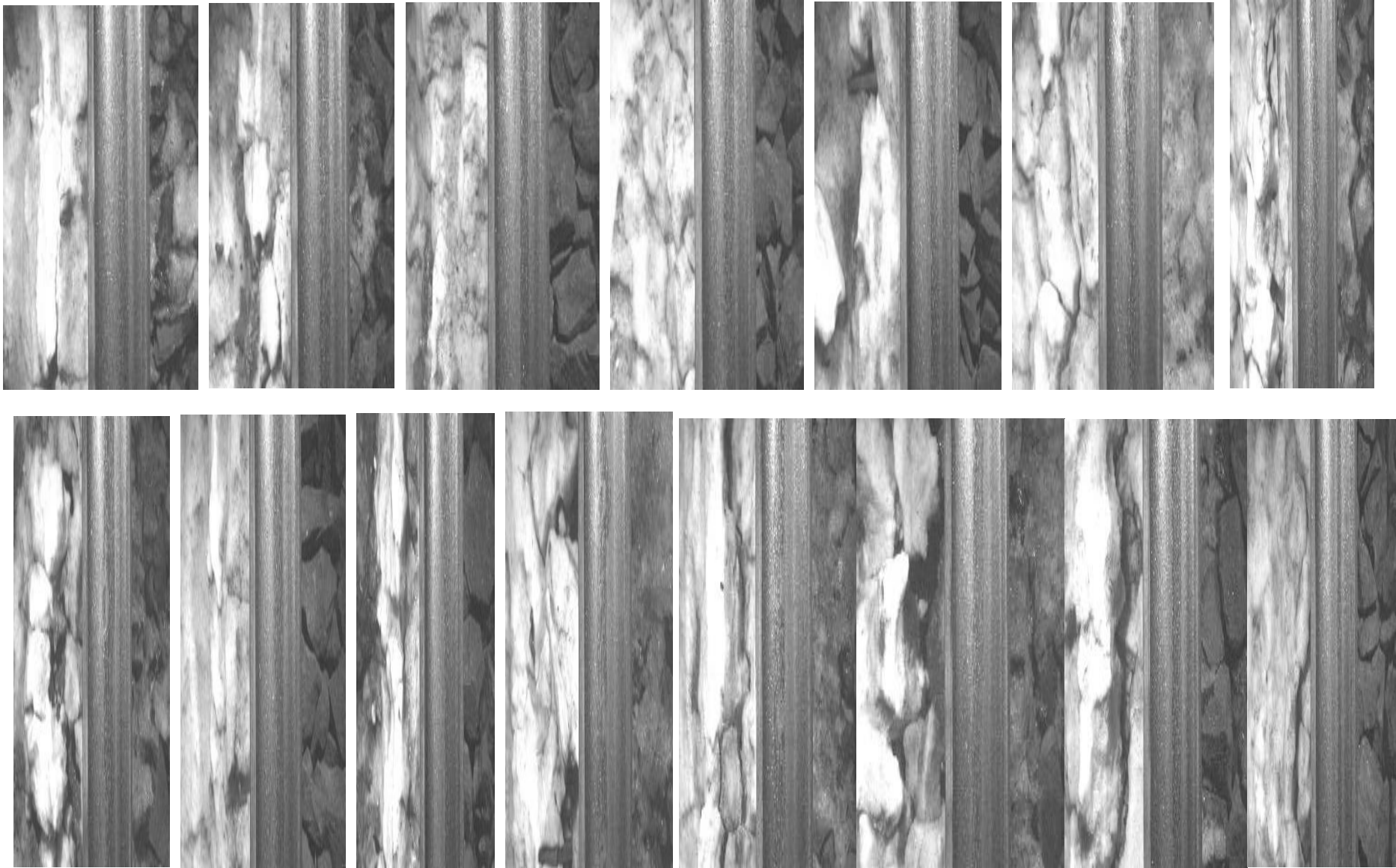
Figure A.C2. 2: Shows right rail field acquired samples of data set 3. (ICRI)

Appendix C3: Function definitions for enhancement

```matlab
%%%%%%%%%%Image processing algorithm%%%%%%%%%%%%%%%

function [ Ienh,Seg,FE] = IMP_FINAL(E1,S1,F1)% Function definition


%%%%%%%%%%%%%%%%%%Enhancement%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

if E==1% if enhancement option is specified to 1 perform LMAF enhancement
RM=0;% specify required mean of input image
RV=1;% specify required variance of input image
bsln=64; % specify LN block size
W=64; % specify LMAF block size
i3=NL(double(i21),RM,RV,bsln); % LN function call

[ Ienh ] = LMAF( IM,W ); % LMAF function call
Figure,imshow(Ienh); % display LMAF image (i3)
title('Linear moving averaging filtered image'); % title of LMAF image


elseif E==2% else if enhancement option is specified to 2 perform VM
            %enhancement
RM=0.2; % specify required mean of input image
RV=1.7; % specify required variance of input image
bsln=64; % specify LN block size
W=5; % specify VM block size
i3=NL(double(i21),RM,RV,bsln); % LN function call


[Ienh] = VisbMeth(i3,W); % VM function call
Figure,imshow(Ienh); % display VM image (Ienh)
title('Visibility measure enhanced image'); % title of VM image


elseif E==3% else if enhancement option is specified to 3 perform LN
            %enhancement

RM=0.2; % specify required mean of input image
RV=1.7; % specify required variance of input image
bsln=64; % specify LN block size

[Ienh]=NL(double(i21),RM,RV,bsln); % LN function call
Figure,imshow(Ienh); % display LN image (Ienh)
title('Locally normalized image'); % title of LN image

elseif E==4% else if enhancement option is specified to 4 perform FFT
            %enhancement

RM=0.2; % specify required mean of input image
RV=1.7; % specify required variance of input image
bsln=64; % specify LN block size
bsfft=64; % specify FFT block size

i3=NL(double(i21),RM,RV,bsfft); % LN function call

k=0.25;% FFT enhancement factor
```

274

```matlab
[Ienh]=fftenhance(i3,k,bs1f);% FFT function call
Figure,imshow(Ienh); % display FFT image (Ienh)
title('FFT enhanced image'); % title of FFT image
end



%%%%%%%%%%%%%%%%%%%%%Segmentation%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%



if S==1% if segmentation option is specified to 1 perform Wolf segmentation
bss1=32;% Wolf block size

[Seg1] = Wolf(Ienh,bss1,[],1);% Wolf function call
Figure,imshow(Seg1);% display Wolf segmented image
title('Wolfs thresholding algorithm');%title of Wolf segmented image

elseif S==2% else if segmentation option is specified to 1 perform OGT
            %segmentation

[Seg2] = OGT(double(Ienh)); % OGT function call
Figure,imshow(Seg2); % display OGT segmented image
title('OGT thresholding method'); %title of OGT segmented image



elseif S==3% else if segmentation option is specified to 1 perform AIT
            %segmentation

[Seg3] = AIT(Ienh,opthr(double(Ienh))); % AIT function call
Figure,imshow(Seg3); % display AIT segmented image
title('Automatic iterative threshold'); %title of OGT segmented image
end

%%%%%%%%%%%%%%%%%%%%%%Feature
extraction%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%



if F==1% if feature extraction option is specified to 1 perform SDHSD
method

[FE, count] = SDHSD3(Seg,FE_thresh) % SDHSD function call by specifying
standard
                                %deviation threshold
Figure,imshow(FE1); % display SDHSD image
title('SDHSD image'); %title of SDHSD image



elseif F==2% else if feature extraction option is specified to 2 perform
            %Gabor method

fmax=50;% filter frequency
E=0.009;% dissipated energy threshold

[FE2] = Gabor_F(Seg2,fmax,E)% Gabor function call
Figure,imshow(FE2); % display Gabor processed image
title('Texture based feature extraction'); % title of Gabor processed image

elseif F==3% else if feature extraction option is specified to 3 perform
```

275

```matlab
            %Morphological feature extraction method
    Max_thr_area=800;%specify maximum allowable defect area defect[mm^2]
    Min_thr_area=300; %Specify minimum allowable defect area [mm^2]

[ FE3, count ] = MORPHFEXTR(Seg,Max_thr_area,Min_thr_area) %Morphological
feature
                                              %extraction
                                              %function call
Figure,imshow(FE3); %display morphological feature extraction image
title('Morphological based feature extraction'); %title of morphological
feature
                                    %extraction image
end
end

%%%%%%%%%%%%%%%%% LMAF %%%%%%%%%%%%%%%%%%%%%%%%%%%%


function [ Ienh ] = LMAF( IM,W )

Im=im2double(IM);% double precision representation of intensity values
[w h]=size(Im);% determine size of image

for j=1:W:w%column loop window pixel intensities window wise
for i=1:W:h %row loop window pixel intensities window wise

if i+(W-1) <= h & j+(W-1) <= w %if pixel within specified window size
    if ~isa(Im(j:j+(W-1),i:i+(W-1)),'double'), Im(j:j+(W-1),i:i+(W-1)) =
double(Im(j:j+(W-1),i:i+(W-1))); %ensure double precision for each pixel
location
                                  %within block
    end %end if

    im2=Im(j:j+(W-1),i:i+(W-1));
    im2=double(im2);
    N(j:j+(W-1),2)=[j:j+(W-1)]'; %loop window pixels in new location
        end %end if


for jj=1:w%loop to select seed pixel
for ii=1:h%loop to select seed pixel
for jj2=1:W
    if N(jj2,2)-(ceil(N(jj2,2)/2))<=1
     a(jj2,1)=1;
     b(jj2,1)=N(jj2,2)+ceil(N(jj2,2)/2);%end point of centralization


    elseif N(jj2,2)+ceil(N(jj2,2)/2)>=W
     a(jj2,1)=N(jj2,2)-(ceil(N(jj2,2)/2));%start point for centralization
                                       %selected pixel by N/2 and N/2+1
    b(jj2,1)=W;%end point of centralization

    else
        a(jj2,1)=N(jj2,2)-ceil(N(jj2,2)/2);%start point for
centralization
                                       %selected pixel by N/2 and
                                       N/2+1
        b(jj2,1)=N(jj2,2)+ceil(N(jj2,2)/2);%end point of centralization
```

```matlab
    end %end if

for jj3=a(jj2,1):b(jj2,1) % loop for collecting elements to be averaged
    LMF2= Im(jj3,ii);
end
LMF2=LMF2';
MM(jj2)=std(LMF2);
del(jj2)=MM(jj2)*1.001;

    end%end for loop

LMF(jj,ii)=sum(sum(LMF2))/N(jj2,2);%averaging


 Imdel(jj,ii)=LMF(jj,ii)-del(jj2); %perform white noise treatment by offset
of del
 Icon(jj,ii)=Imdel(jj,ii)-Im(jj,ii);   %obtain relative gray value image



 if Icon(jj,ii)>=0 %thresholding for Icon>0=Icon while Icon<0=0 try loops
instead?

    I_LMF(jj,ii)=Icon(jj,ii);
 else
    I_LMF(jj,ii)=0;
 end%end if

 Ienh(jj,ii)=255-((I_LMF(jj,ii)*255)/Icon(jj,ii)); %enhanced output

end%end for loop
end%end for loop
end%end for loop
end%end for loop
end %end function
%%%%%%%%%%%%%%%%%% Visibility Measure %%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [Venh] = VisbMeth(i21,W)

im=im2double(i3); % double precision representation of image


[w,h] = size(im); % determine size of image
w=floor(w/W)*W; % round step size to fit image size
h=floor(h/W)*W; % round step size to fit image size

vm=zeros(size(i21)); % initialize with zeros



for i=1:W:w % loop to extract window sample
for j=1:W:h %loop to extract window sample

if i+(W-1) <= w & j+(W-1) <= h
    if ~isa(im(i:i+(W-1),j:j+(W-1)),'double'), im(i:i+(W-1),j:j+(W-1)) =
double(im(i:i+(W-1),j:j+(W-1)));
    end
    im2=im(i:i+(W-1),j:j+(W-1));%create sub image in the form of im2
```

277

```matlab
end
    mu= mean(im2(:)); %mean of pixel neighbourhood

        vm1=im2- mu; % visibility measure equation
        vm2=im2+ mu; % visibility measure equation
        vm3=vm1./vm2; % visibility measure equation



        vm(i:i+(W-1),j:j+(W-1))= vm3;% %standard deviation of pixel
neighborhood

[w1 h1]=size(vm); % size of VM determined intensities
w1=floor(w1/W)*W; % round step size to fit image size
h1=floor(h1/W)*W; % round step size to fit image size

 for ii=1:w1% loop to fill each intensity value of the original image
 for jj=1:h1% loop to fill each intensity value of the original image
        if im(ii,jj)<=mu % VM function
Venh(ii,jj)=vm(ii,jj); % final enhanced image

        else
        Venh(ii,jj)=0; % final enhanced image

            end% if loop
        end% end loop
 end% end loop
end% end loop
end% end loop


end % end function

%%%%%%%%%%%%% Local normalize %%%%%%%%%%%%%%%%%%%%%%

function n = NL(im, reqmean, reqvar,W)

[w,h] = size(im);
w=floor(w/W)*W; % round step size to fit image size
h=floor(h/W)*W; % round step size to fit image size
n=zeros(size(im)); % initialise output to zero


    if ~(nargin == 1 | nargin == 4)
       error('No of arguments must be 1 or 4');
    end

    if nargin == 1   % Normalise 0 - 1
    if ndims(im) == 3        % Assume colour image
        hsv = rgb2hsv(im);
        v = hsv(:,:,3);
        v = v - min(v(:));    % Just normalize value component
        v = v/max(v(:));
        hsv(:,:,3) = v;
        n = hsv2rgb(hsv);
    else                     % Assume greyscale
        if ~isa(im,'double'), im = double(im); end
        n = im - min(im(:));
        n = n/max(n(:));
    end
```

```matlab
    else  % Normalise to desired mean and variance

    if ndims(im) == 3        % colour image?
        error('cannot normalise colour image to desired mean and
variance');
    end

for i=1:W:w % loop to traverse block wise
for j=1:W:h% loop to traverse block wise
    if i+(W-1) <= w & j+(W-1) <= h
    if ~isa(im(i:i+(W-1),j:j+(W-1)),'double'), im(i:i+(W-1),j:j+(W-1)) =
double(im(i:i+(W-1),j:j+(W-1))); end
    im2=im(i:i+(W-1),j:j+(W-1)); % create sub image of the block
    im(i:i+(W-1),j:j+(W-1)) = im(i:i+(W-1),j:j+(W-1)) - (mean(im2(:)));
    im(i:i+(W-1),j:j+(W-1)) = im(i:i+(W-1),j:j+(W-1))/std(im2(:));      %
set
                                                        %block to
                                                        %desired
                                                        %mean and
                                                        %standard
                                                        %deviatio
                                                        n %value


    n(i:i+(W-1),j:j+(W-1)) = reqmean + im(i:i+(W-1),j:j+(W-
1))*sqrt(reqvar);
    end% end if
end% end for loop
end% end for loop
end% end function


%%%%%%%%%%%%% FFT enhancement  %%%%%%%%%%%%%%%%%%

function [final]=fftenhance(image,f,W)

I = 255-double(image); % normalize image to 1

[w,h] = size(I); % extract size of image

w1=floor(w/W)*W; % round step size to fit image size
h1=floor(h/W)*W; % round step size to fit image size
final=zeros(size(I)); % initialize output to zero
inner = zeros(size(I)); % initialize normalized intensity variable to zero
to


for i=1:W:w1% FFT block loop
   for j=1:W:h1% FFT block loop

      a=i+(W-1); % FFT block size
      b=j+(W-1); % FFT block size
      F=fft2( I(i:a,j:b) ); % transform image block to frequency domain

      factor=abs(F).^f; % apply enhancement factor in frequency domain

      block = abs(ifft2(F.*factor)); % use inverse FFT to recapture image
```

```matlab
        larv=max(block(:));
        if larv==0
            larv=1;
        end; %end if

        block= block./larv;
        inner(i:a,j:b) = block;
    end; %end loop
end; %end loop

final=inner*255; % change intensity to range 255-0

final=histeq(uint8(final)); % perform histogram equalization
end; % end function
```

Appendix C4: Function definition for segmentation

```matlab
%%%%%%%%%%%%%%%%%%%%% Wolf segmentation %%%%%%%%%%%%%%%%%%%

function [Seg1] = WOLF(im, W)

    K1=1; %Wolf's constant
    K2=1; %Wolf's constant
    Gama=0.001; %Wolf's constant

Im=im2double(im);% double precision representation of intensity values
[w h]=size(Im);% determine size of image

for j=1:W:w%column loop window pixel intensities window wise
for i=1:W:h %row loop window pixel intensities window wise

if i+(W-1) <= h & j+(W-1) <= w %if pixel within specified window size
    if ~isa(Im(j:j+(W-1),i:i+(W-1)),'double'), Im(j:j+(W-1),i:i+(W-1)) =
double(Im(j:j+(W-1),i:i+(W-1))); %ensure double precision for each pixel
location
%within block
    end %end if

    im2=Im(j:j+(W-1),i:i+(W-1));
    im2=double(im2); % create sub image one


m=(mean(im2(:)));% block mean intensity
M=min(im2(:));% minimum scalar value within the block
S1= std(im2(:));
%R=max(S);
end %end if

im3=Im(j:j+(W/2-1),i:i+(W/2-1));
    im3=double(im3); % create sub image 2
S2= std(im3(:));
end %end for
end %end for
Twolf=(1-K1(S1/S2)^Gama)*(m+K2(S1/S2^Gama)*m-M)+K2(S1/S2)^Gama*M; %Wolf
%threshold function
      Seg1 = im2bw(im2, Twolf); % convert gray image to binary
end %end function
```

```matlab
%%%%%%%%%%%%%%%%%%%% OGT segmentation %%%%%%%%%%%%%%%%%%

function [Seg2] = OGT(im, To)

    Wo=count(im>To)/count(im>0); %background occurrence probability
    W1=count(im<To)/count(im>0); %object occurrence probability
    Uo=sum(sum(find(im<To)))/count(im<To); %background mean
    U1=sum(sum(find(im>To)))/count(im>To); %object mean
    Ta=0.001*To;

if (Wo*W1)*(Uo-U1)^2==max(WoW1(Uo-U1^2))
T=To+ta; % corrected threshold
Seg2=im2bw(im,T); % binary segmentation
end %end if
end %end function




%%%%%%%%%%%%%%%%%%%%%%% AIT segmentation %%%%%%%%%%%%%%%%%%%

function [Seg3] = AIT(im)

Im=im2double(im);% double precision representation of intensity values
[w h]=size(Im);% determine size of image

col_c=floor(w/10); % corner column index
rows_c=floor(h/10); % corner row index


corners=[I(1:rows_c,1:col_c); I(1:rows_c,(end-col_c+1):end);...
         I((end-rows_c+1):end,1:col_c);I((end-rows_c+1):end,(end-
col_c+1):end)];
                                       %corner pixel intensity


T=mean(mean(corners)); % mean corner pixel intensity (initial threshold)

while new_T!=T

mean_obj=sum(sum( (I>T).*I ))/length(find(I>T)); % object mean intensity

mean_backgnd=sum(sum( (I<=T).*I ))/length(find(I<=T)); % background mean
                                      %intensity

    new_T=(mean_obj+mean_backgnd)/2; % new threshold is calculated as the
                             %mean of the previous results.


if(new_T==T)
     break; % iteration starts only if the threshold has changed.

  else
     T=new_T; % optimal threshold value
end %end if
end %end while

Seg3=im2bw=(im,T);
```

```
end %end function
```

Appendix C5: Function definitions for feature extraction

```
%%%%%%%%%%%%% SDHSD %%%%%%%%%%%%%%%%%%%%

function [count  DL] = SDHSD3(im,FE_thresh)
 im7=im;
[labeledImage maxlabel2] = bwlabel(im7); %label blobs
[nL1, minLabel, maxLabel1] = renumberregions(labeledImage);%Re-number
regions
[nL2, maxlabel2] = makeregionsdistinct(nL1, 8);%Ensures labeled segments is
                                             %distinct
count=0; %initialize count for detected features

for blobIndex=1:maxlabel2 %loop for each detected defect
boundaries = bwboundaries(nL2==blobIndex); %extract blob boundary
BB=cell2mat(boundaries); %save boundary coordinates in variable BB


x = BB(:, 2); % x coordinate of boundary
y = BB(:, 1); % y coordinate of boundary


% center of each blob
xc= (1/length(x))*sum(sum(x));
yc= (1/length(x))*sum(sum(y));


%modify max distance function to distance between each (xi,yi) pixel and
%(xc,yc)

    for k = 1 : length(x)
        sq=sqrt( (x(k) - xc) .^ 2 + (y(k) - yc) .^ 2 );
        di(k) = ceil(sq);

        Ndi=sum(di(:) == ceil(sqrt( (- xc) .^ 2 + (- yc) .^ 2 )));
        pi=Ndi/length(x);
    end% for loop

mudi=sum(sum(di(:)))/length(x); %average distance of all pixels

stdsdh(blobIndex)= ceil(sqrt((1/length(x))*(pi-mudi)^(2))); %standard
deviation

if stdsdh(blobIndex)<=FE_thresh
count=count+1; %increase detected defect counter
end% for loop
end% for function
%%%%%%%%%%%%% Gabor %%%%%%%%%%%%%%%%%%%%

function [R] = Gabor_F(img,fmax,T)

u=5*4;% ellipse minor axis
v=8*4;% ellipse major axis
```

```matlab
m=32;n=32; %block size
fmax = 50;% filter maximum frequency
gama =sqrt(0.0005);%Gabor filter lateral sharpness control parameter
eta =sqrt(0.0005); %Gabor filter longitudinal sharpness control parameter


for i = 1:u
    fu = fmax/((sqrt(2))^(i-1));
    alpha = fu/gama; %Gabor filter lateral sharpness control parameter
    beta = fu/eta; %Gabor filter longitudinal sharpness control parameter

    for j = 1:v

        tetav = ((j-1)/v)*pi;

        for x = 1:m
            for y = 1:n

 xprime = (x-((m+1)/2))*cos(tetav)+(y-((n+1)/2))*sin(tetav); %filter
                                                             %orientation
 yprime = -(x-((m+1)/2))*sin(tetav)+(y-((n+1)/2))*cos(tetav); %filter
                                                             %orientation
gFilter(x,y) =(fu^2/(pi*gama*eta))*exp(-
((alpha^2)*(xprime^2)+(beta^2)*(yprime^2)))*exp(1i*2*pi*fu*xprime);
%spatial
                                               %filter
                                          representation


            end% end for loop
        end% end for loop
    end% end for loop
end% end for loop

 D = conv2(img,gFilter,'same'); %filtration by convolution
E=D.^(2); %magnify filtered image
[w h]=size(D);

for iii=1:h
    for jjj=1:w
if T>abs(E(jjj,iii))%use energy threshold for binarisation of detected
defects
    R(jjj,iii)=1;%object pixel
else %mod
    R(jjj,iii)=0;%background pixel
end% end if
    end% end for loop
end% end for loop


end% end function


%%%%%%%%%% Morphological feature extraction %%%%%%%%%%%%

function [ count] = MORPHFEXTR(Seg2)

[m n]=size(Seg2); %size of the original image
```

283

```matlab
[labeledImage maxlabel2] = bwlabel(Seg2); % label blob
[nL1, minLabel, maxLabel1] = renumberregions(labeledImage);%RENUMBERREGIONS
[nL2, maxlabel2] = makeregionsdistinct(nL1, 8);%Ensures labeled segments is
                                              %distinct

%extract blob boundary cordinates
boundaries = bwboundaries(labeledImage);
numberOfBoundaries = size(boundaries, 1);
count=0; %defect counter initialized to zero

for blobIndex=1:maxlabel2%loop for each detected defect
 Im = (nL2==blobIndex);

%extract blob area and orientation
Measurements(blobIndex) = regionprops(Im, 'Area','Orientation');
 Am(blobIndex) = getfield(Measurements(blobIndex), 'Area');
 Am(blobIndex)=Am(blobIndex)*((150/926).^2);%Calibration of detected area

Om(blobIndex) = getfield(Measurements(blobIndex), 'Orientation');
    thisBoundary = boundaries{blobIndex};
    x = thisBoundary(:, 2); % x = columns.
    y = thisBoundary(:, 1); % y = rows.

% Find which two bounary points are farthest from each other.
    maxDistance = -inf;
    for k = 1 : length(x)
        distances = sqrt( (x(k) - x) .^ 2 + (y(k) - y) .^ 2 );
        [thisMaxDistance, indexOfMaxDistance] = max(distances);
        if thisMaxDistance > maxDistance
            maxDistance = thisMaxDistance;
            index1 = k;
            index2 = indexOfMaxDistance;
        end% end if
    end% end for loop

    DL(blobIndex)=maxDistance; % defect length

% condition for true area of blob
if ((mean(Am)-0.5*mean(Am))<Am(blobIndex)&& Am(blobIndex)<max(Am))
   T_A_loc(blobIndex)=1;
else
    T_A_loc(blobIndex)=0;
end

%defect orientation must not be perfectly parallel or perpendicular
if  (0~=Om(blobIndex) && Om(blobIndex)~=90 && Om(blobIndex)~=360)
%      True_defect_O = (nL2==blobIndex);
   T_O_loc(blobIndex)=1;
else
    T_O_loc(blobIndex)=0;
end


if T_O_loc(blobIndex)== T_A_loc(blobIndex)==1
    True_defect = (nL2==blobIndex);

   count=count+1; % increase counter when condition above satisfied
    DL3(count)=DL(blobIndex)*(150/926); %calibrate defect length
```

```matlab
    Om(count)=Om(blobIndex);
end

end%end fuction

%Measure geometrical features of detected blob
for i= 1:u_range
    Im = (nL2==i);
 Measurements(i) = regionprops(Im,
'Area','Perimeter','Orientation','MajorAxisLength','MinorAxisLength');
 Am(i) = getfield(Measurements(i), 'Area');
 Pm(i) = getfield(Measurements(i), 'Perimeter');
 Om(i) = getfield(Measurements(i), 'Orientation');
 MajAxm(i) = getfield(Measurements(i), 'MajorAxisLength');
 MinrAxm(i) = getfield(Measurements(i), 'MinorAxisLength');

end

 % calibration of detected geometrical data
 DL=DL.*(150/926);
 Am=Am.*((150/926)^2);
 Pm=Pm.*(150/926);
 MajAxm=MajAxm.*(150/926);
 MinrAxm=MinrAxm.*(150/926);

% percentile representation of detected data
pp=1;
 for p=10:10:90;
 DL_S(pp) = prctile(DL,p);
 Am_S(pp) = prctile(Am,p);
 Pm_S(pp) = prctile(Pm,p);
 Om_S(pp) = prctile(abs(Om),p);
 MajAxm_S(pp) = prctile(MajAxm,p);
 MinrAxm_S(pp)=prctile(MinrAxm,p);
 pp=pp+1;
 end
 PL=[DL_S' Am_S' Om_S'];
```

Appendix C6: Simulated results for data set 3

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
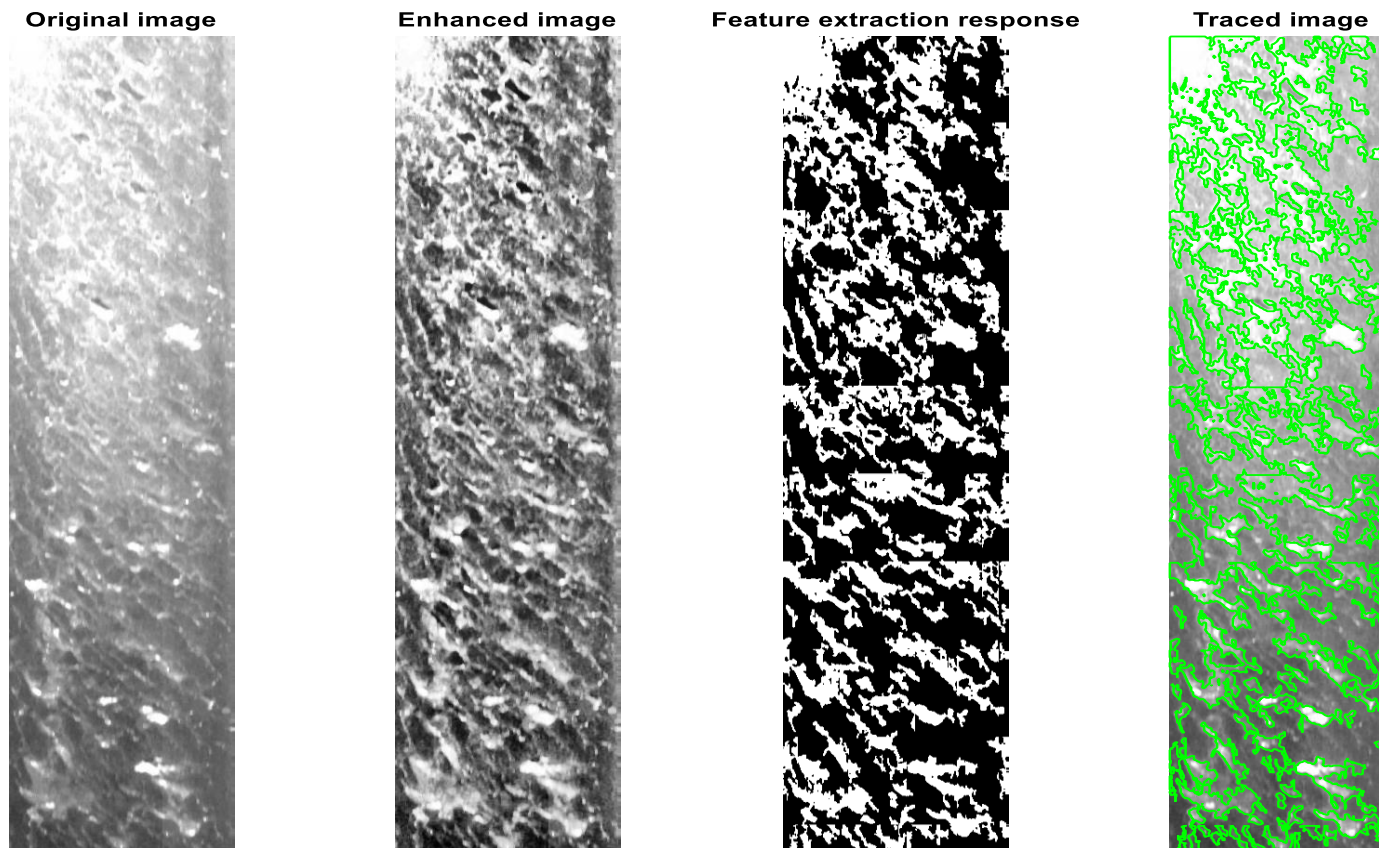
Figure A.C6. 1: Shows original, enhanced, extracted feature, and traced images for FA-H1.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
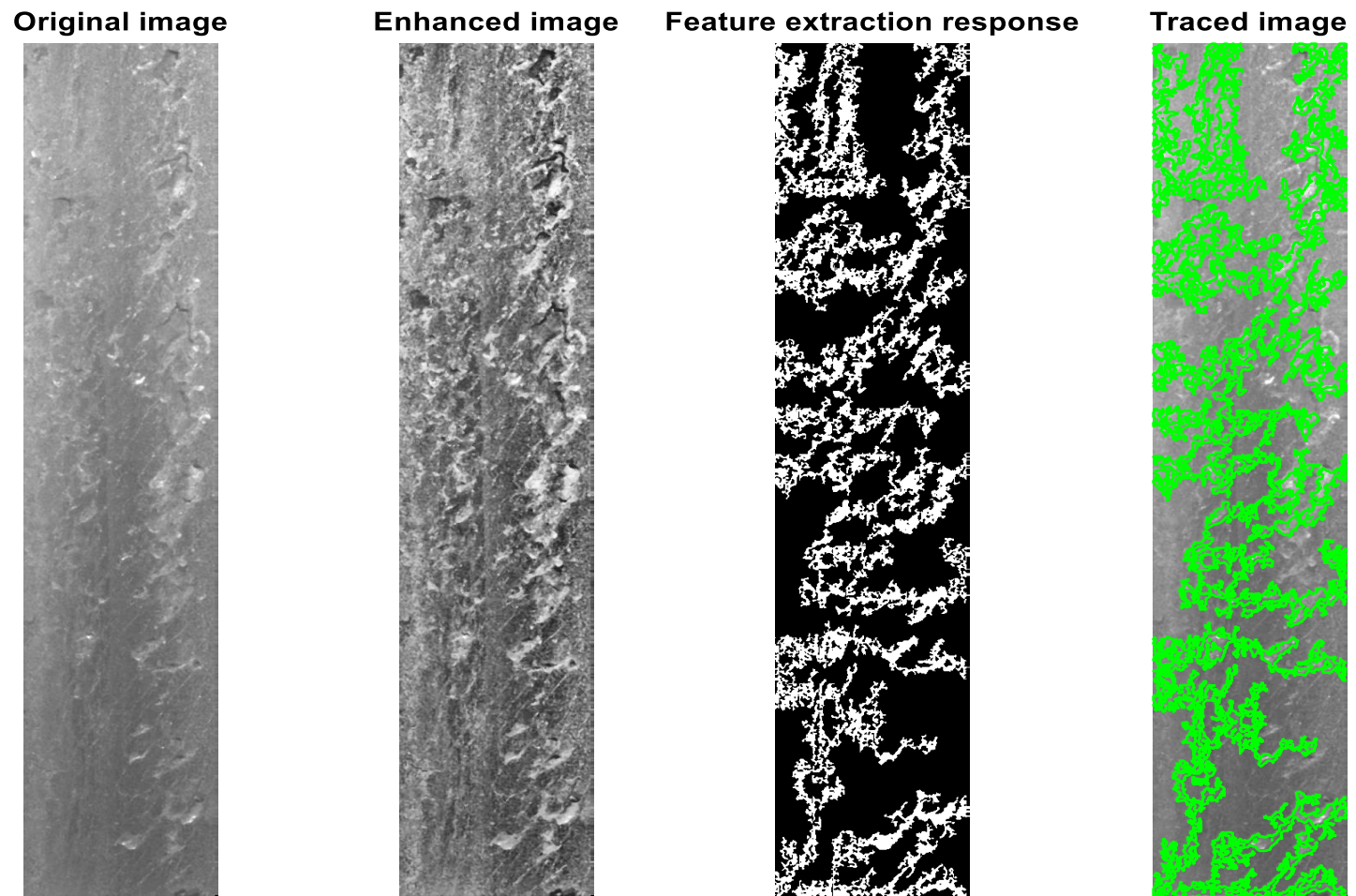
Figure A.C6. 2: Shows original, enhanced, extracted feature, and traced images for FA-H2.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
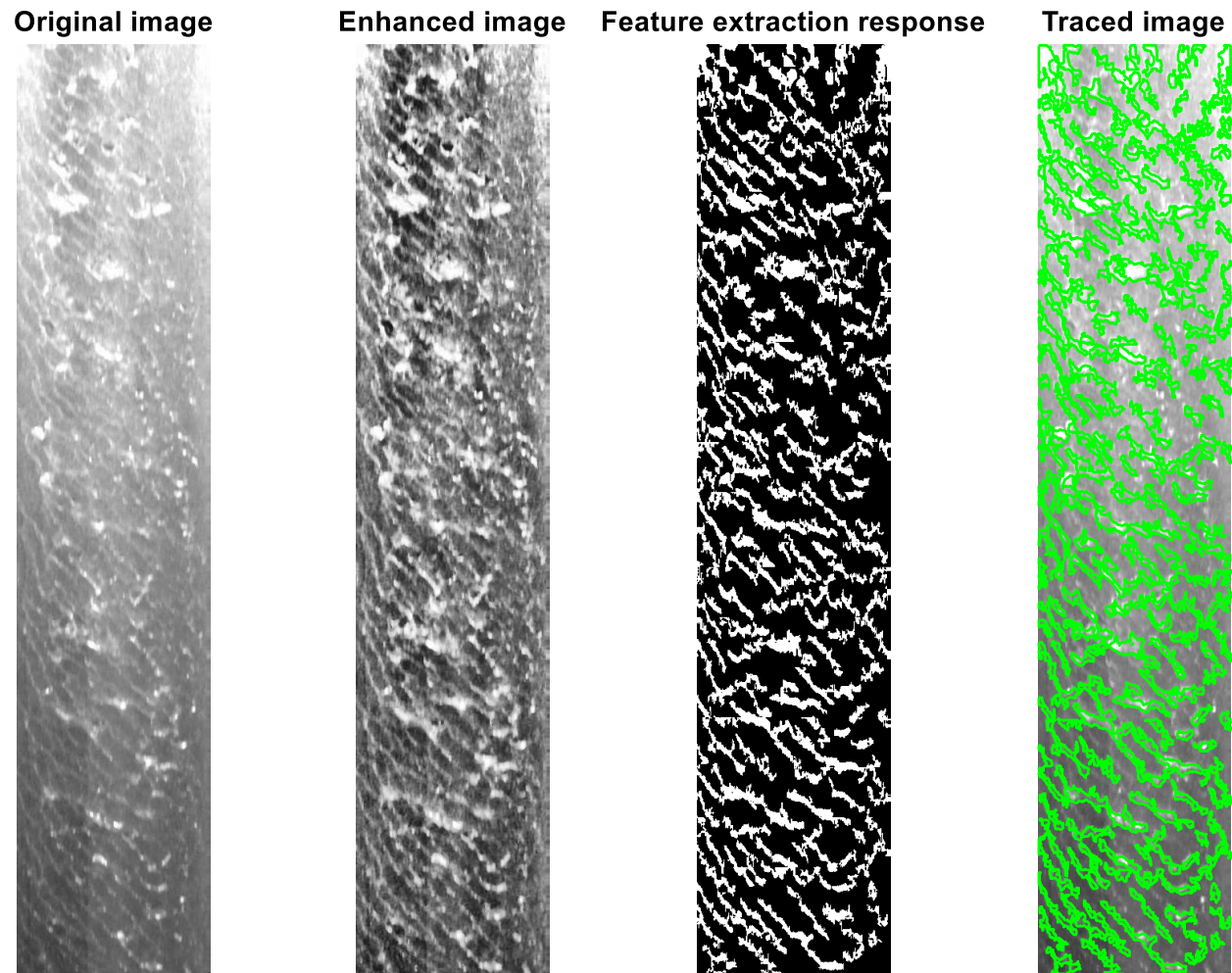
Figure A.C6. 3: Shows original, enhanced, extracted feature, and traced images for FA-H3.

Original image　　　**Enhanced image**　**Feature extraction response**　**Traced image**
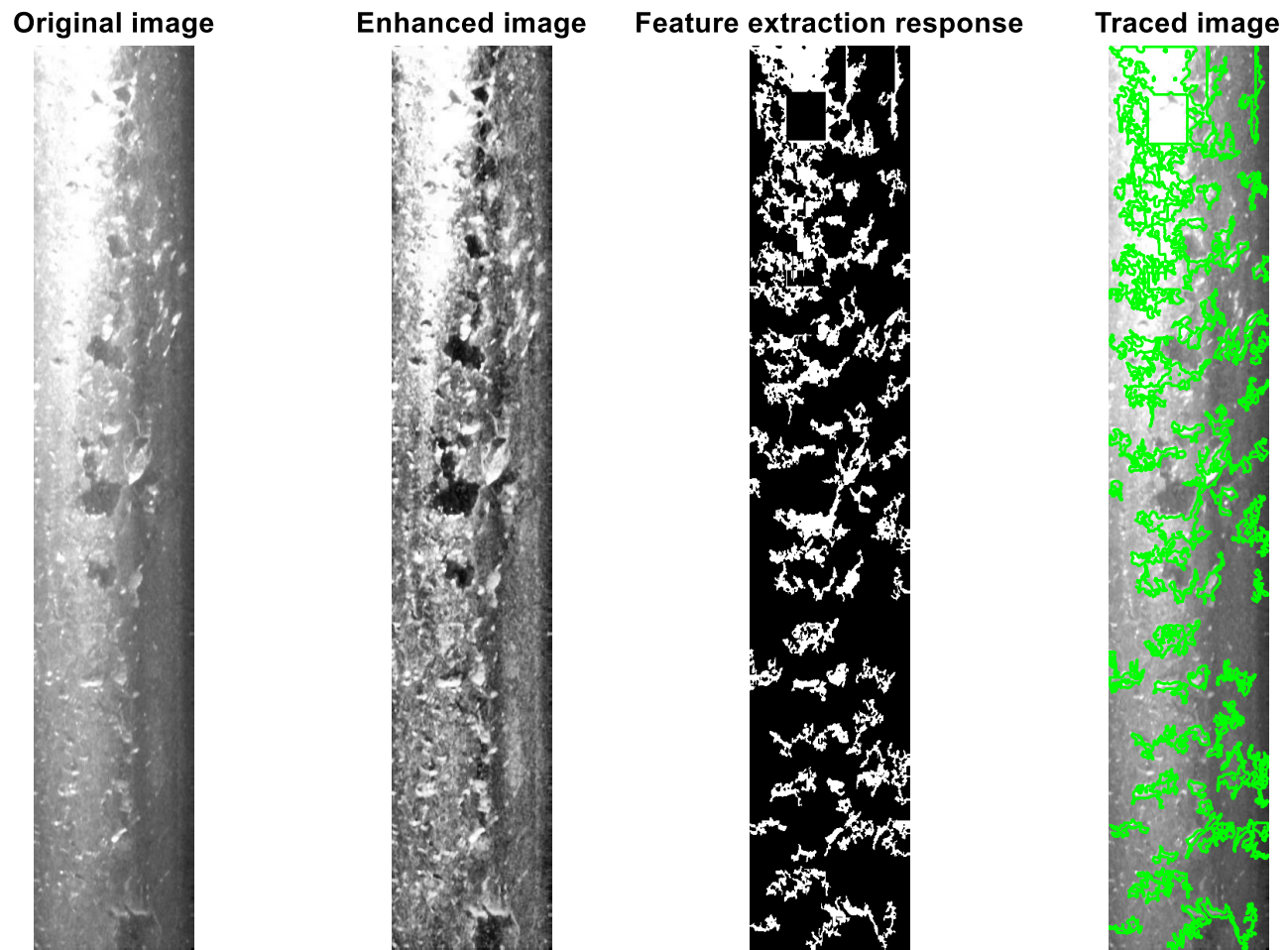
Figure A.C6. 4: Shows original, enhanced, extracted feature, and traced images for FA-H4.

Original image   **Enhanced image** **Feature extraction response**   **Traced image**
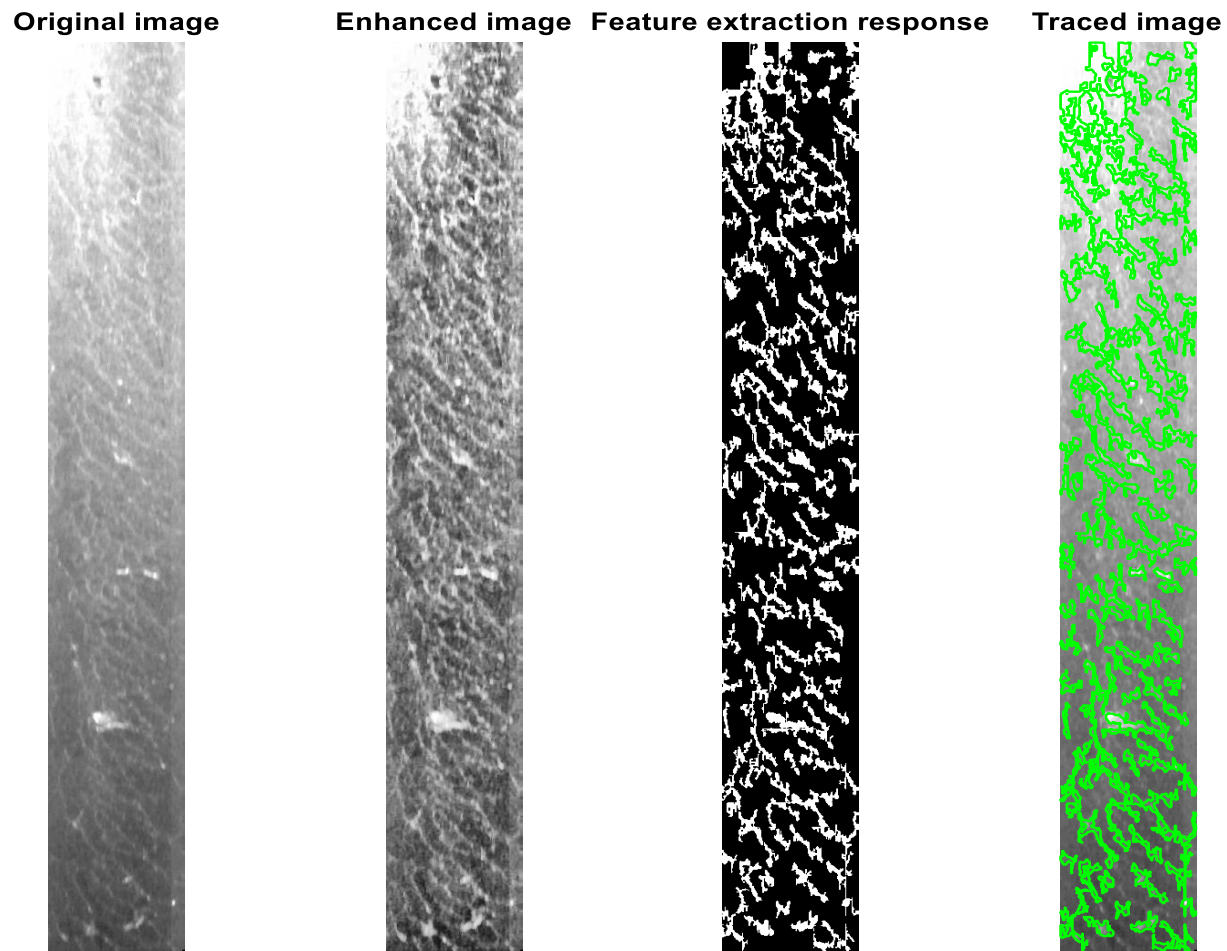
Figure A.C6. 5: Shows original, enhanced, extracted feature, and traced images for FA-H5.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
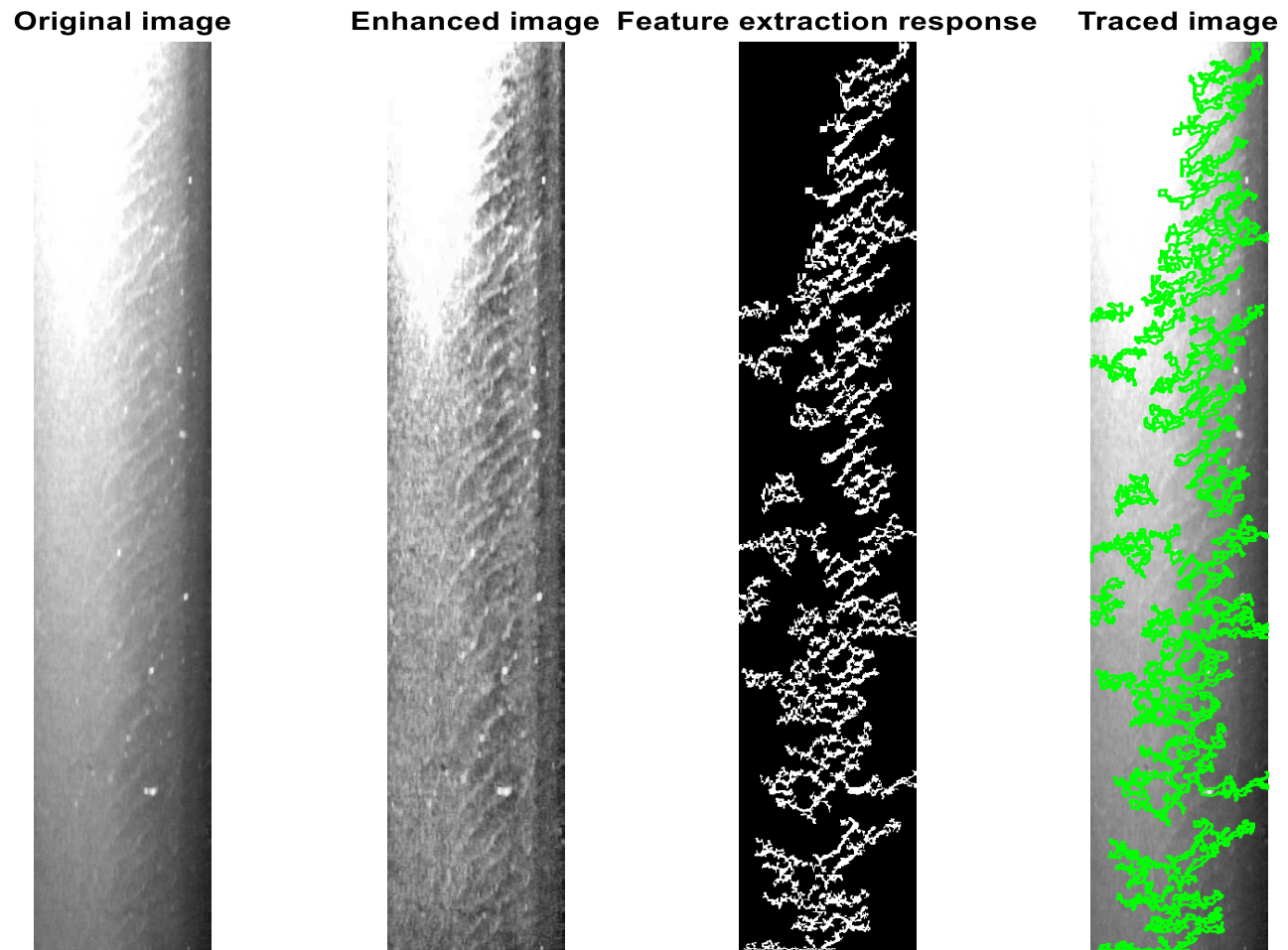
Figure A.C6. 6: Shows original, enhanced, extracted feature, and traced images for FA-M1.

Original image **Enhanced image** **Feature extraction response** **Traced image**
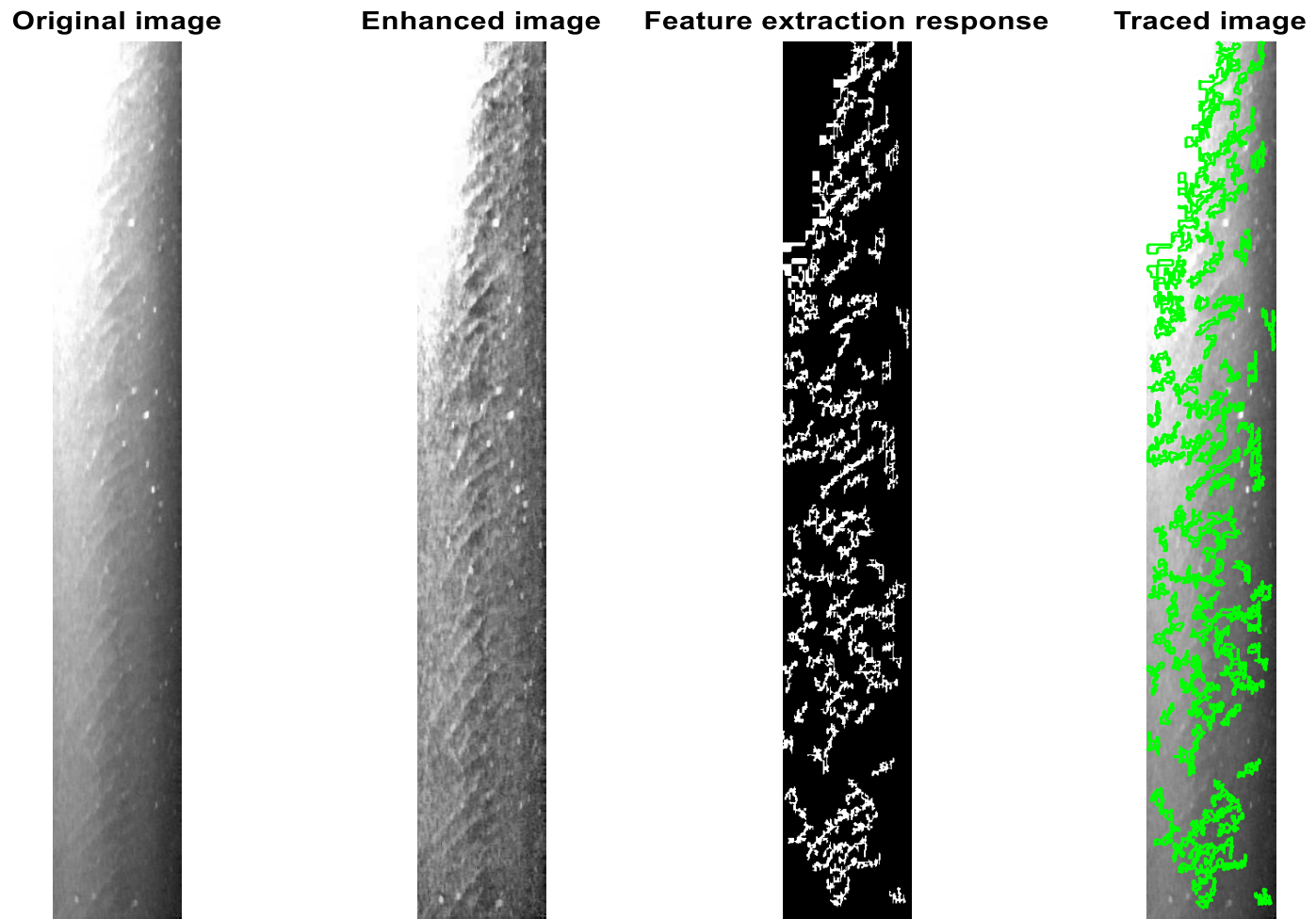
Figure A.C6. 7: Shows original, enhanced, extracted feature, and traced images for FA-M2.

Original image **Enhanced image** **Feature extraction response** **Traced image**
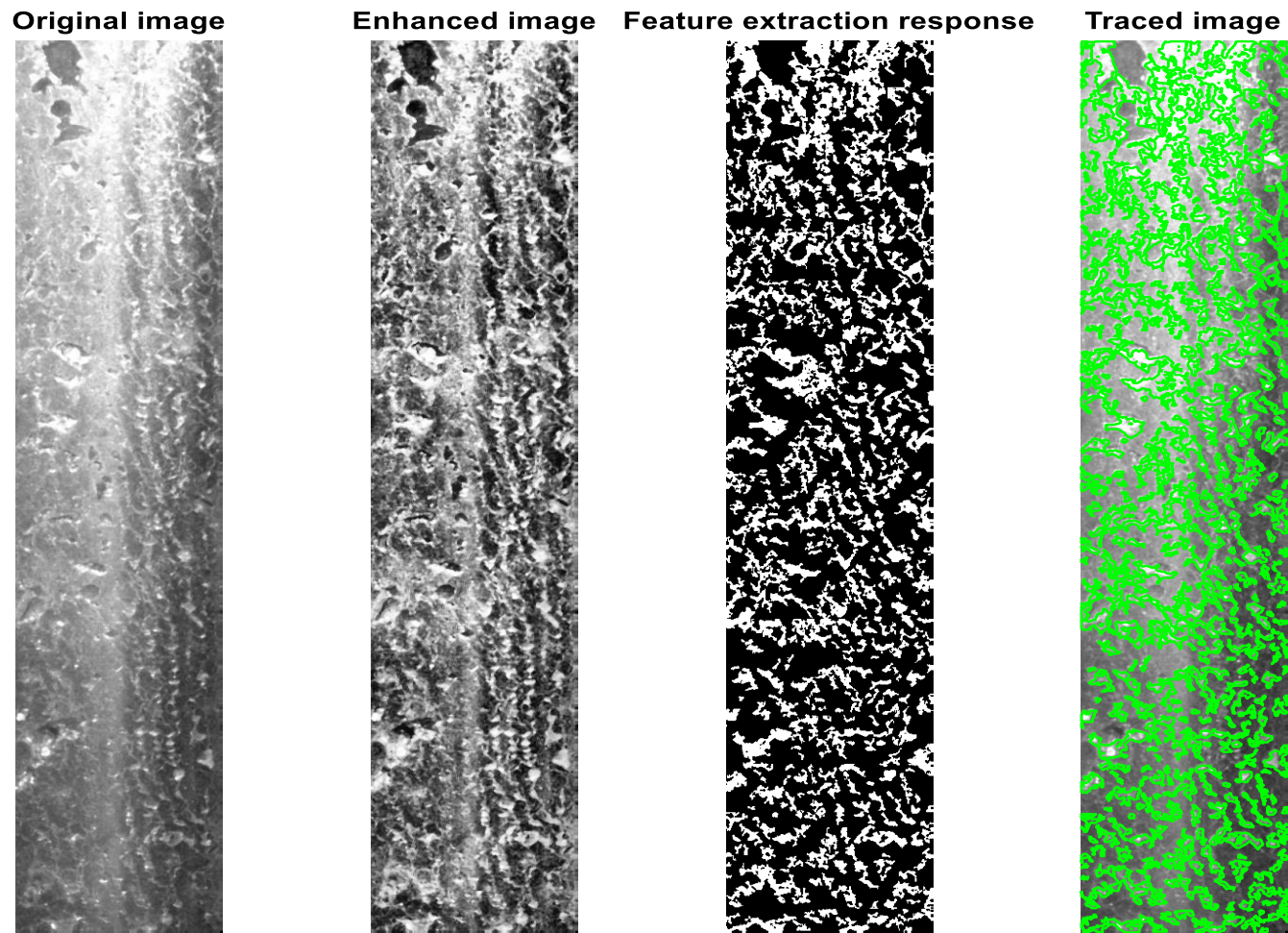
Figure A.C6. 8: Shows original, enhanced, extracted feature, and traced images for FA-M3.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
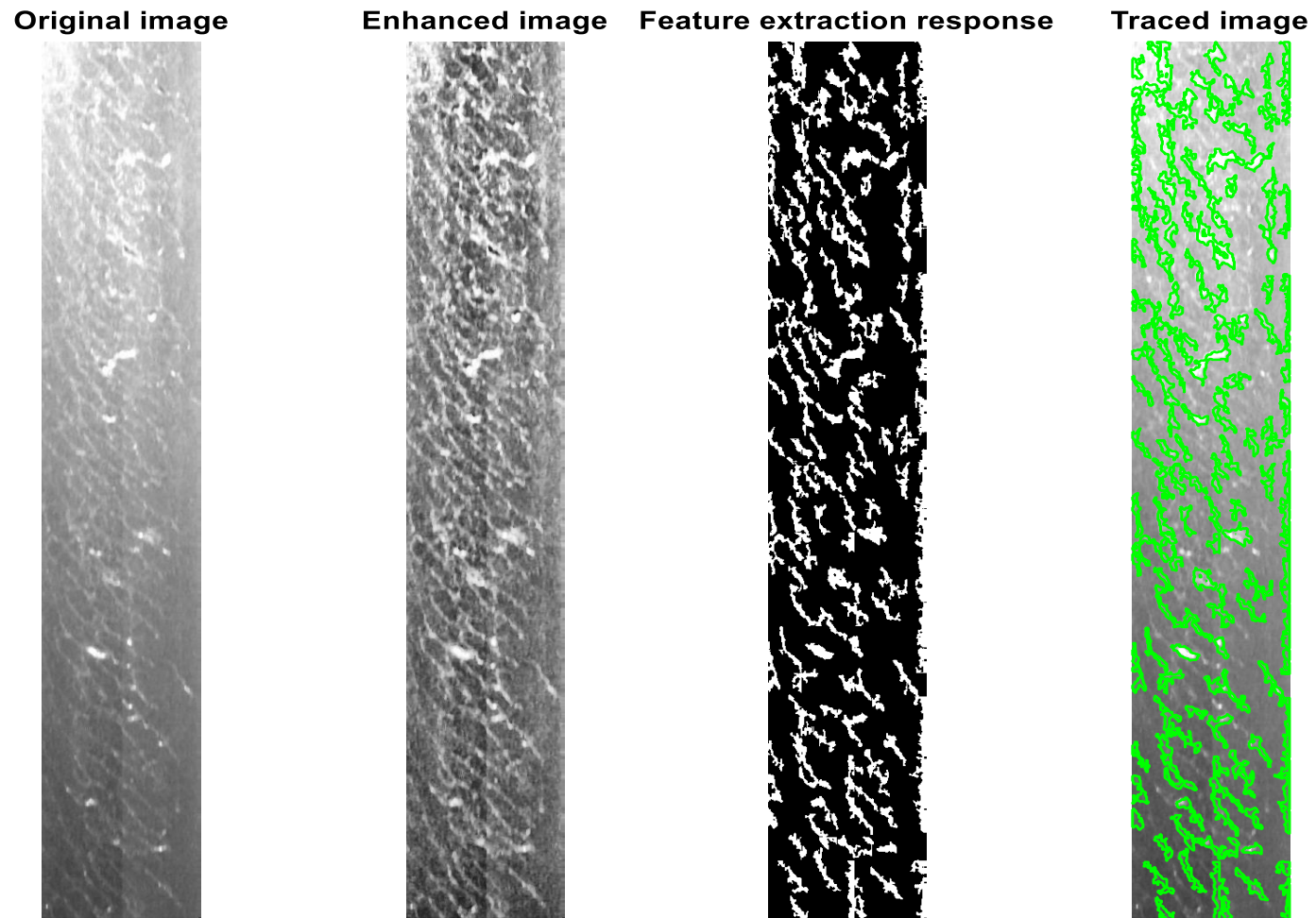
Figure A.C6. 9: Shows original, enhanced, extracted feature, and traced images for FA-M4.

Original image **Enhanced image** **Feature extraction response** **Traced image**
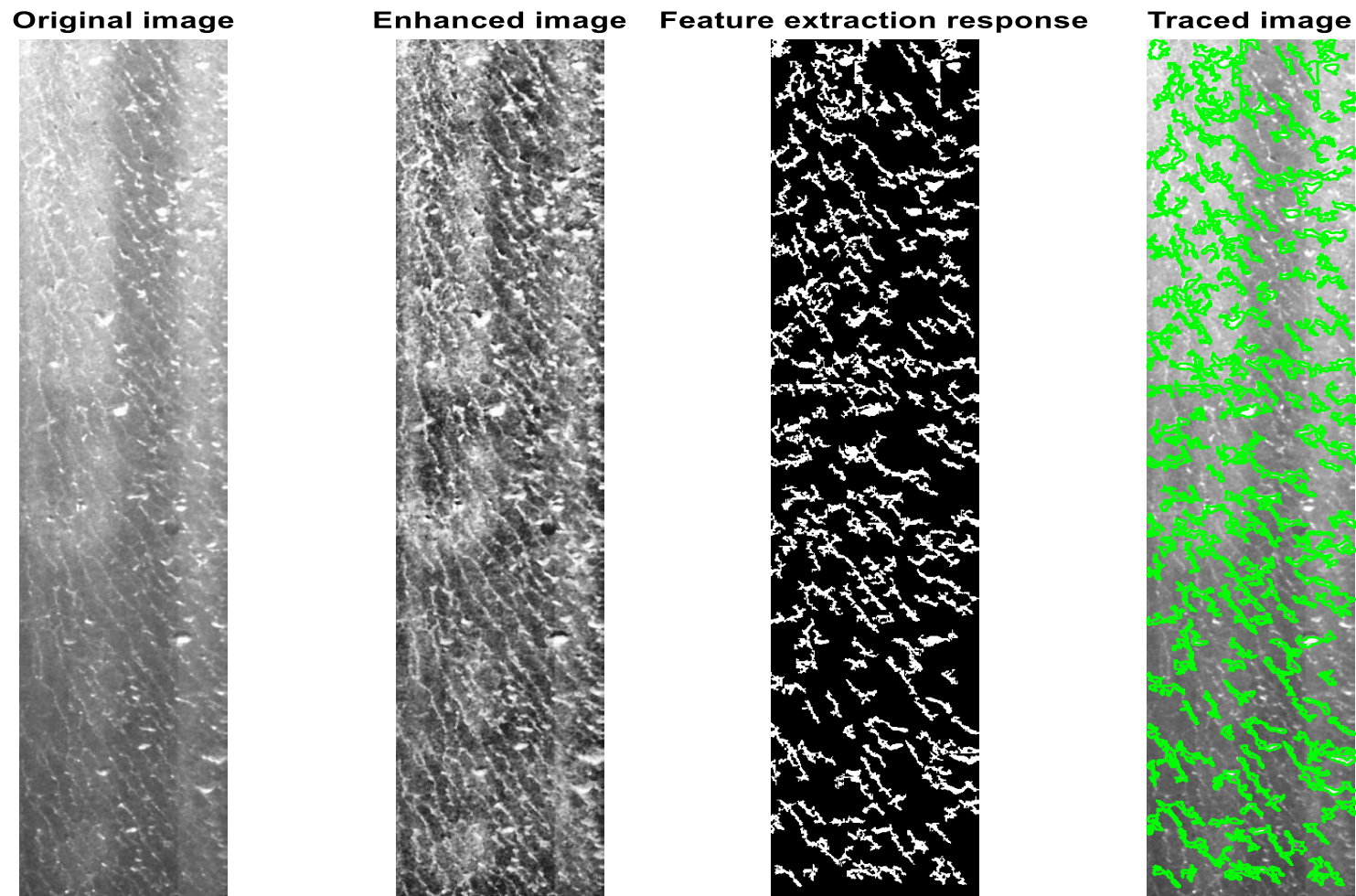
Figure A.C6. 10: Shows original, enhanced, extracted feature, and traced images for FA-M5.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
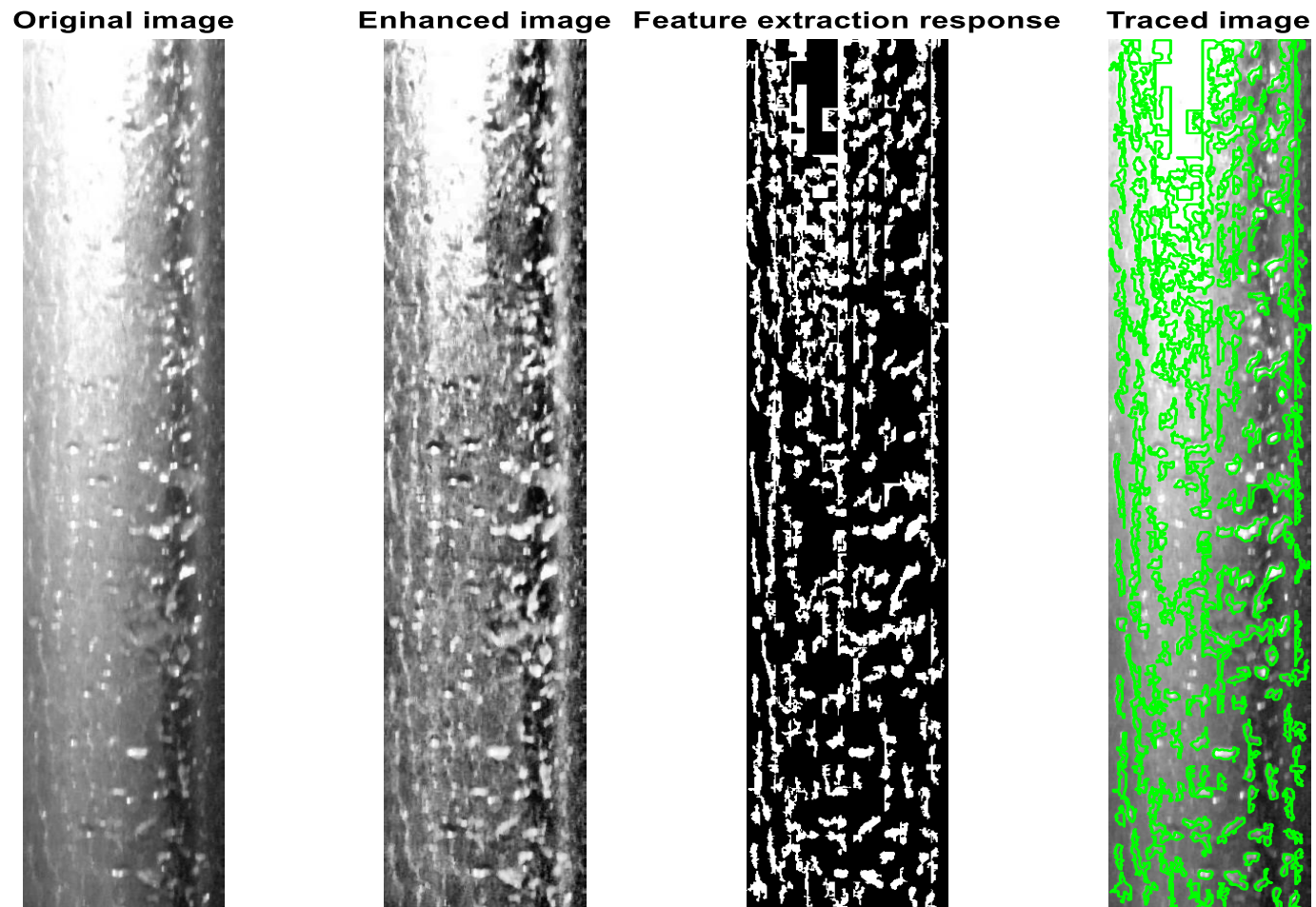
Figure A.C6. 11: Shows original, enhanced, extracted feature, and traced images for FA-L1.
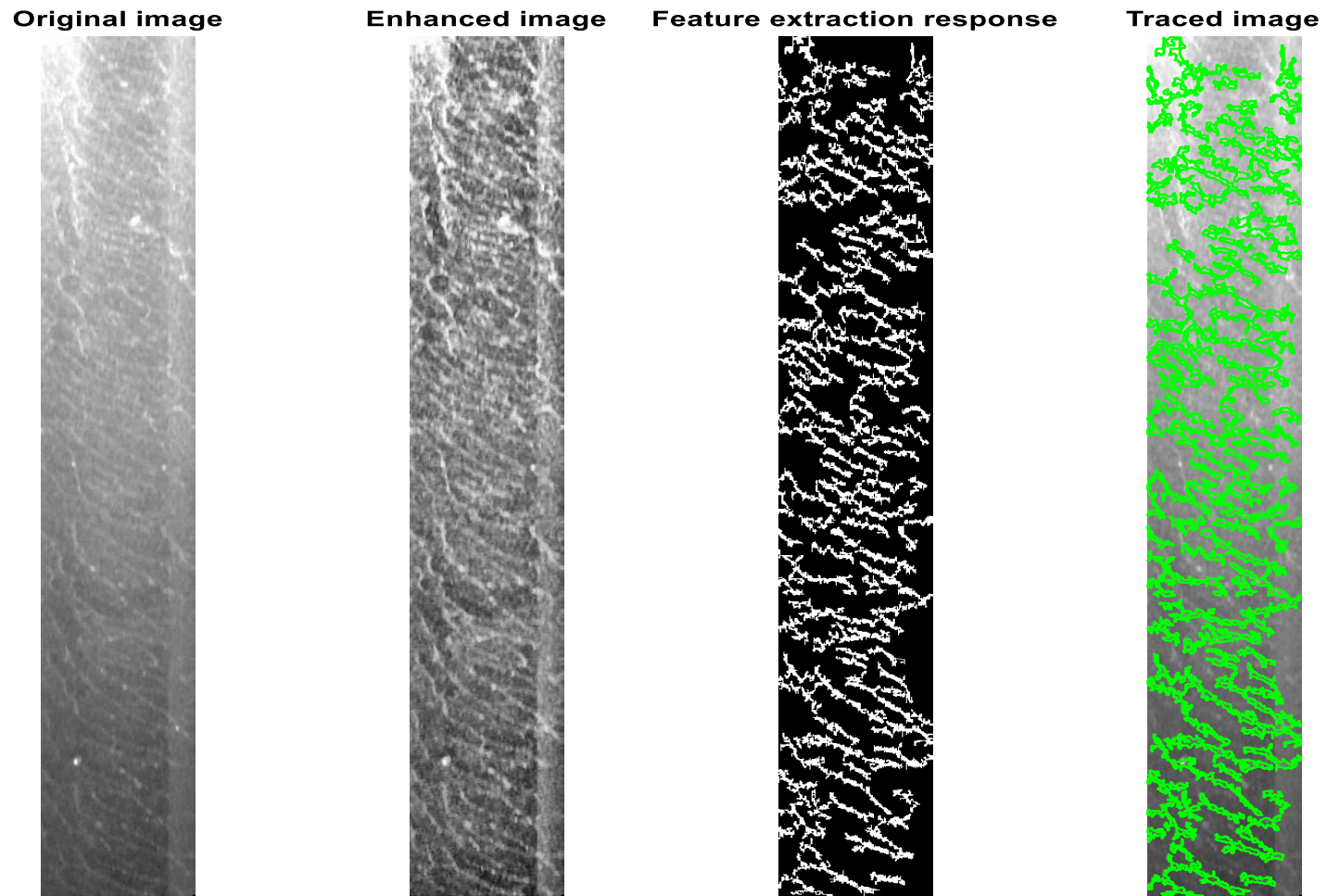
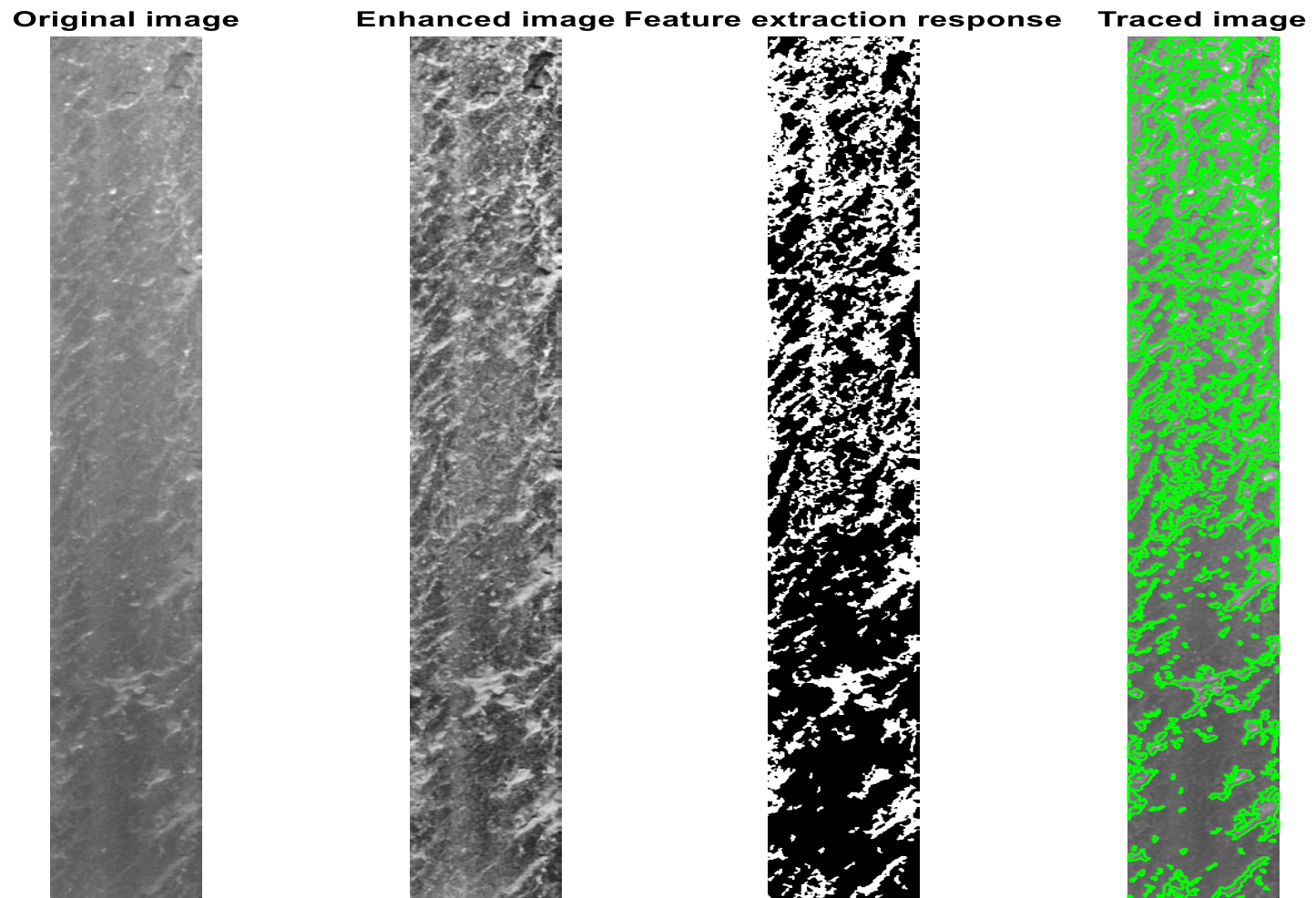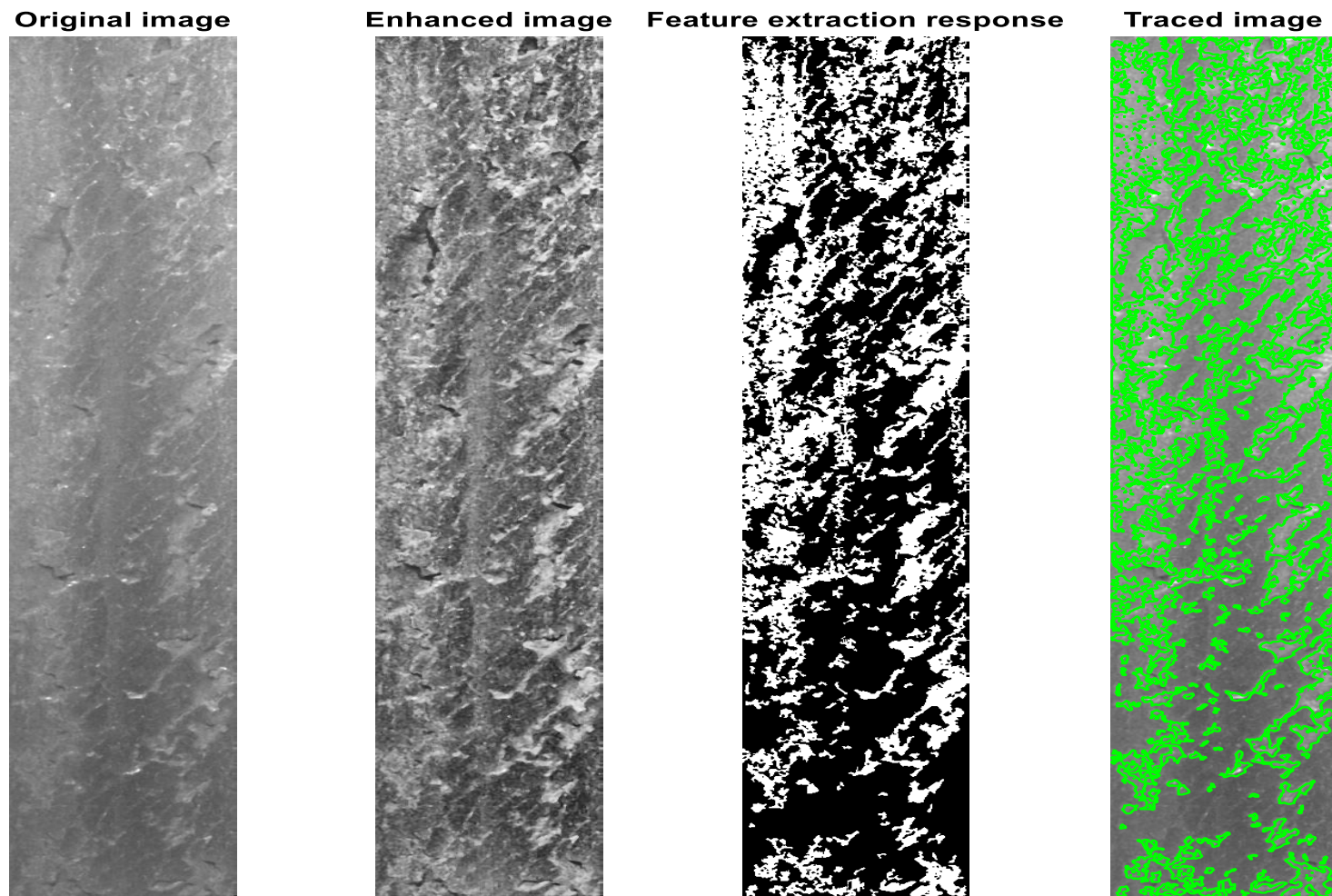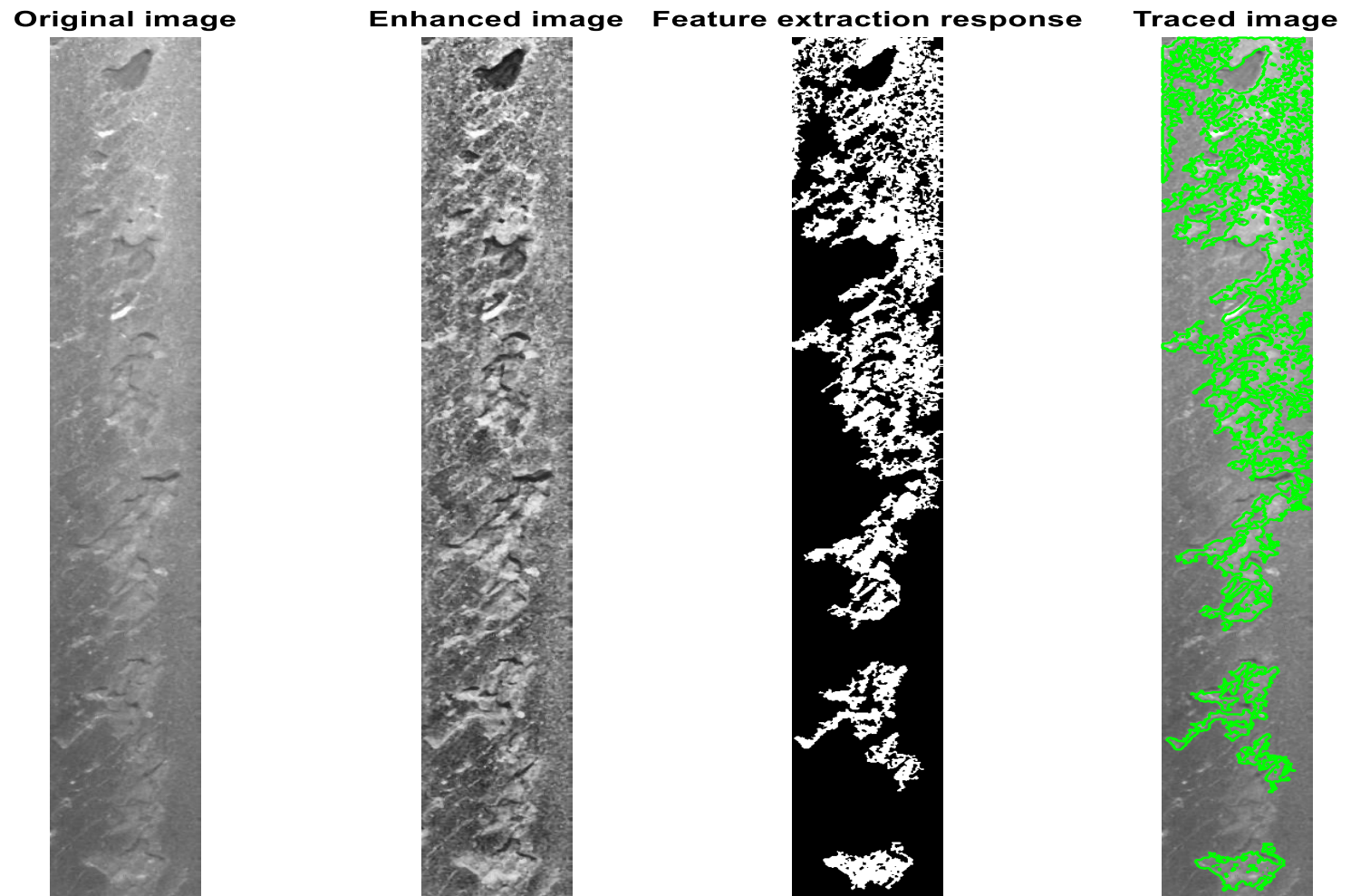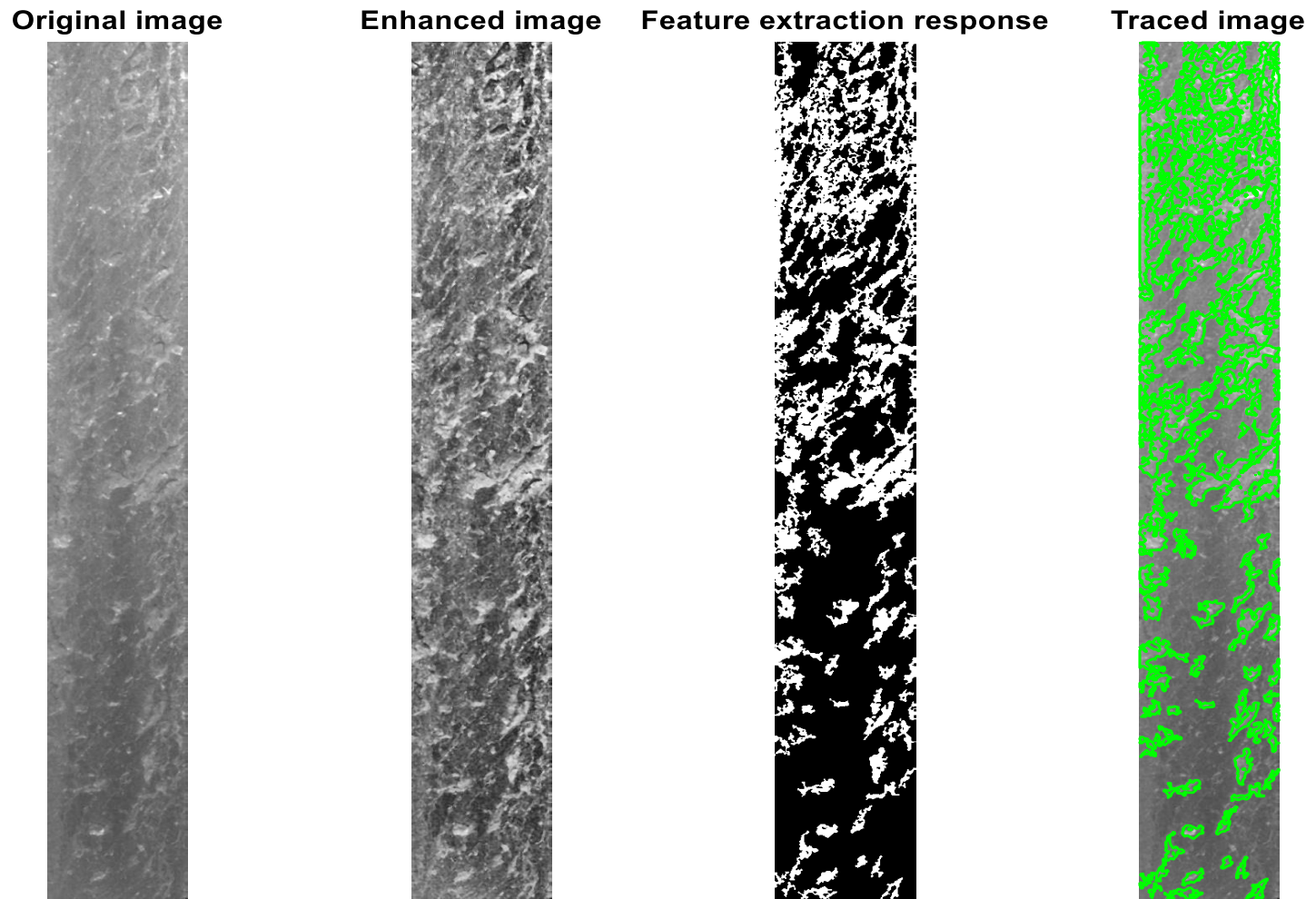**Figure A.C6. 12: Shows original, enhanced, extracted feature, and traced images for FA-L2.**

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
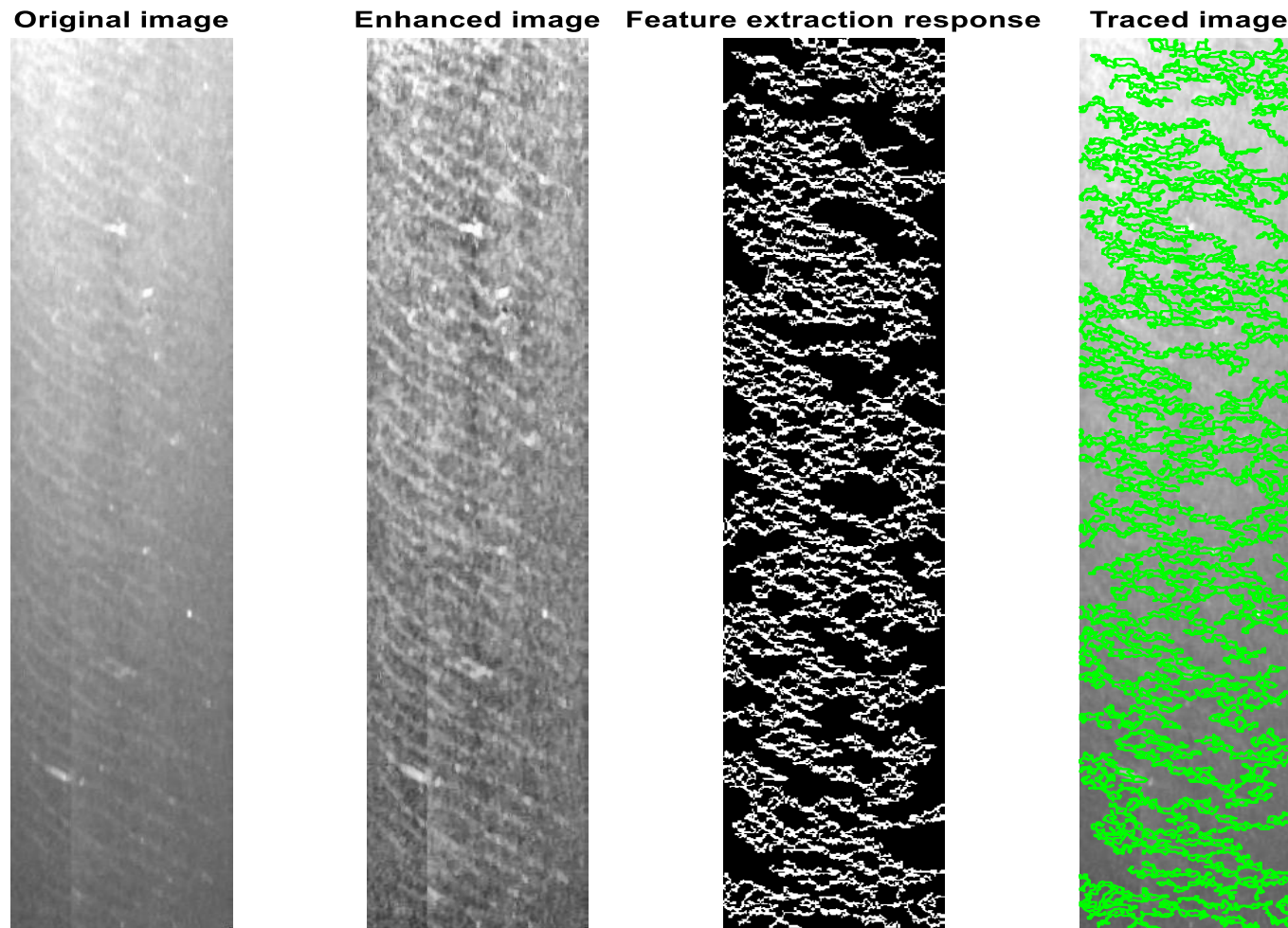
Figure A.C6. 13: Shows original, enhanced, extracted feature, and traced images for FA-L3.

Original image　　**Enhanced image**　　**Feature extraction response**　　**Traced image**
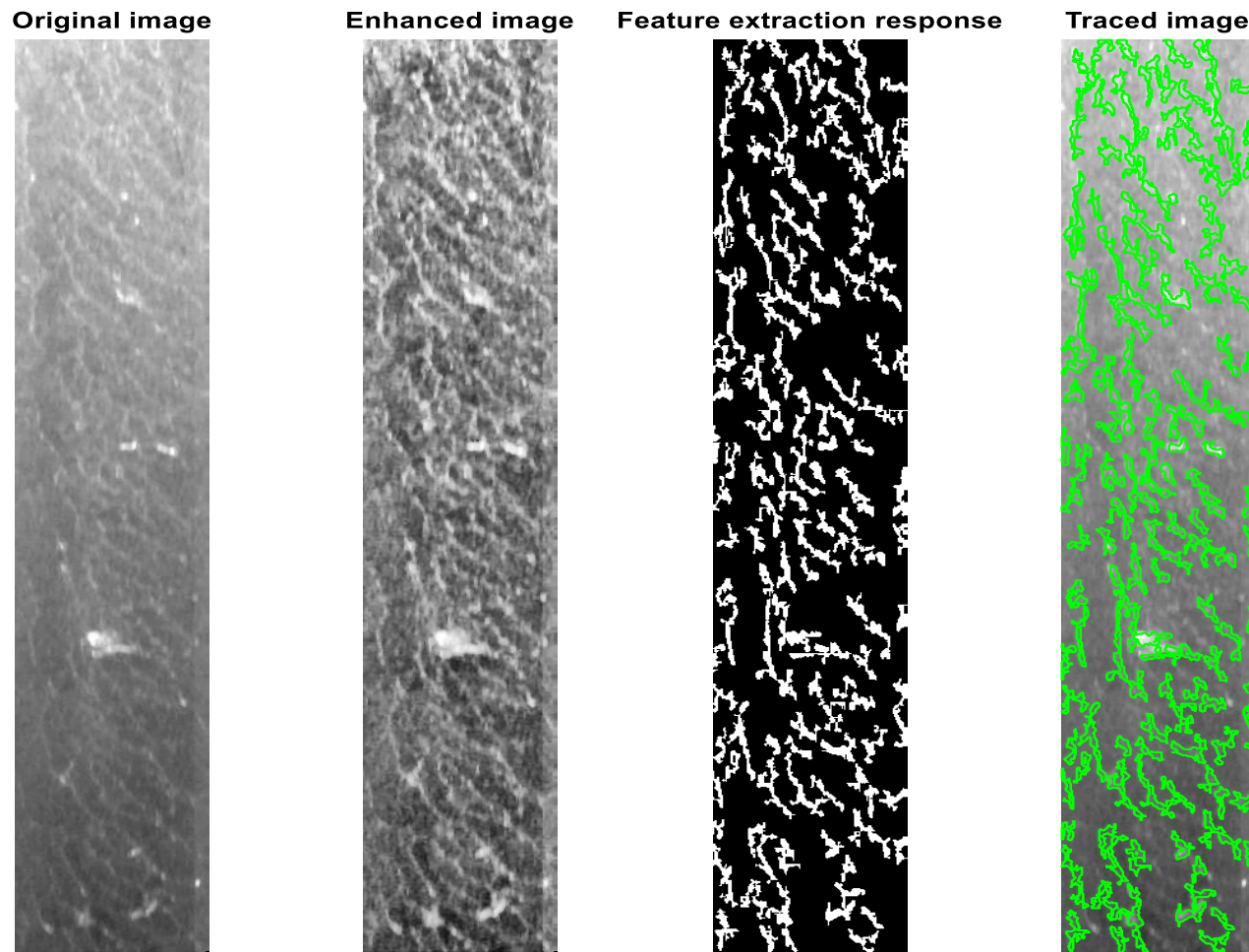
Figure A.C6. 14: Shows original, enhanced, extracted feature, and traced images for FA-L4.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
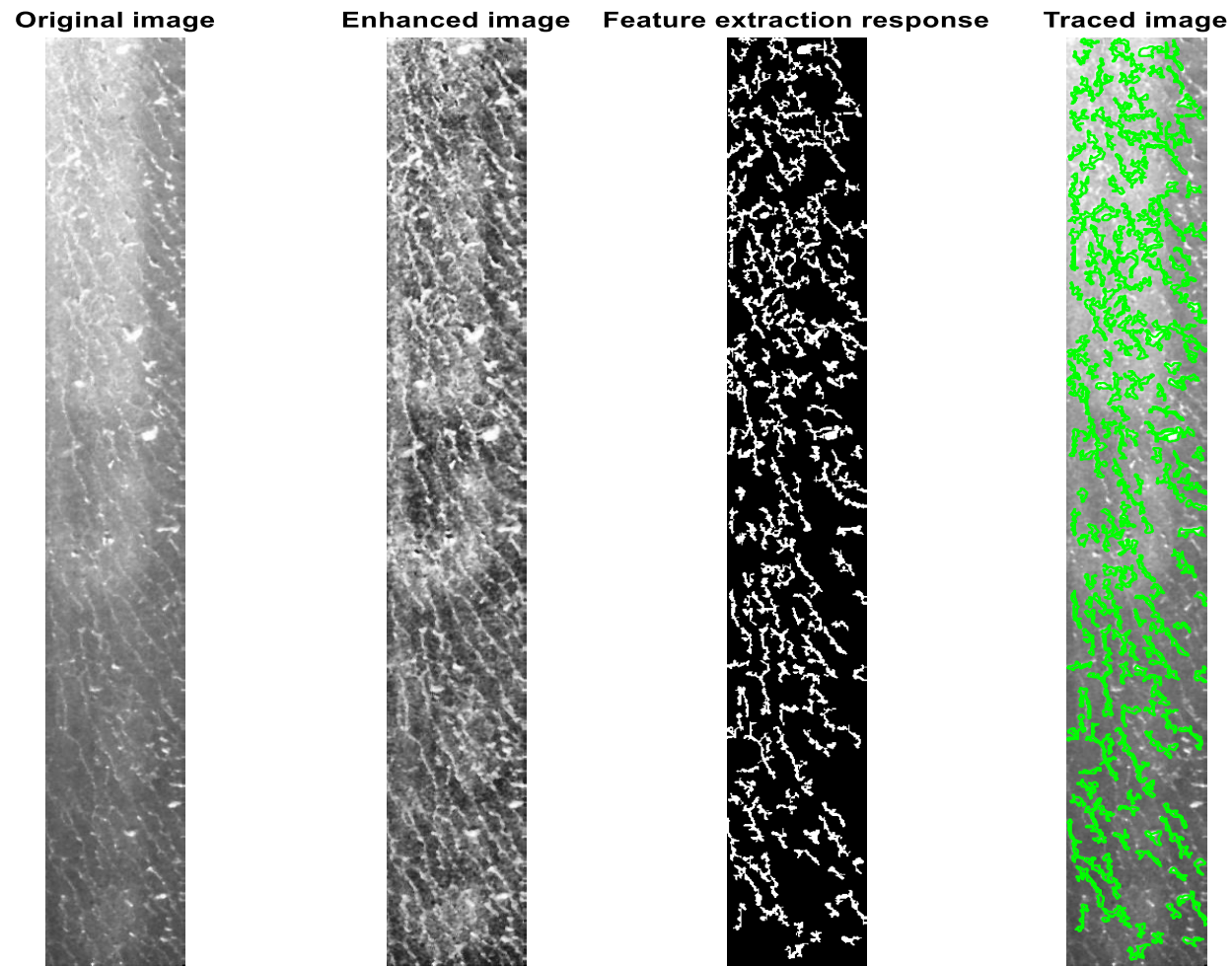
Figure A.C6. 15: Shows original, enhanced, extracted feature, and traced images for FA-L5.

Appendix C7: Simulated results for data set 3

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
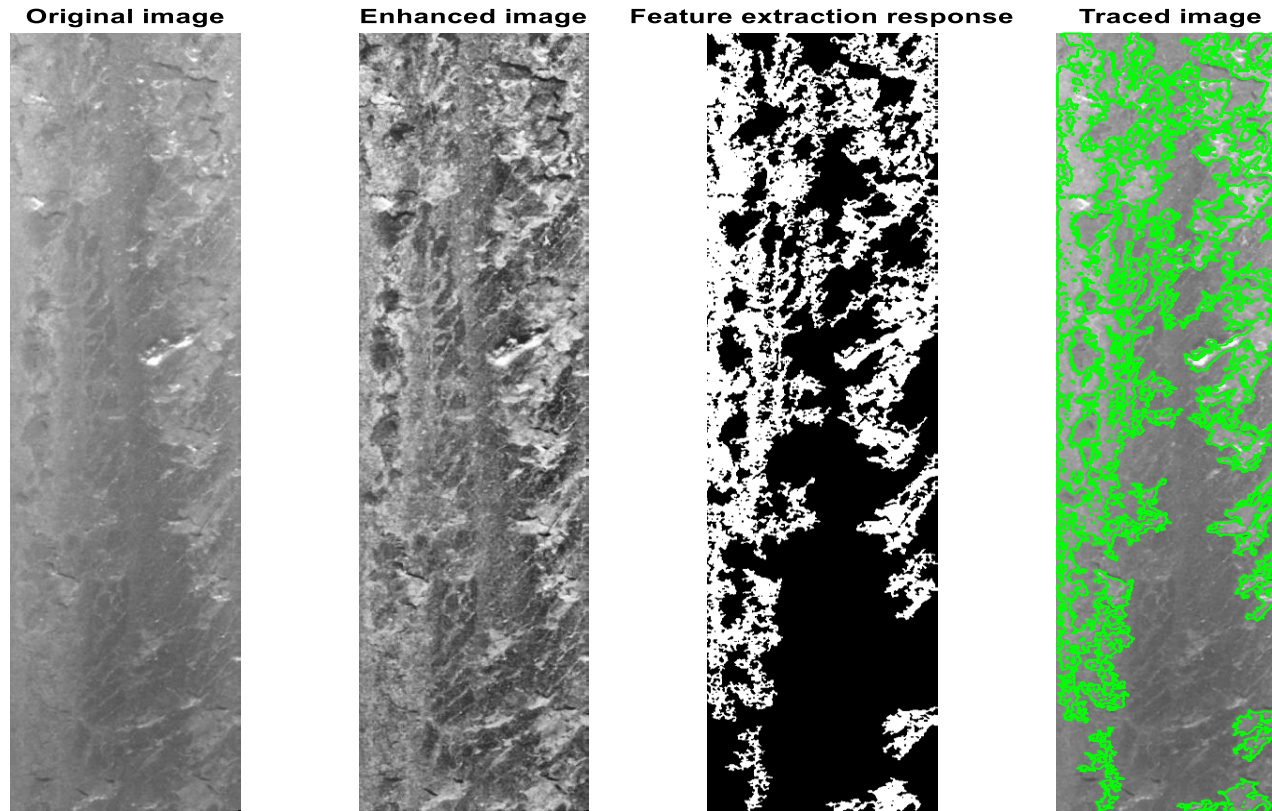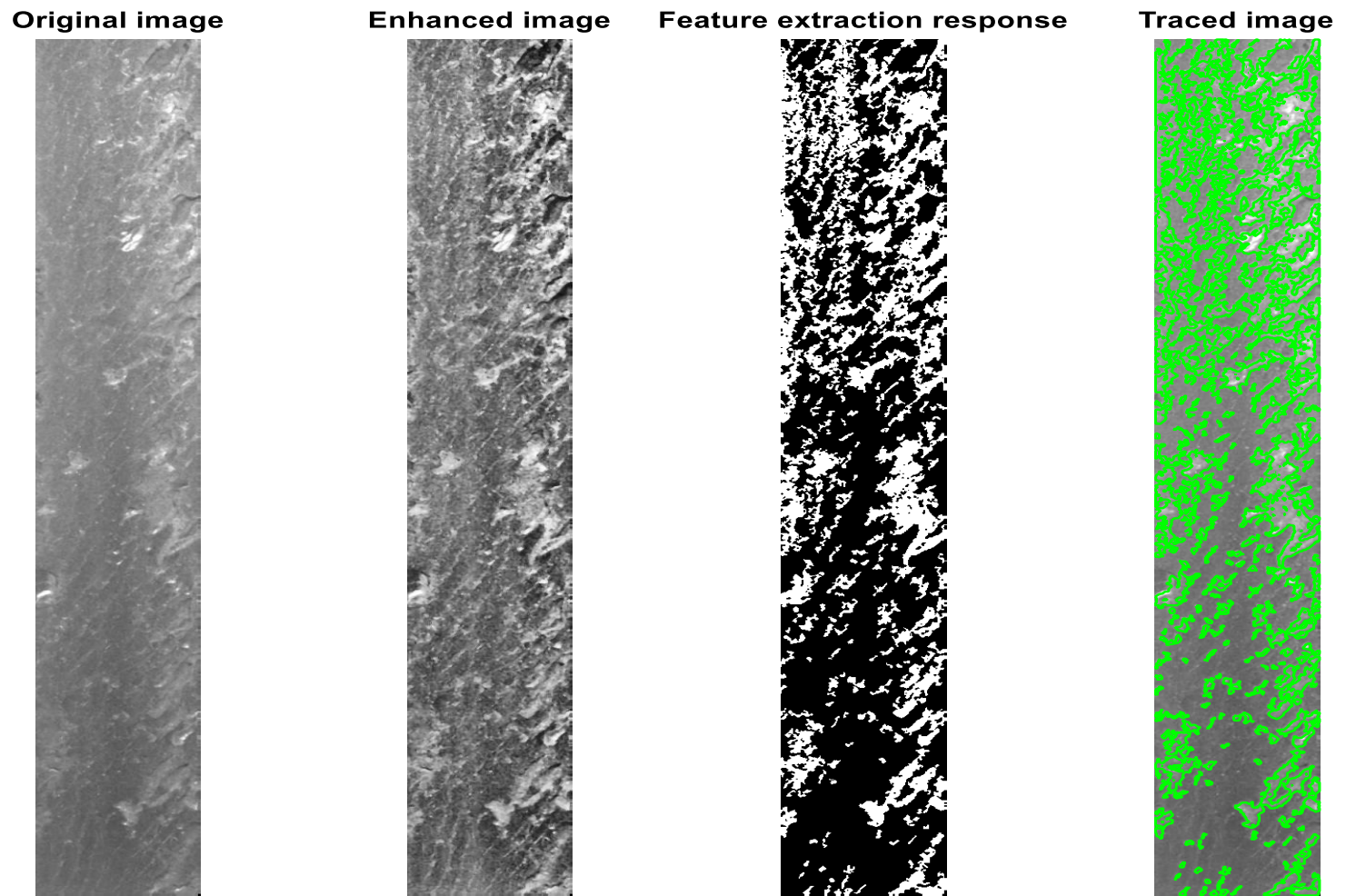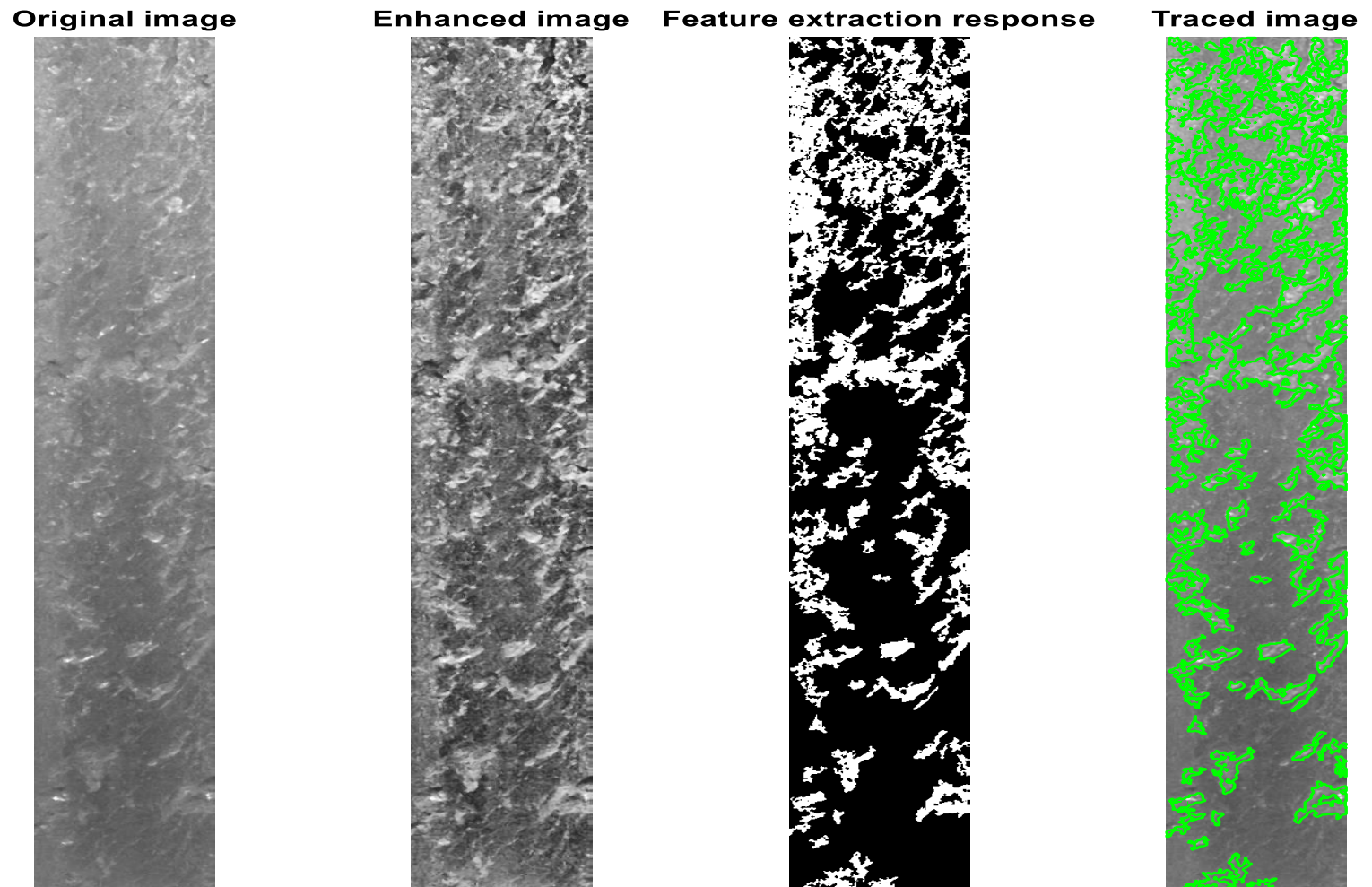
Figure A.C7. 1: Shows original, enhanced, extracted feature, and traced images for L-4060ft.
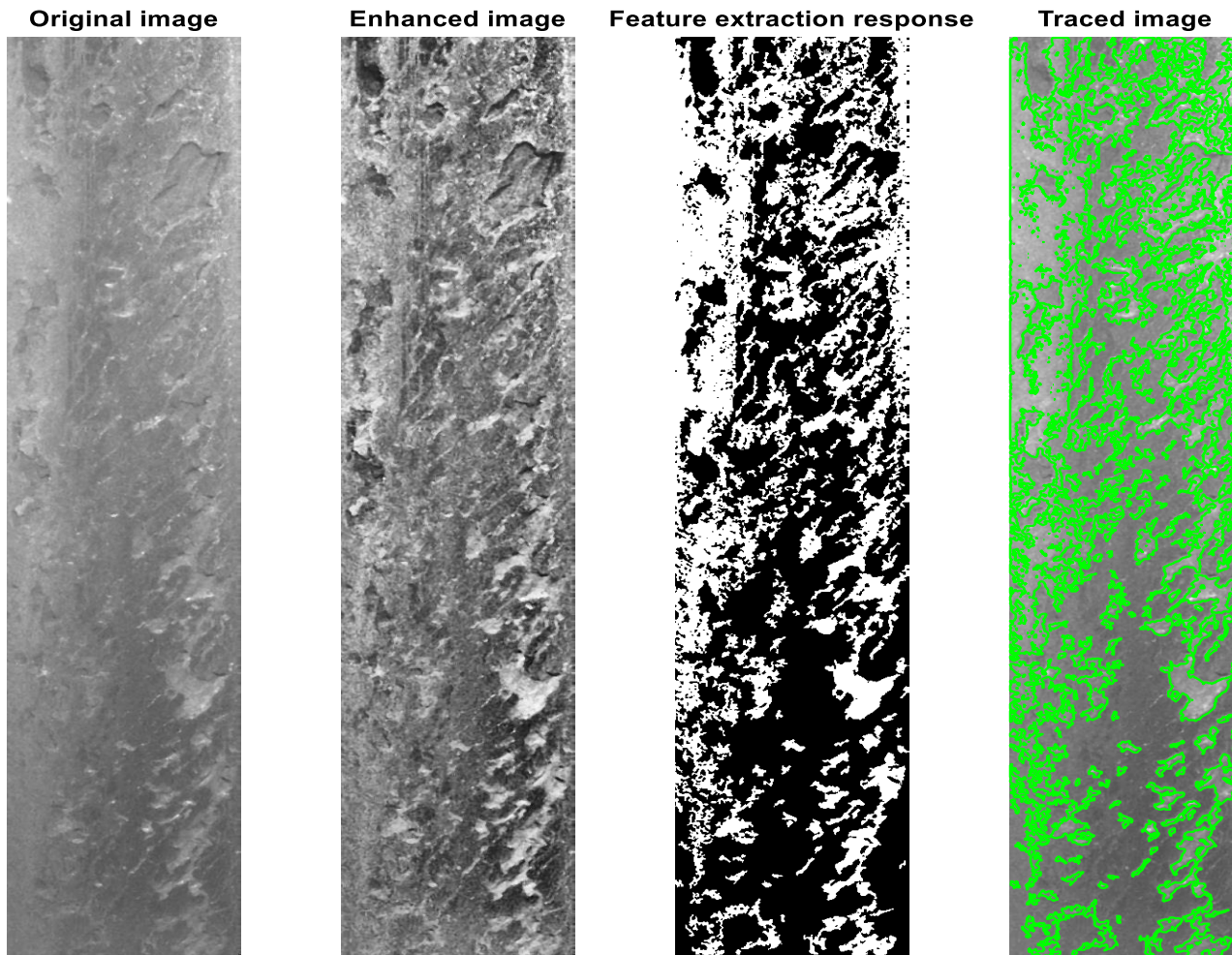
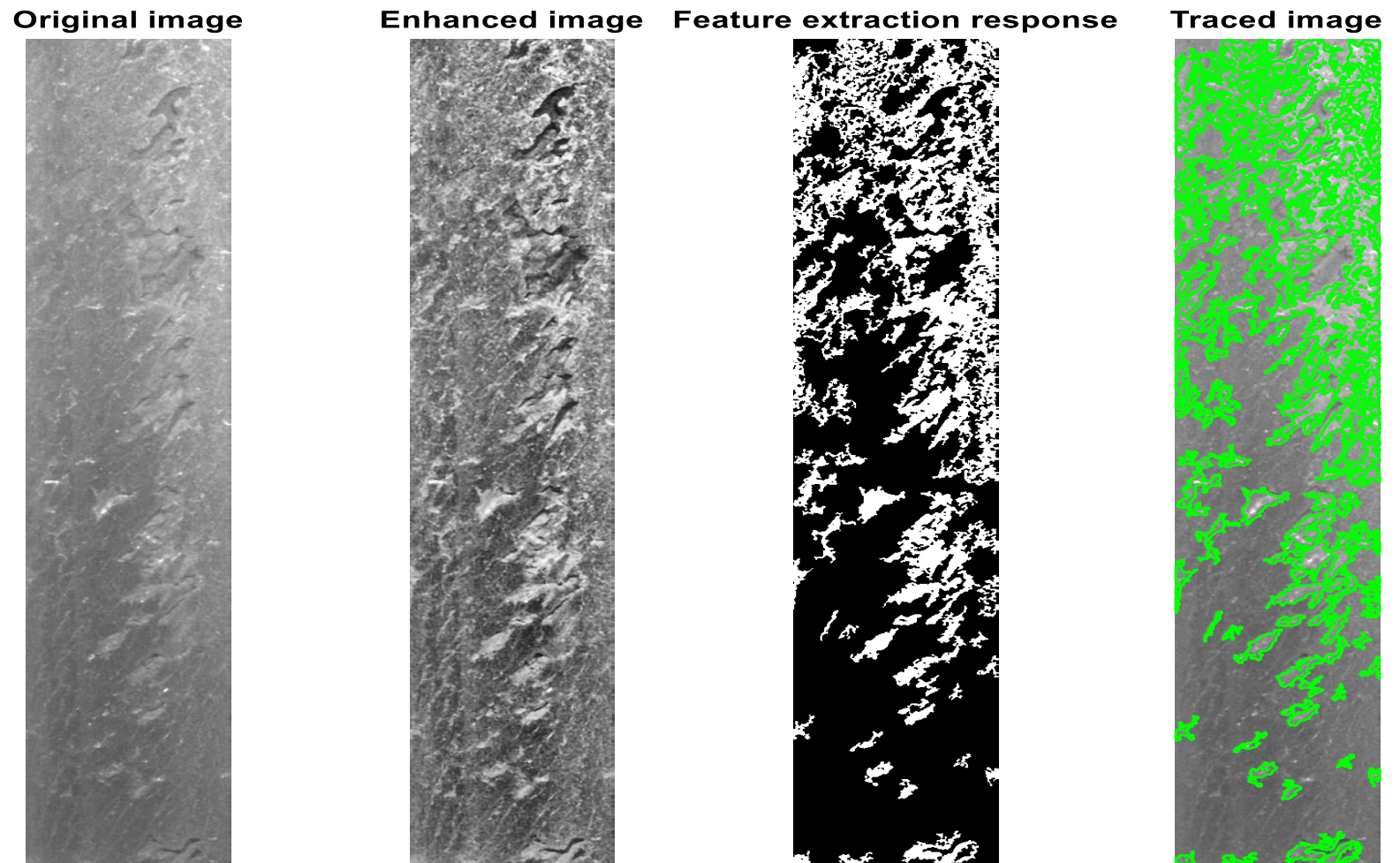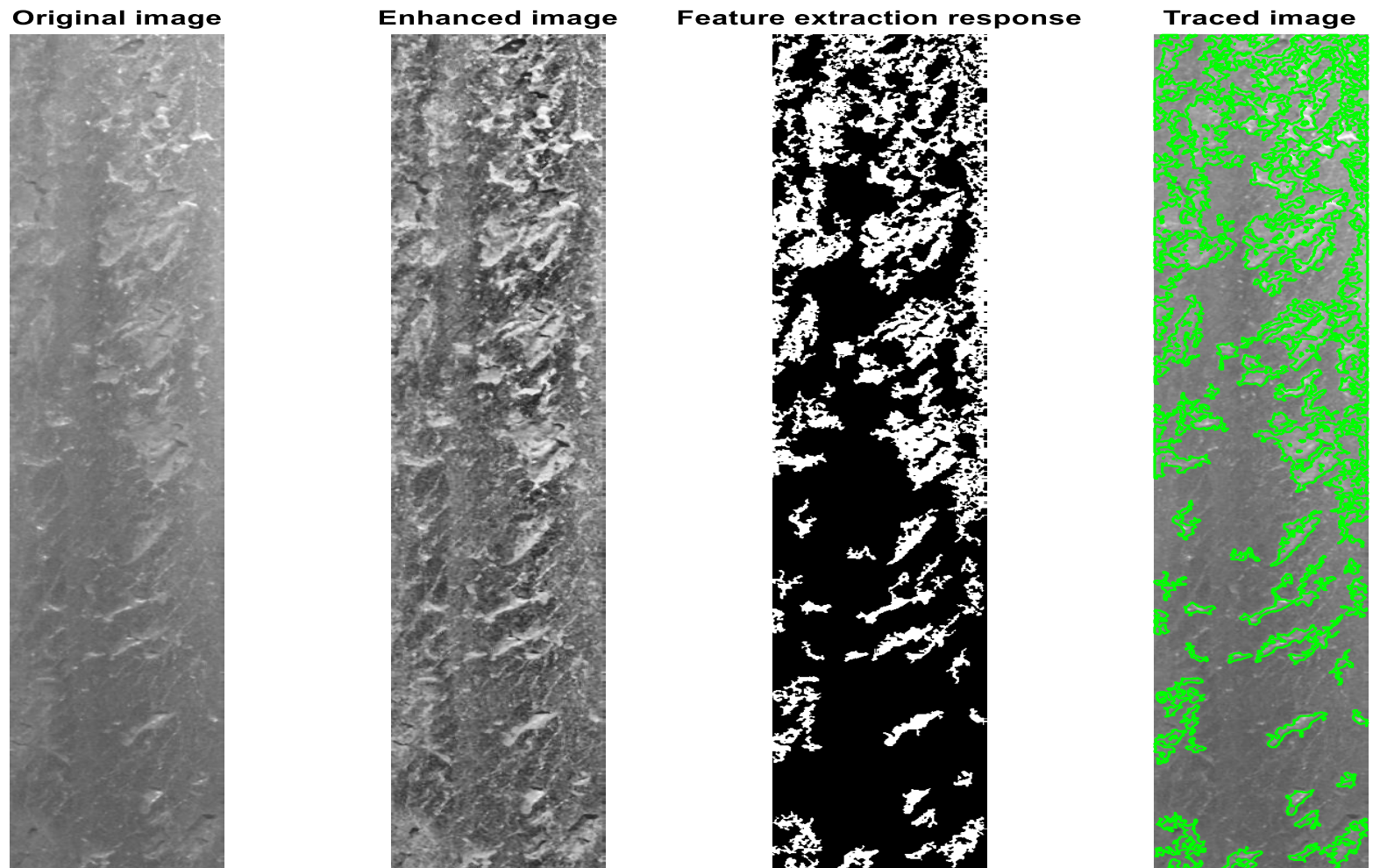**Figure A.C7. 2: Shows original, enhanced, extracted feature, and traced images for L-4063ft.**

Original image **Enhanced image** **Feature extraction response** **Traced image**
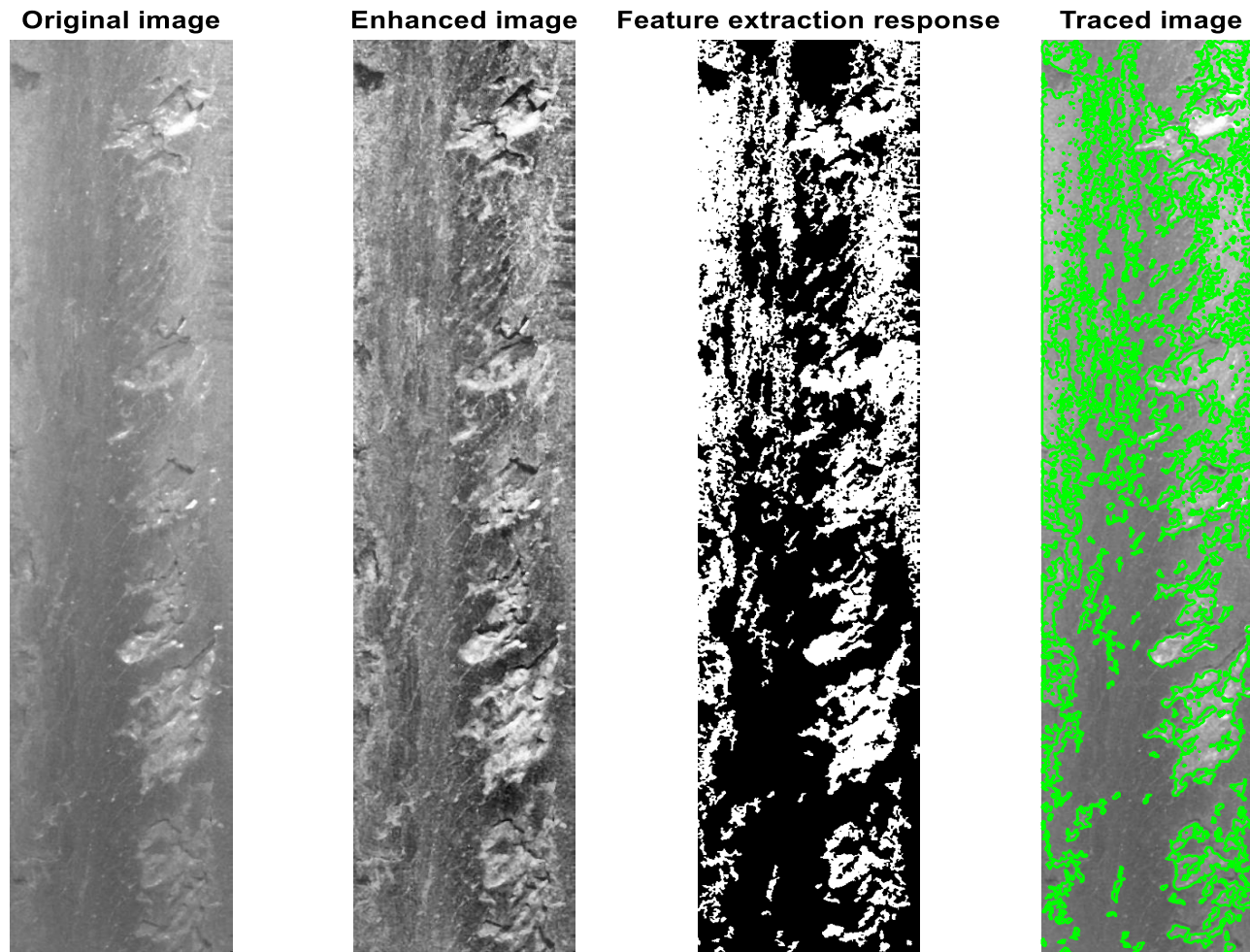
Figure A.C7. 3: Shows original, enhanced, extracted feature, and traced images for L-4110ft.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
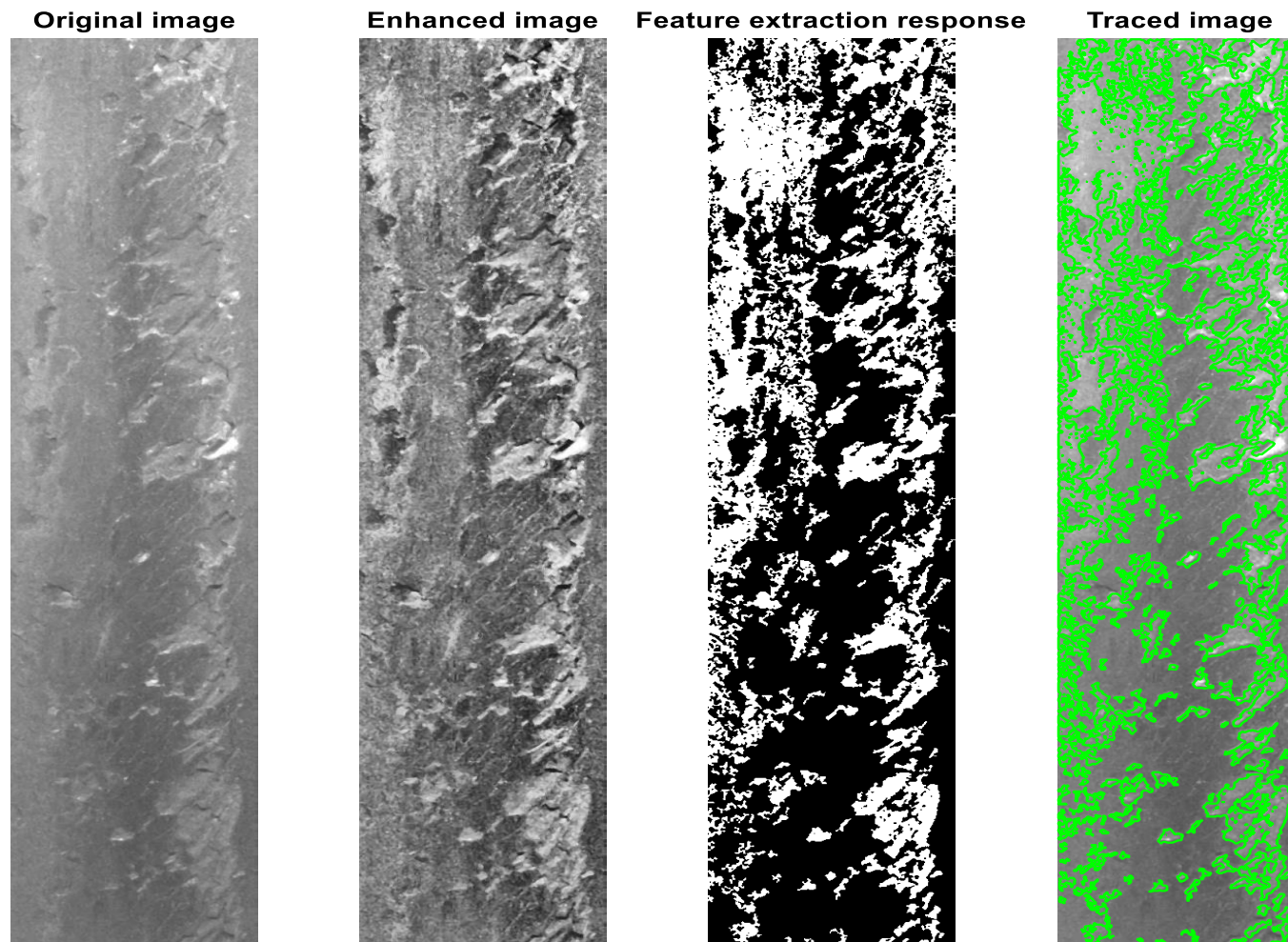
Figure A.C7. 4: Shows original, enhanced, extracted feature, and traced images for L-4195ft.

Original image　　**Enhanced image**　**Feature extraction response**　**Traced image**
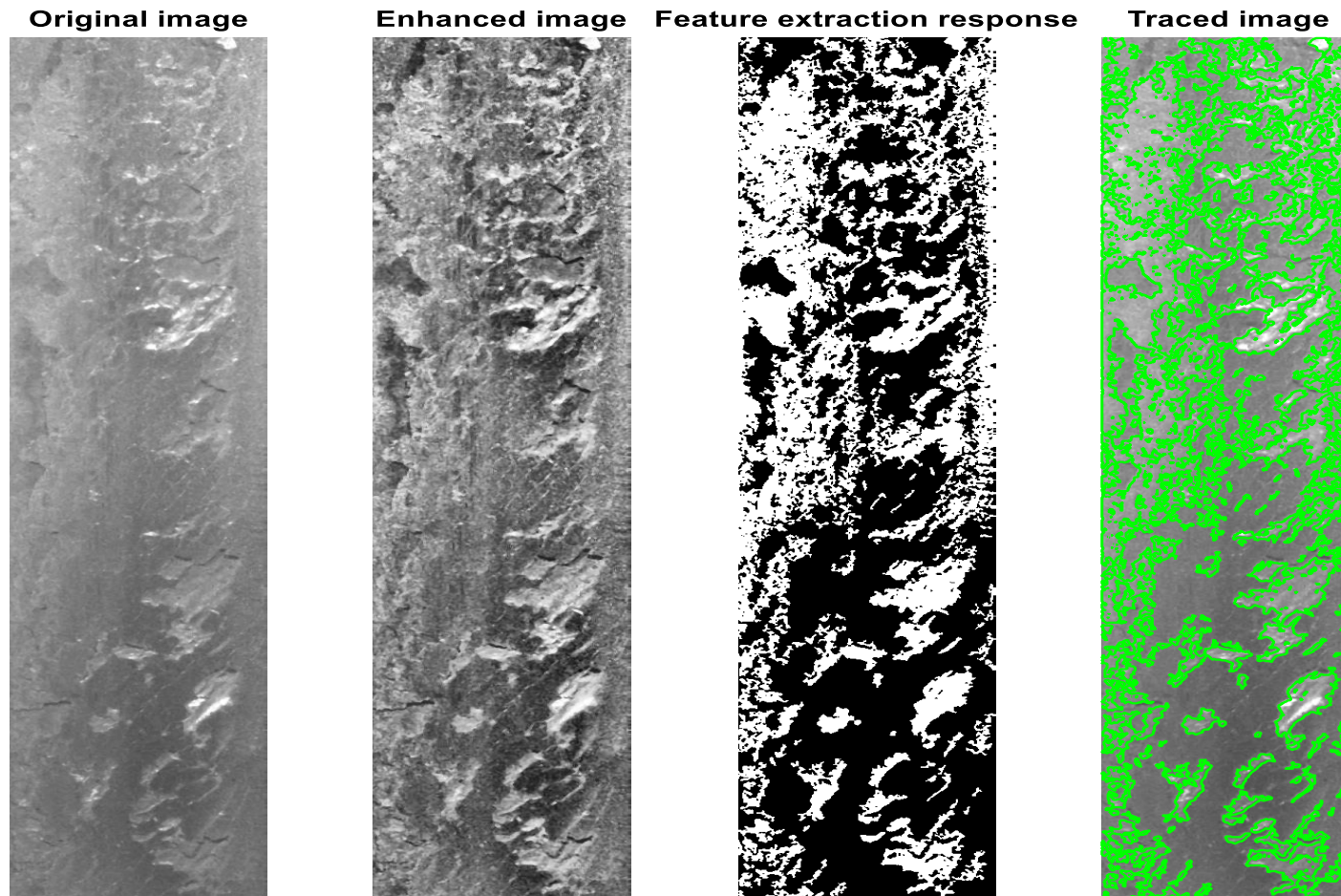
Figure A.C7. 5: Shows original, enhanced, extracted feature, and traced images for L-4270ft.

Original image **Enhanced image** **Feature extraction response** **Traced image**

Figure A.C7. 6: Shows original, enhanced, extracted feature, and traced images for L-3948ft.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**

Figure A.C7. 7: Shows original, enhanced, extracted feature, and traced images for L-3953ft.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**

Figure A.C7. 8: Shows original, enhanced, extracted feature, and traced images for L-3960ft.

Original image **Enhanced image** **Feature extraction response** **Traced image**

Figure A.C7. 9: Shows original, enhanced, extracted feature, and traced images for L-4103ft.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**

Figure A.C7. 10: Shows original, enhanced, extracted feature, and traced images for L-4563ft.

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**

Figure A.C7. 11: Shows original, enhanced, extracted feature, and traced images for L-3945ft.

Original image **Enhanced image** **Feature extraction response** **Traced image**

Figure A.C7. 12: Shows original, enhanced, extracted feature, and traced images for L-3978ft.

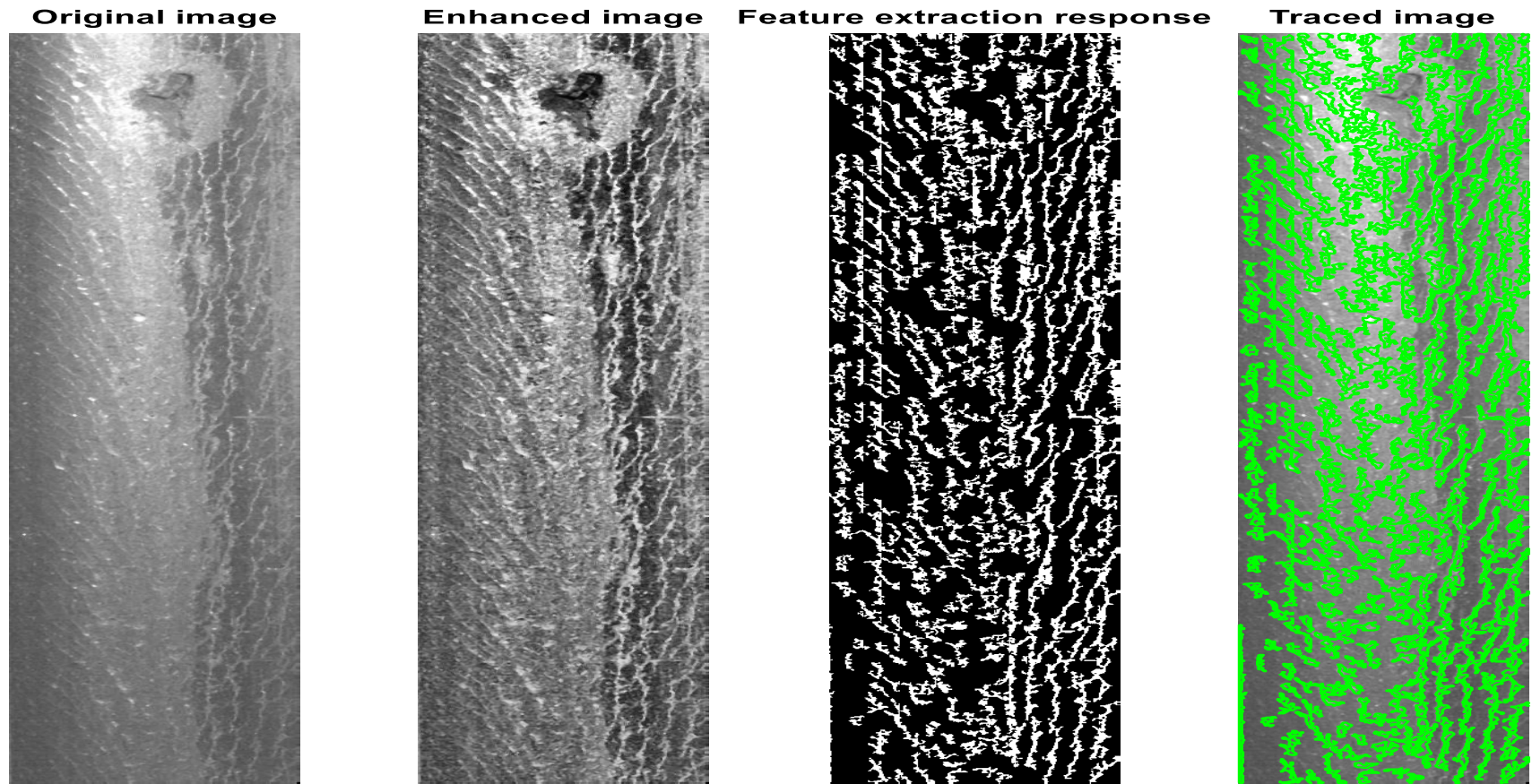**Original image** **Enhanced image** **Feature extraction response** **Traced image**

Figure A.C7. 13: Shows original, enhanced, extracted feature, and traced images for L-4025ft.
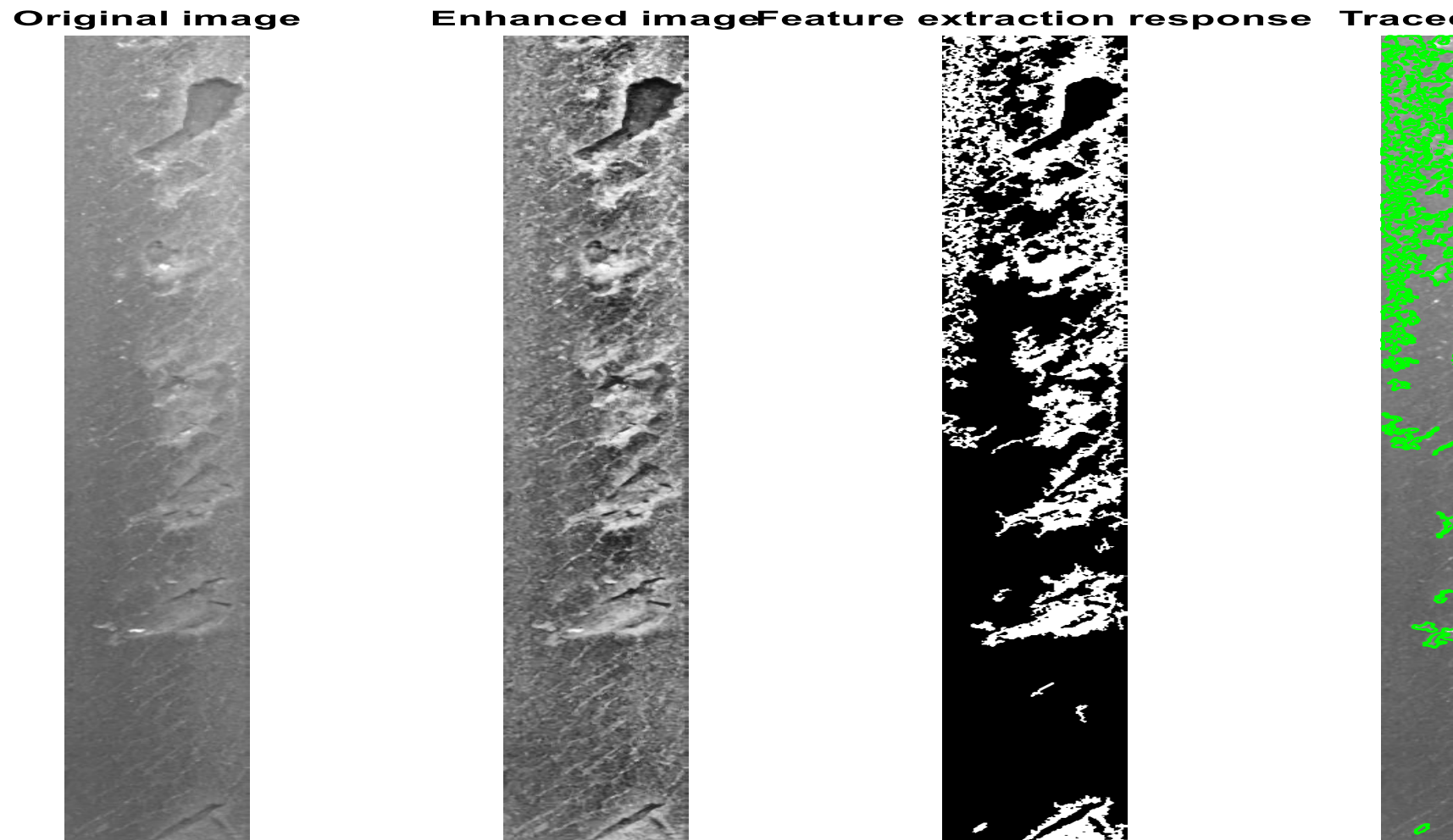
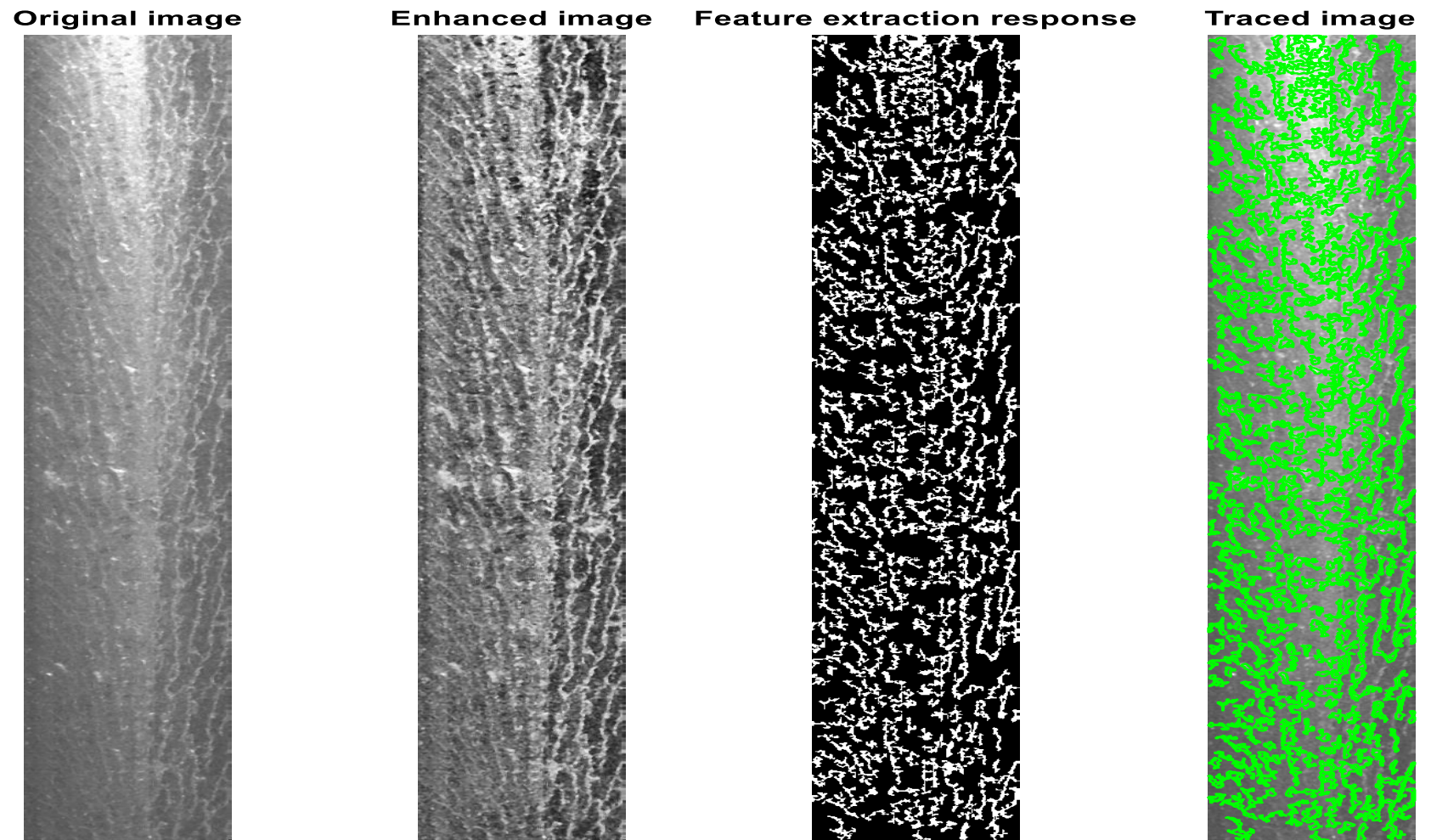**Figure A.C7. 14: Shows original, enhanced, extracted feature, and traced images for L-4025ft.**
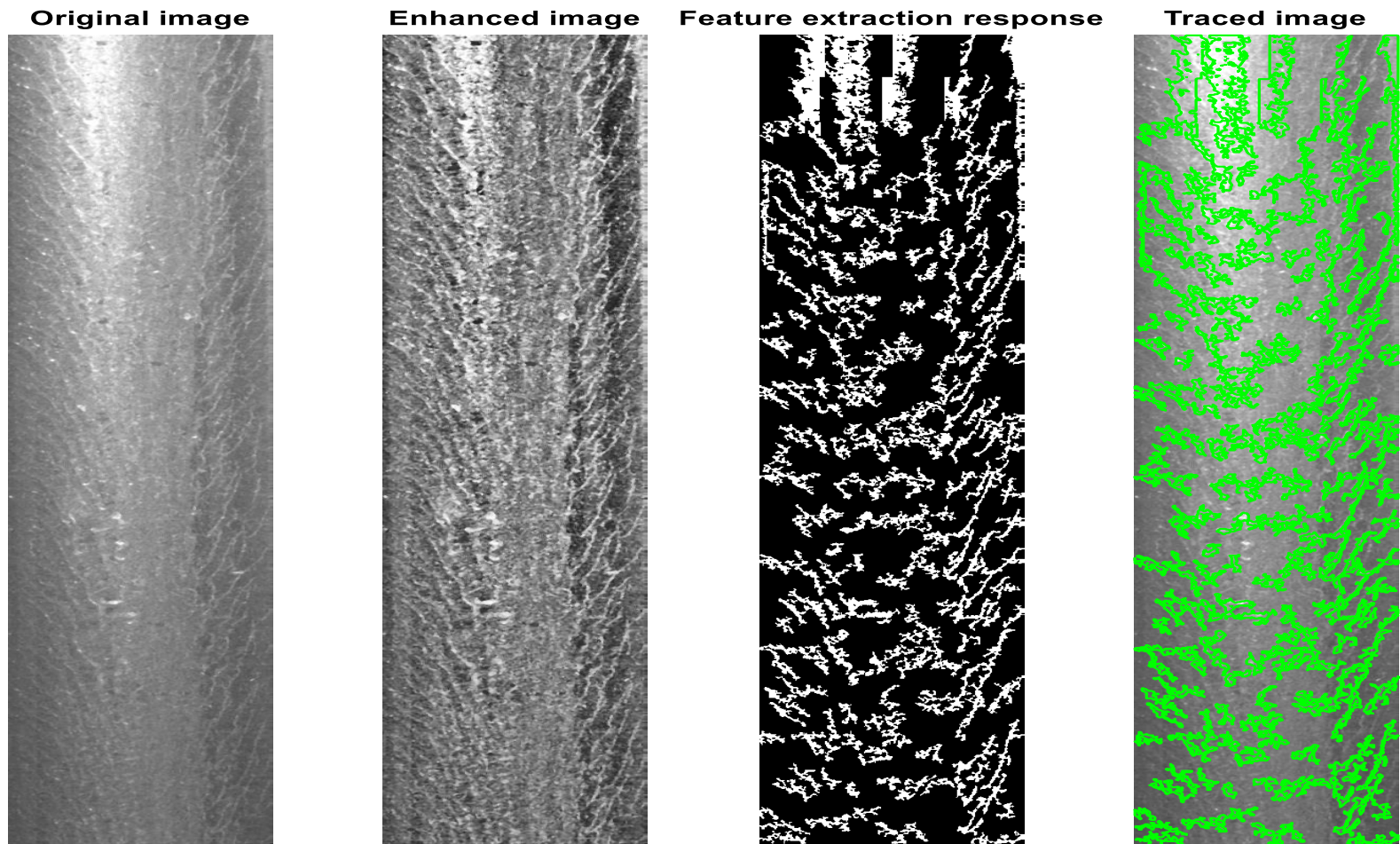
Figure A.C7. 15: Shows original, enhanced, extracted feature, and traced images for L-4213ft.
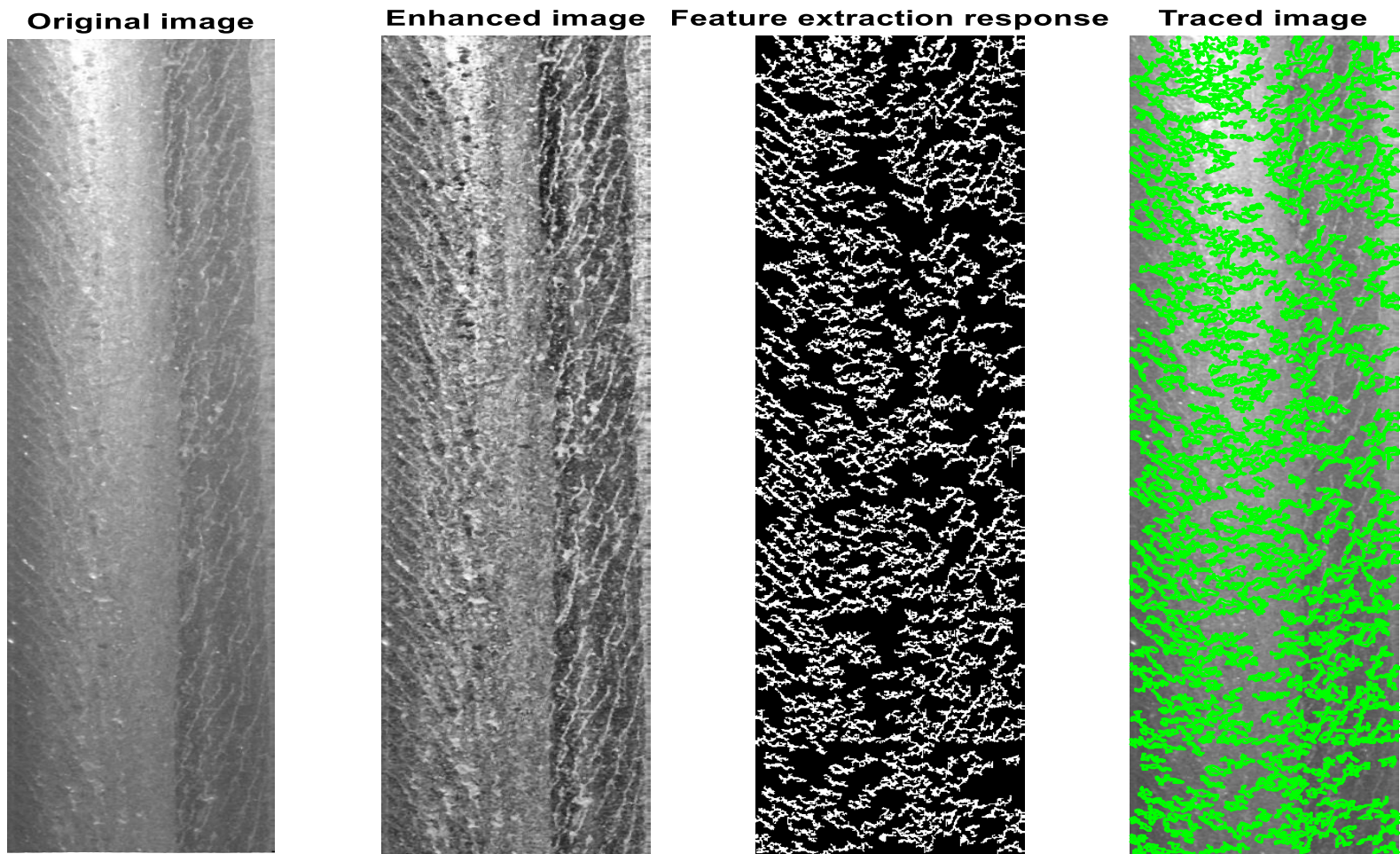
Figure A.C7. 16: Shows original, enhanced, extracted feature, and traced images for R-4060ft.
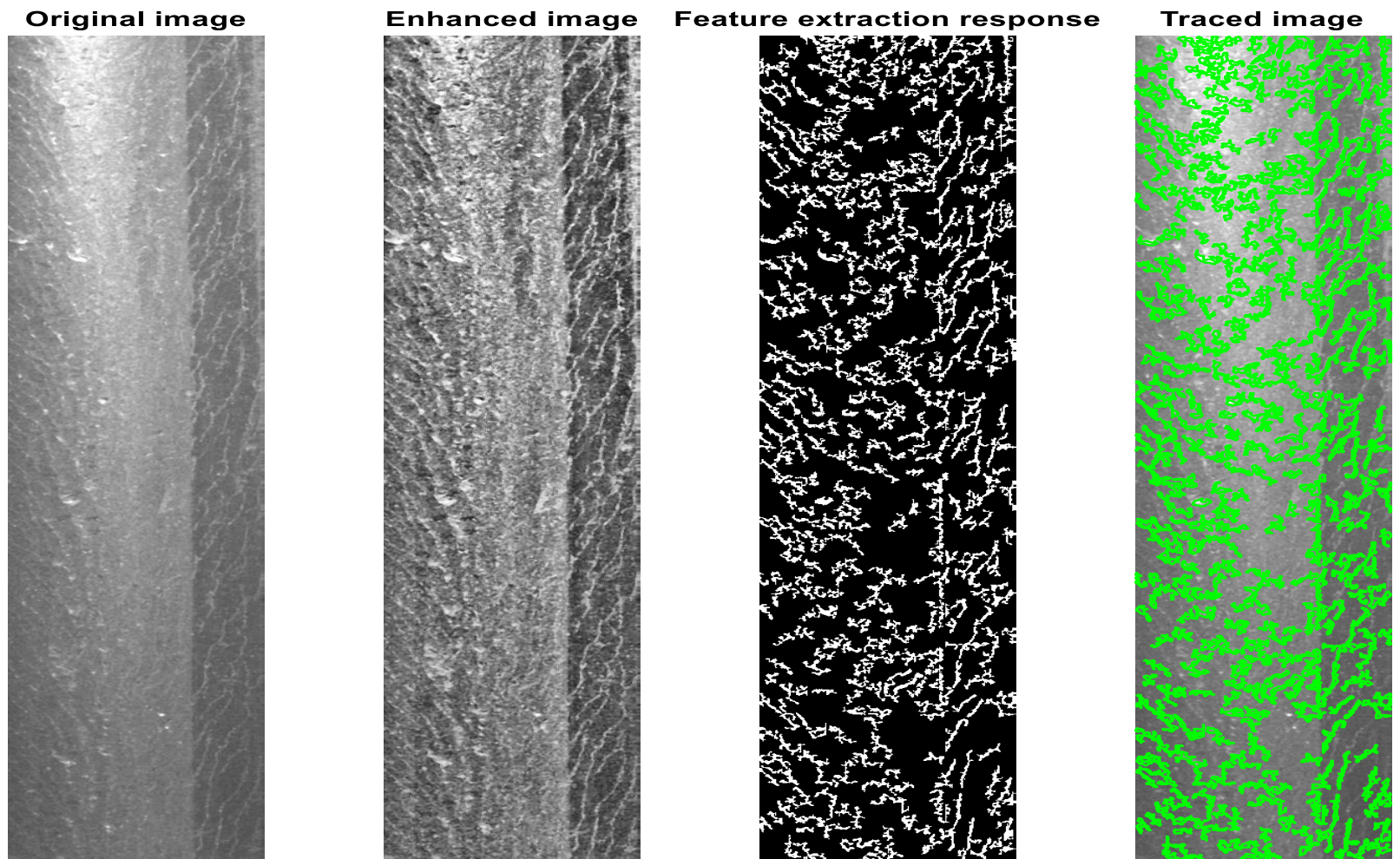
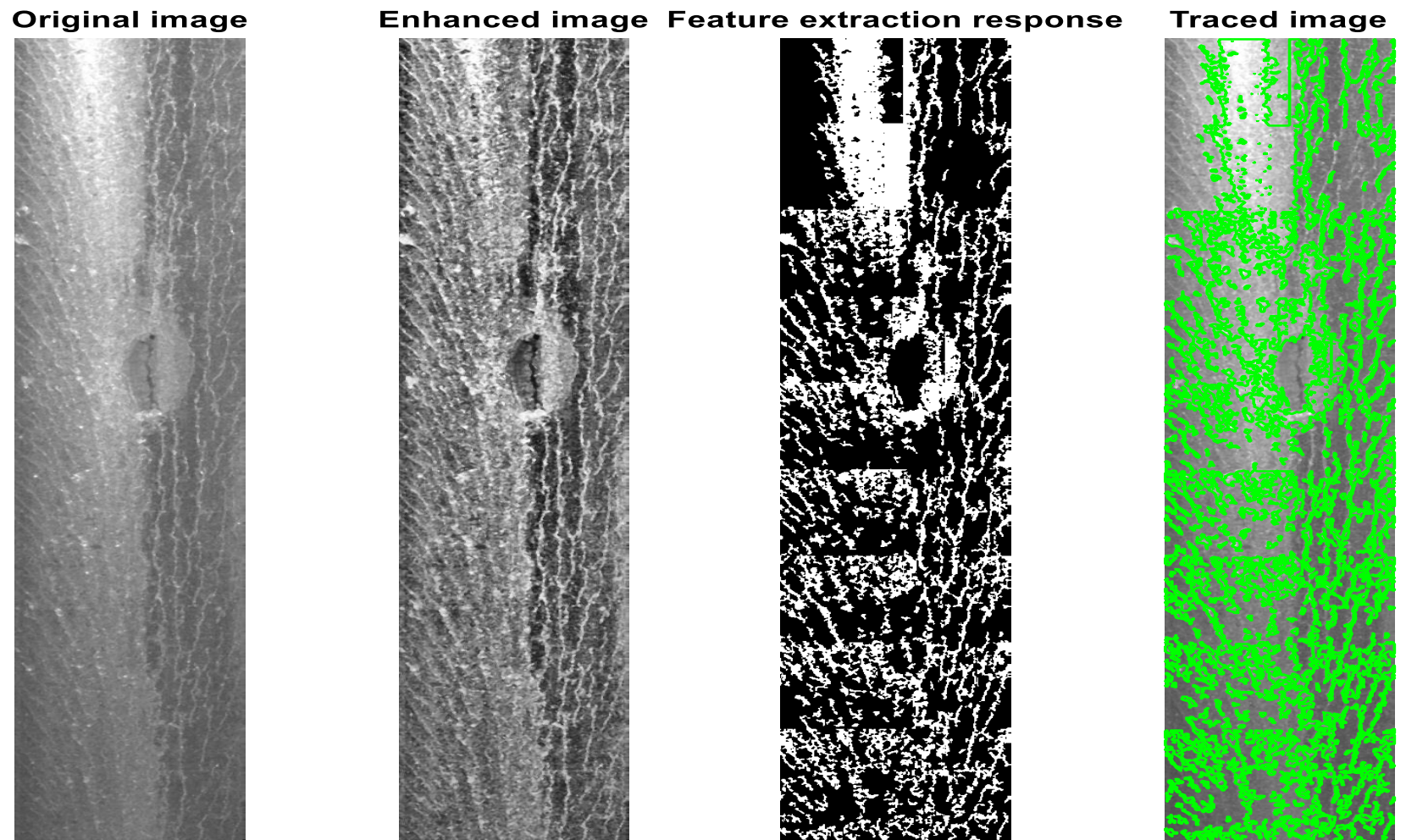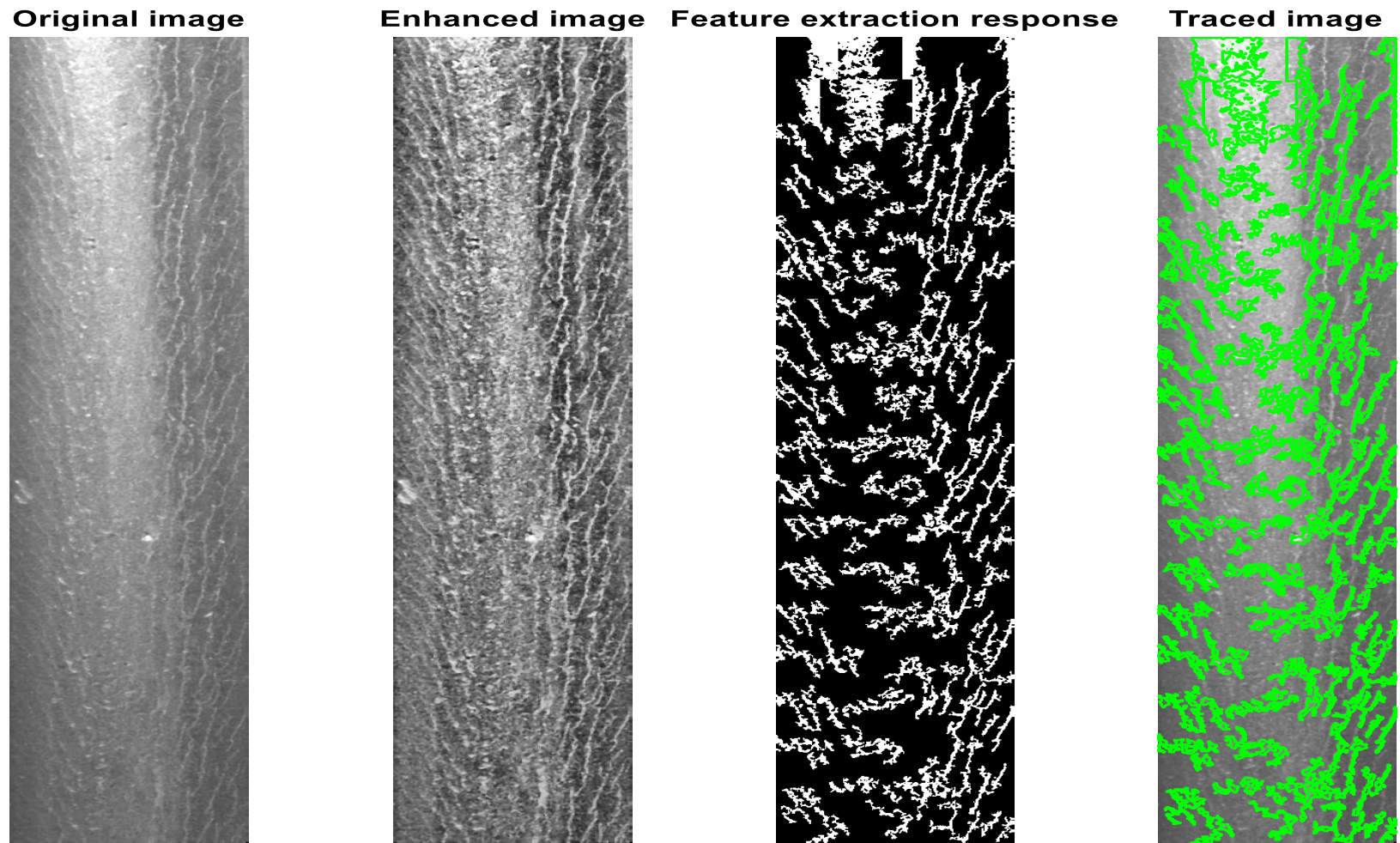**Figure A.C7. 17: Shows original, enhanced, extracted feature, and traced images for R-4063ft.**

Original image **Enhanced image** **Feature extraction response** **Traced image**
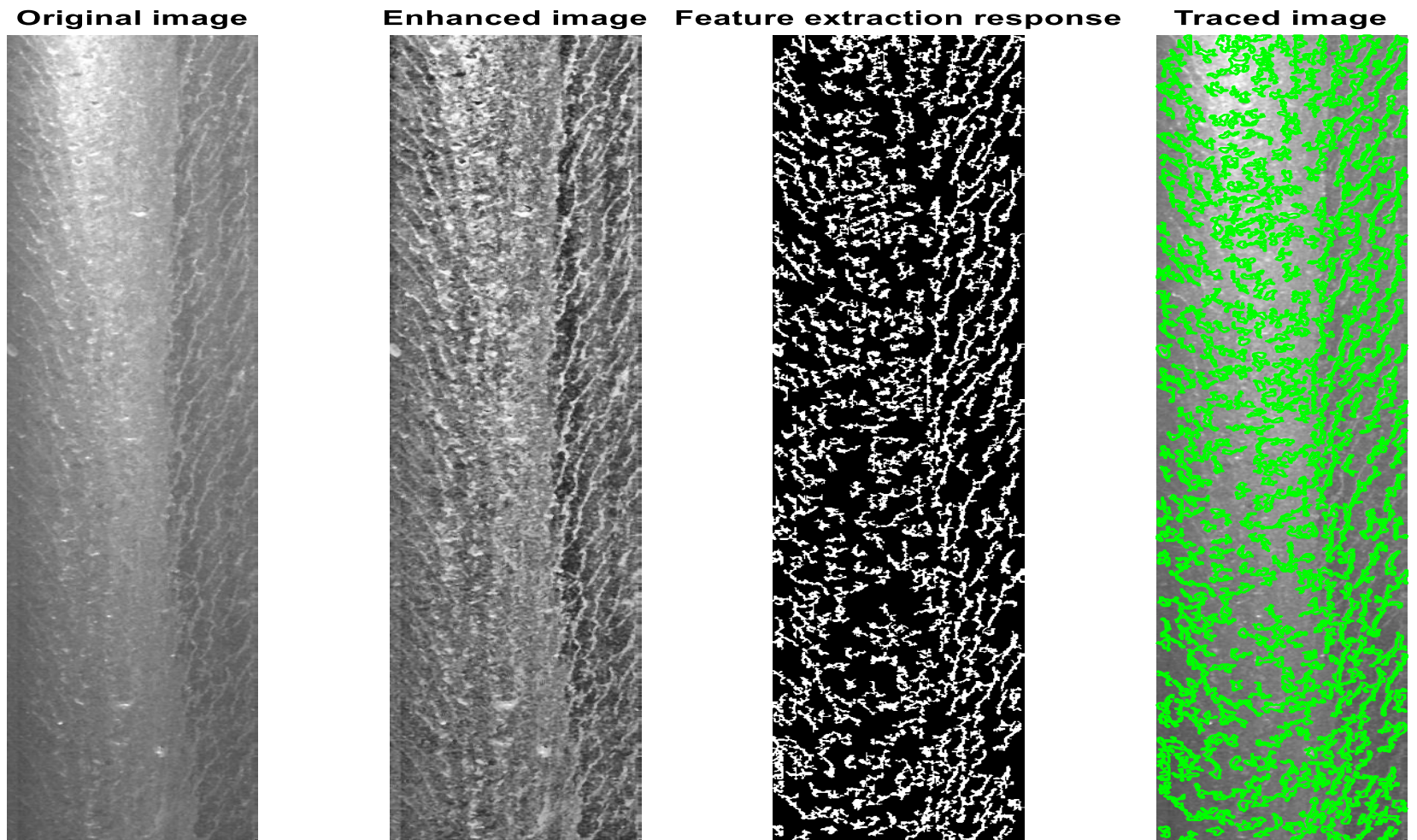
Figure A.C7. 18: Shows original, enhanced, extracted feature, and traced images for R-4110ft.

Original image　　**Enhanced image**　**Feature extraction response**　　**Traced image**
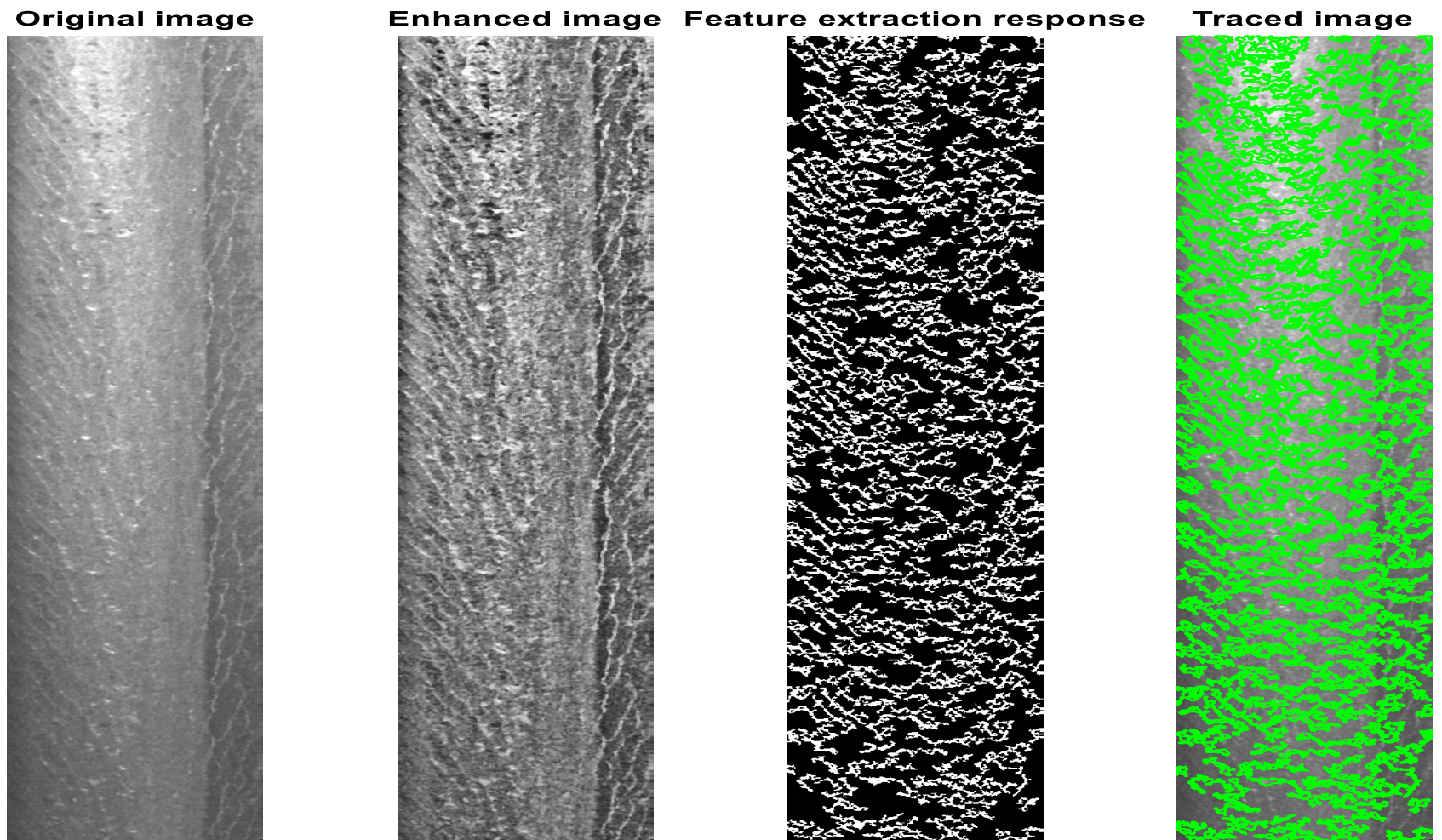
Figure A.C7. 19: Shows original, enhanced, extracted feature, and traced images for R-4195ft.
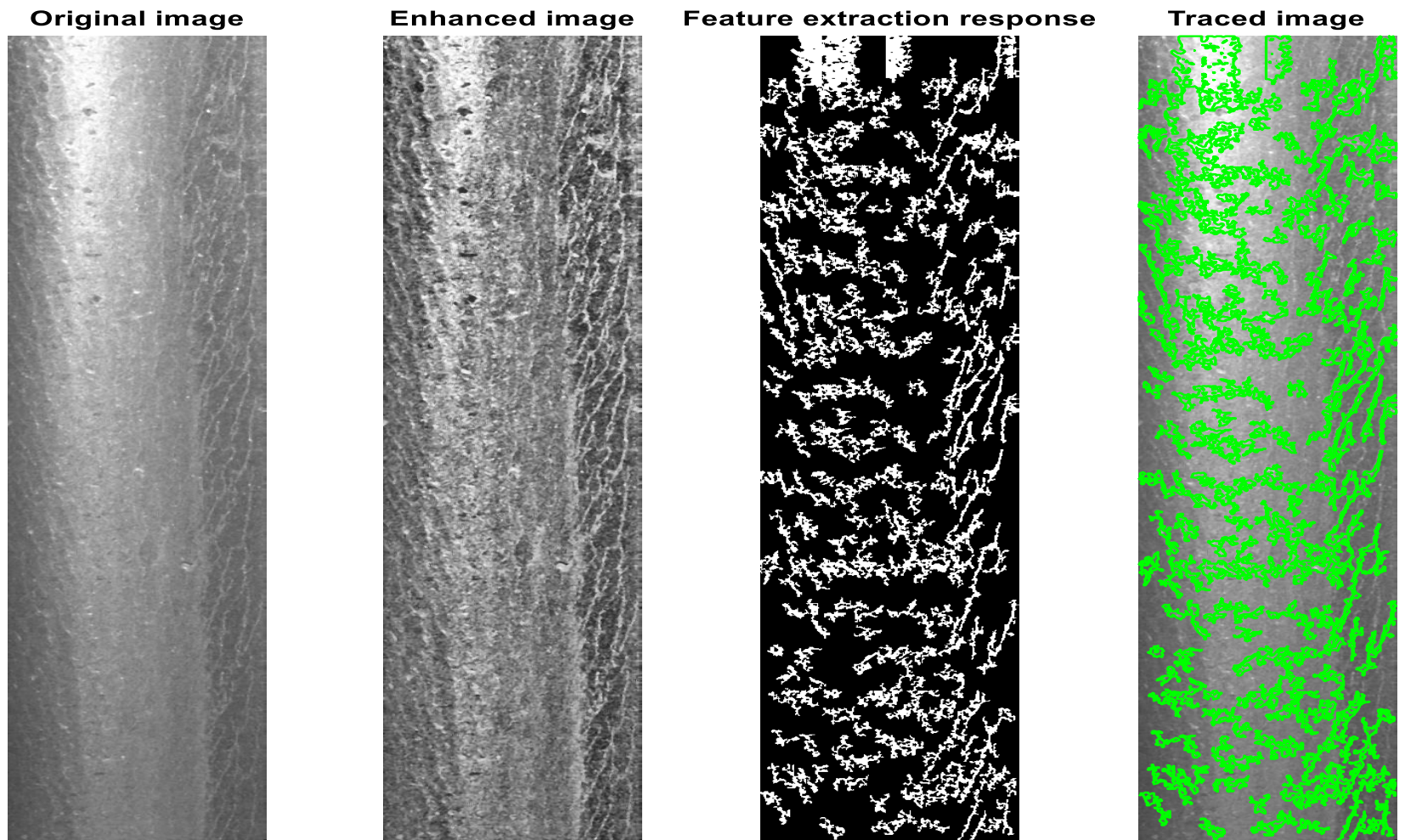
Figure A.C7. 20: Shows original, enhanced, extracted feature, and traced images for R-4270ft.

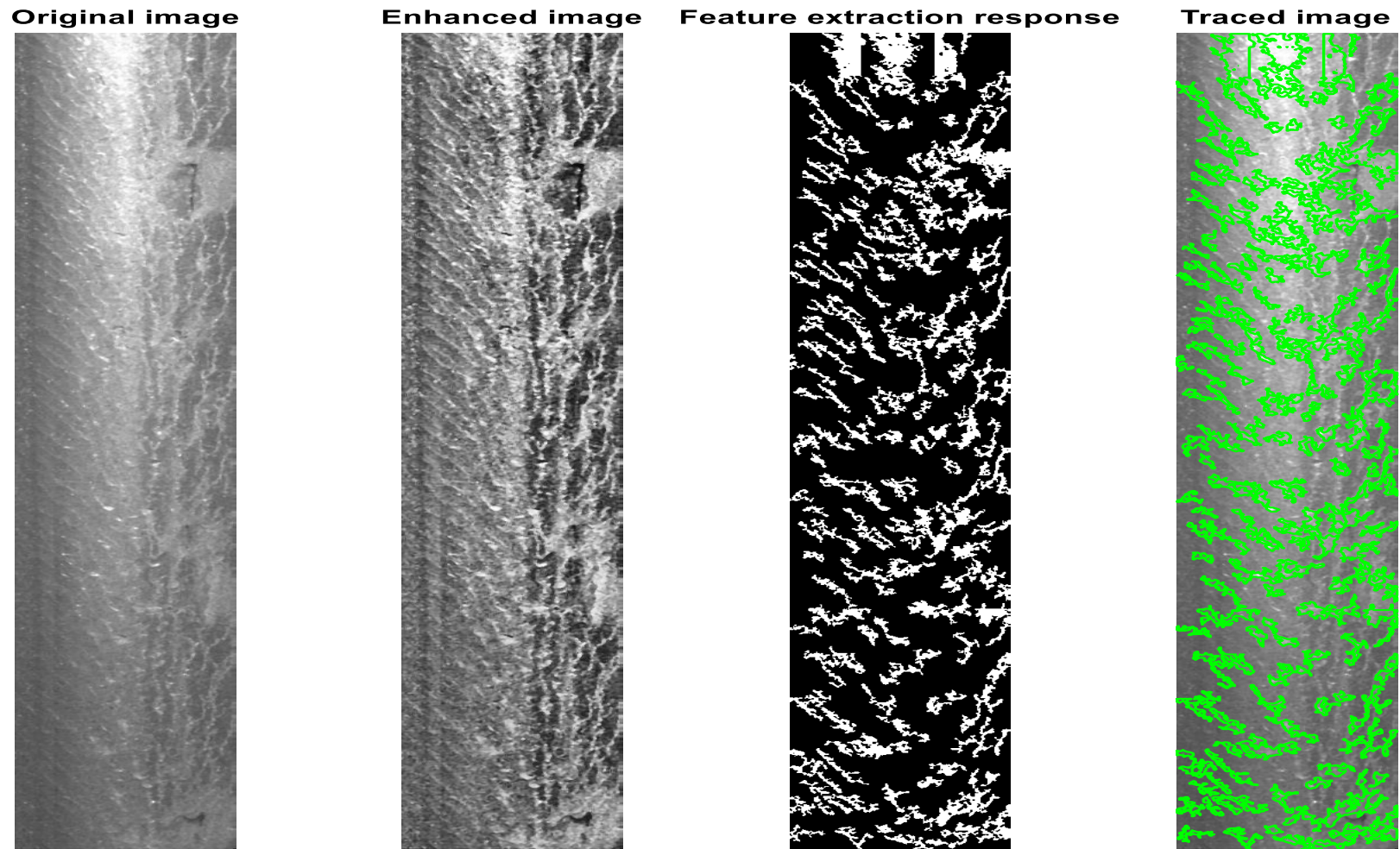**Figure A.C7. 21: Shows original, enhanced, extracted feature, and traced images for R-43948ft.**

Original image   **Enhanced image**   **Feature extraction response**   **Traced image**
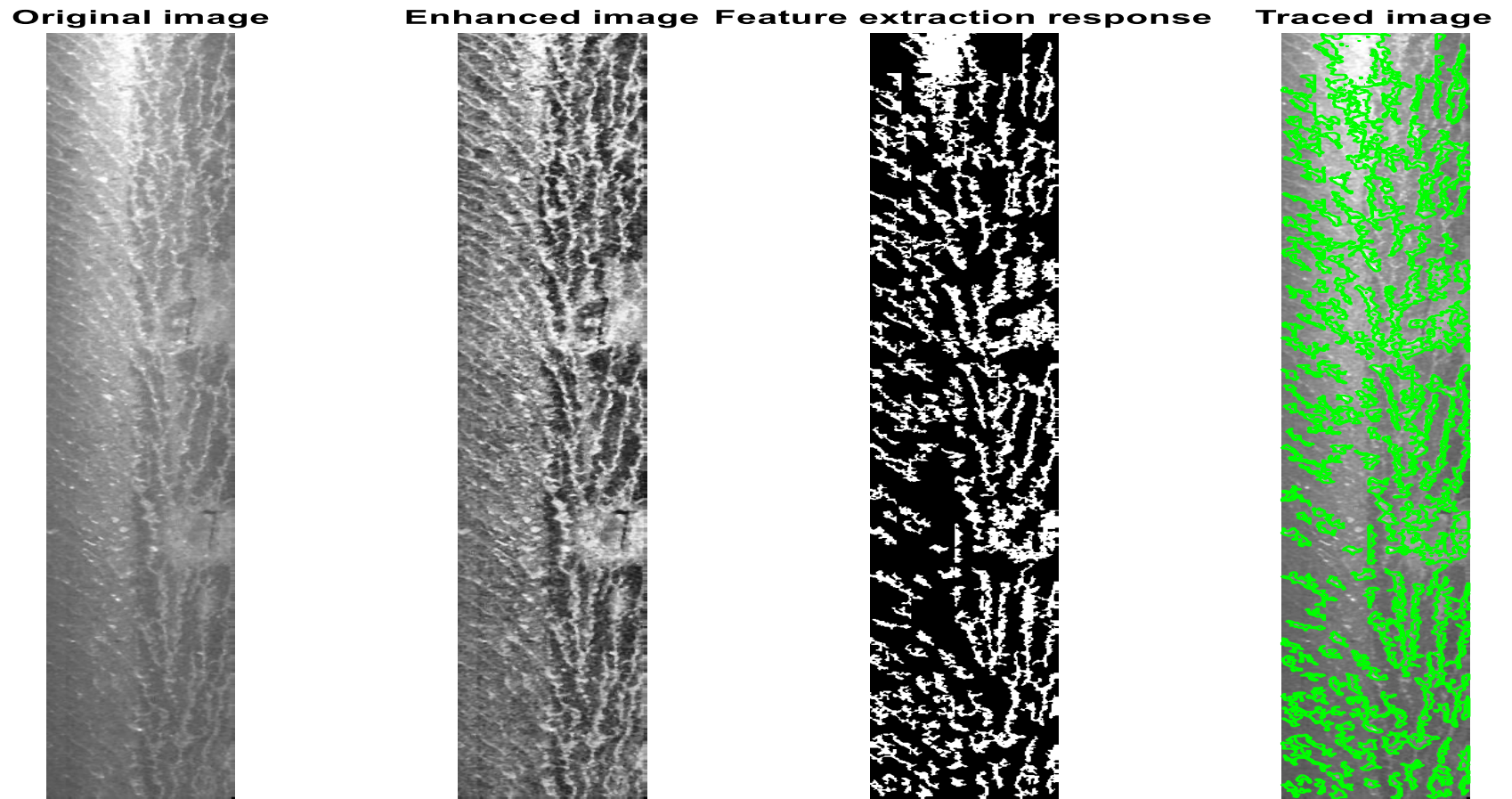

Figure A.C7. 22: Shows original, enhanced, extracted feature, and traced images for R-3953ft.

Figure A.C7. 23: Shows original, enhanced, extracted feature, and traced images for R-3960ft.

Figure A.C7. 24: Shows original, enhanced, extracted feature, and traced images for R-4103ft.

Figure A.C7. 25: Shows original, enhanced, extracted feature, and traced images for R-4563ft.
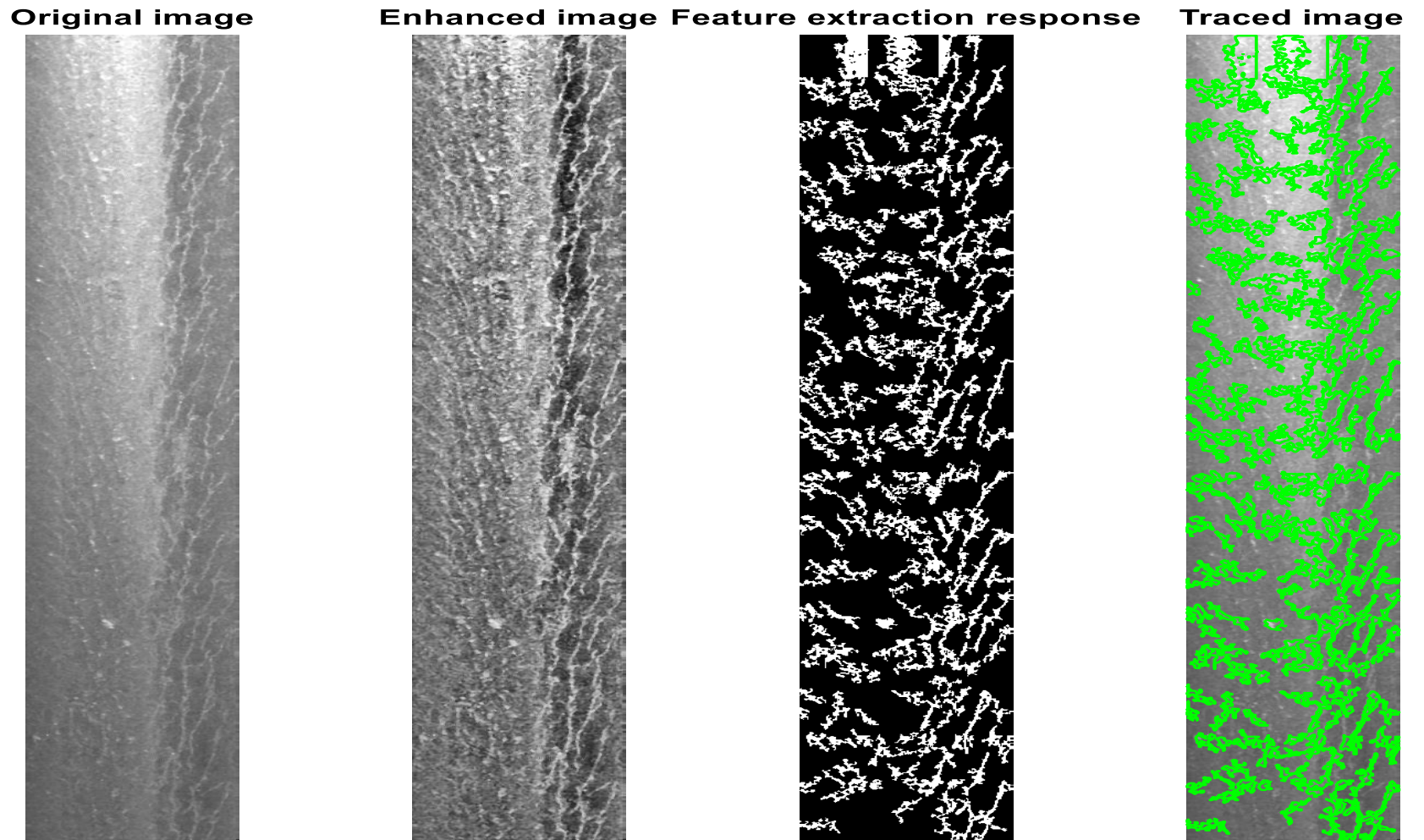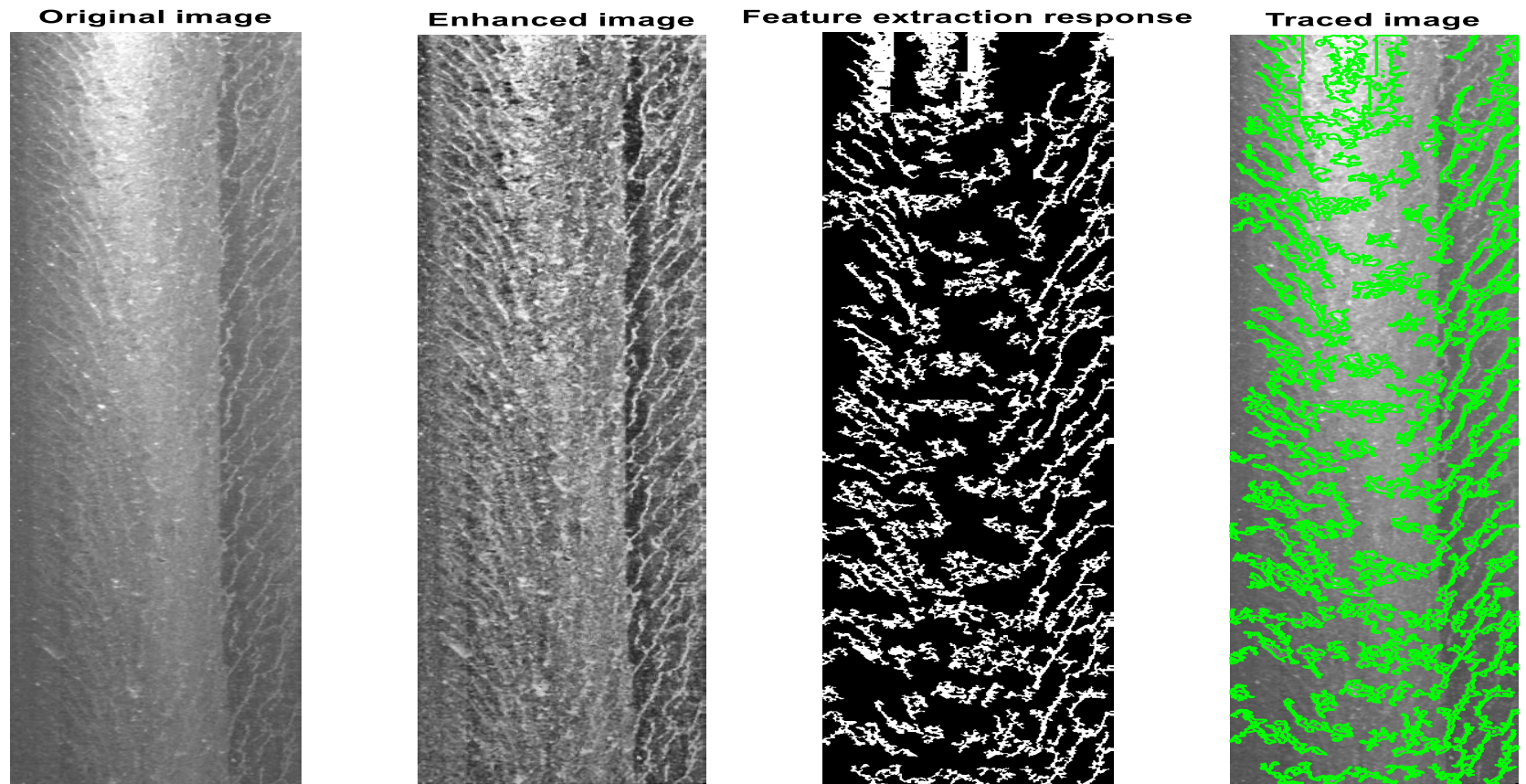
Original image   **Enhanced image** **Feature extraction response**   **Traced image**

Figure A.C7. 26: Shows original, enhanced, extracted feature, and traced images for R-3943ft.

Original image **Enhanced image** **Feature extraction response** **Traced image**

Figure A.C7. 27: Shows original, enhanced, extracted feature, and traced images for R-3945ft.

Figure A.C7. 28: Shows original, enhanced, extracted feature, and traced images for R-3978ft.
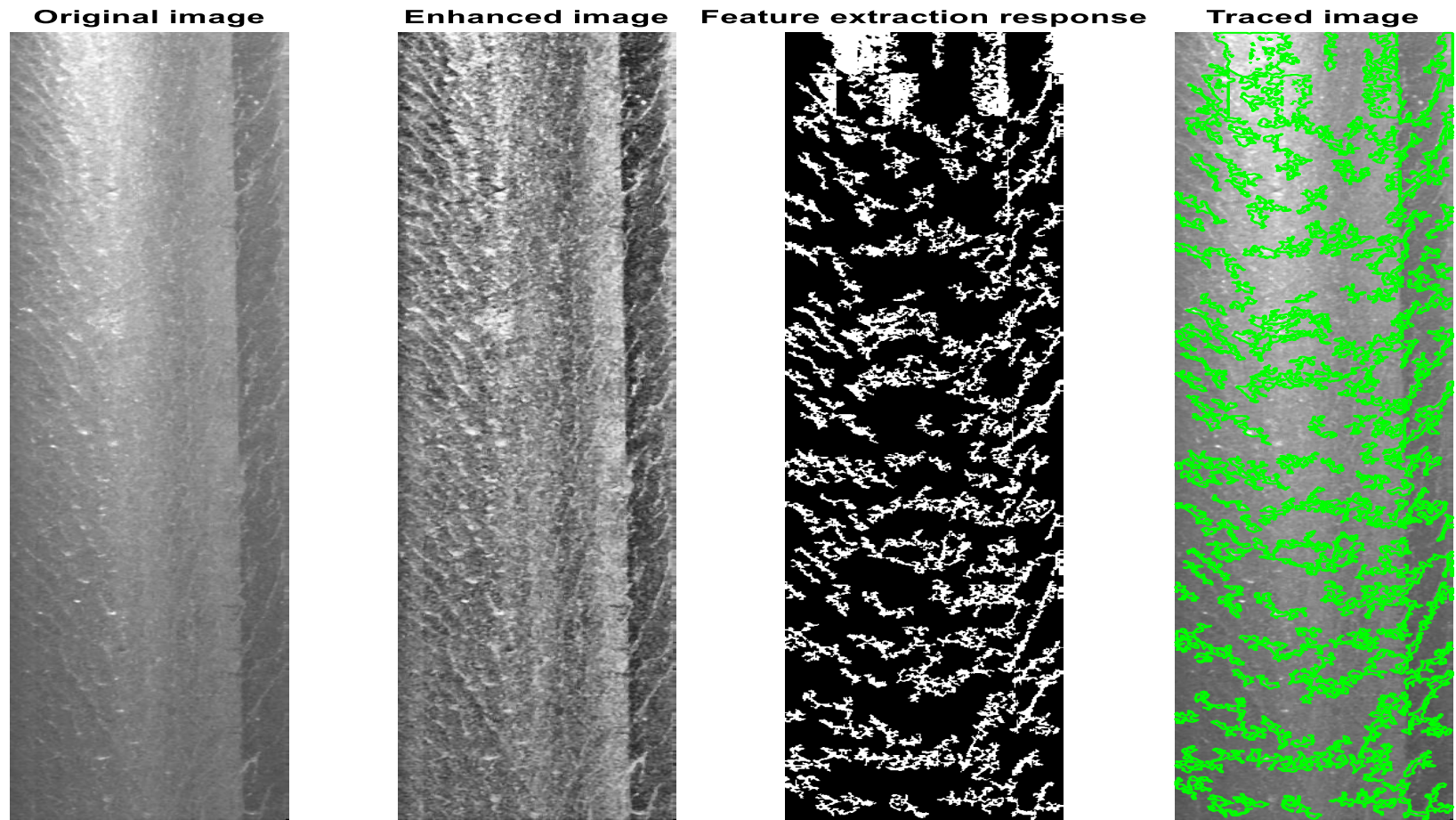
Figure A.C7. 29: Shows original, enhanced, extracted feature, and traced images for R-4025ft.

Original image **Enhanced image** **Feature extraction response** **Traced image**
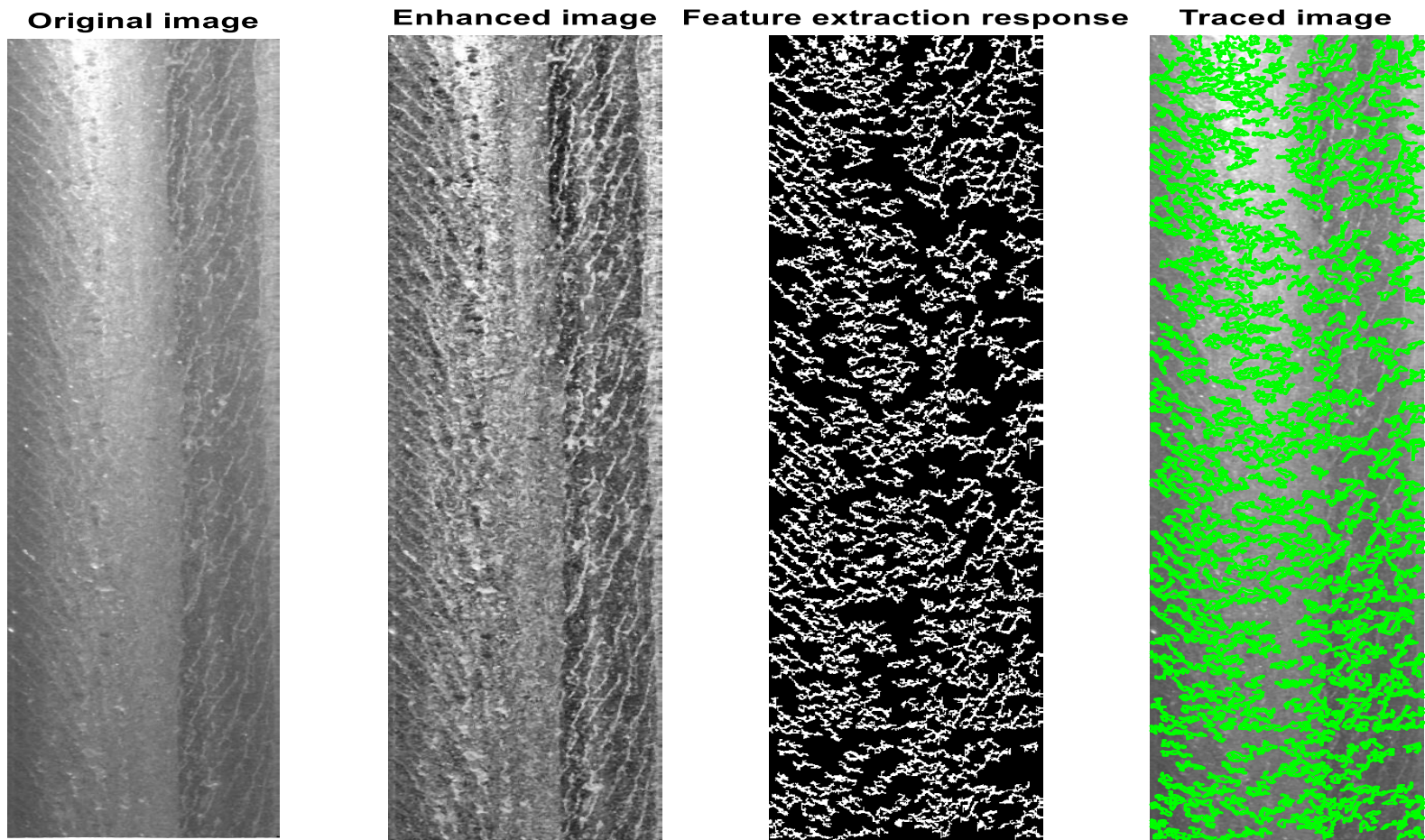

Figure A.C7. 30: Shows original, enhanced, extracted feature, and traced images for R-4213ft.

Appendix D: Intermediary validation of local fracture mechanics model

Appendix D1: Pure Mode-I and Mode-II SIF formulations

Pure mode I and II loading case studies are performed on a cube geometry ($10 \times 10 \times 10\ mm$) containing defect types a-c. These defect types have long served as fundamental building blocks of fracture mechanics investigations. The choice of an appropriate defect type depends on the scope of the research, examples in existing literature include:

a) Triangular edge through crack is mostly applied in the case of 2D analysis of rail web fracture mechanics Kotsikos, G., & Grasso, M. (2012).

b) Elliptical edge through crack also applicable to 2D analysis of rail foot fracture mechanics Jeong, D., & Orringer, O. (1989)

c) Elliptical centre crack applicable for 2D or 3D analysis of railhead fracture mechanics. Livieri, P., & Seagala, F. (2012).

The FE models are independently loaded for mode I and II fracture as depicted with a 500N boundary and edge load in incremental steps of 50N (see Figure A.D1.1 and A.D1.2.). The boundary loads are applied on the cubes top and bottom plane in positive and negative z-axis direction respectively. While the edge load is applied on the edges parallel to the crack major semi-axis on the front plane of the cube. A fixed constraint is applied on the edges of the end plane. Such that pure Mode-I case utilises the edges parallel to the major semi-axis while the pure mode-II case utilises the edges parallel to the minor semi-axis. The combined effect of the pure mode I and II loading is established using equivalent SIF of equation 4.23. The outputs of SIF mathematical models in equations 4.4 and 4.14, are further used to account for the influence of considering only the length of the crack or also including the opening of the crack mouth respectively. For defect types a, b and c, the crack length is set to 5mm with an opening of 1mm, while depth of 10mm, 10mm and 2.5mm is adopted respectively.

Mode-I

Mode-I

(a)   (b)   (c)
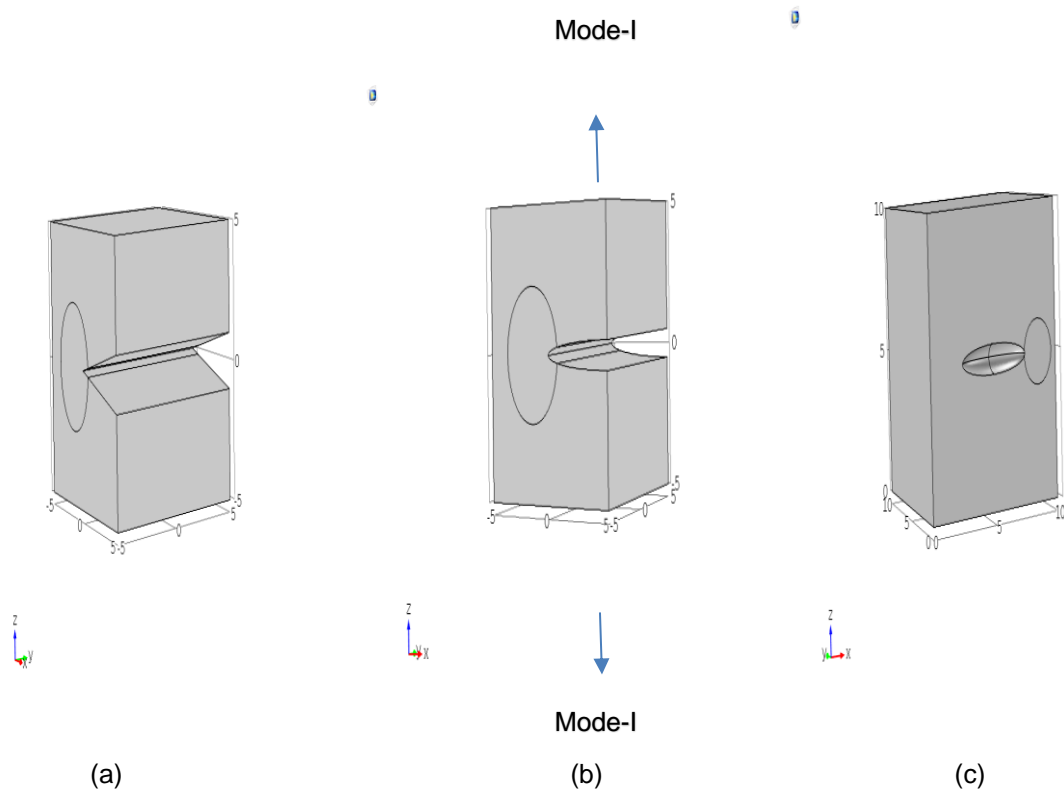
Figure A.D1. 1: Pure tensile fracture mechanism of edge triangular through crack (a), edge elliptical through crack (b), and centred elliptical thickness crack (c).
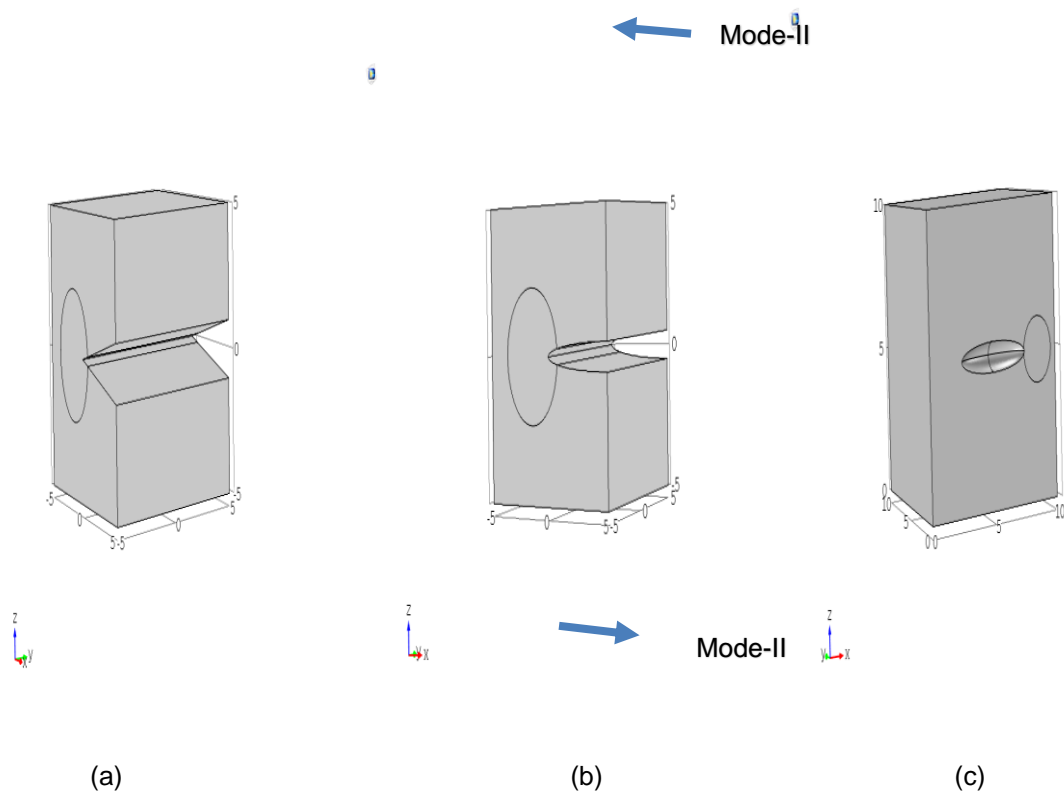
Mode-II

Mode-II

(a)   (b)   (c)

Figure A.D1. 2: Pure shear fracture mechanism of edge triangular through crack (a), edge elliptical through crack (b), and centred elliptical thickness crack (c).

It can be seen from the FE simulated results of Figure A.D1.3 and A.D1.4 below, that the peak stress induced in both loading cases is dominant at the crack tip (for all crack types investigated).
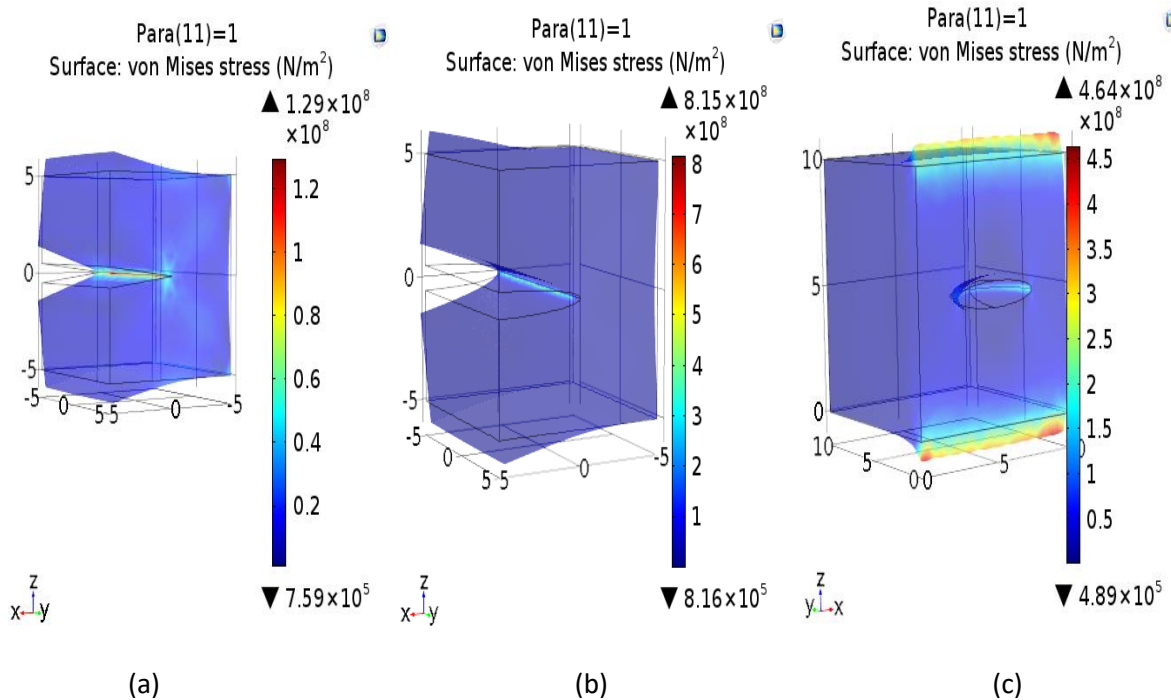

Figure A.D1. 3: Shows the simulated mode-I Von Misses stress distribution for of edge triangular through crack (a), edge elliptical through crack (b), and centred elliptical thickness crack (c).
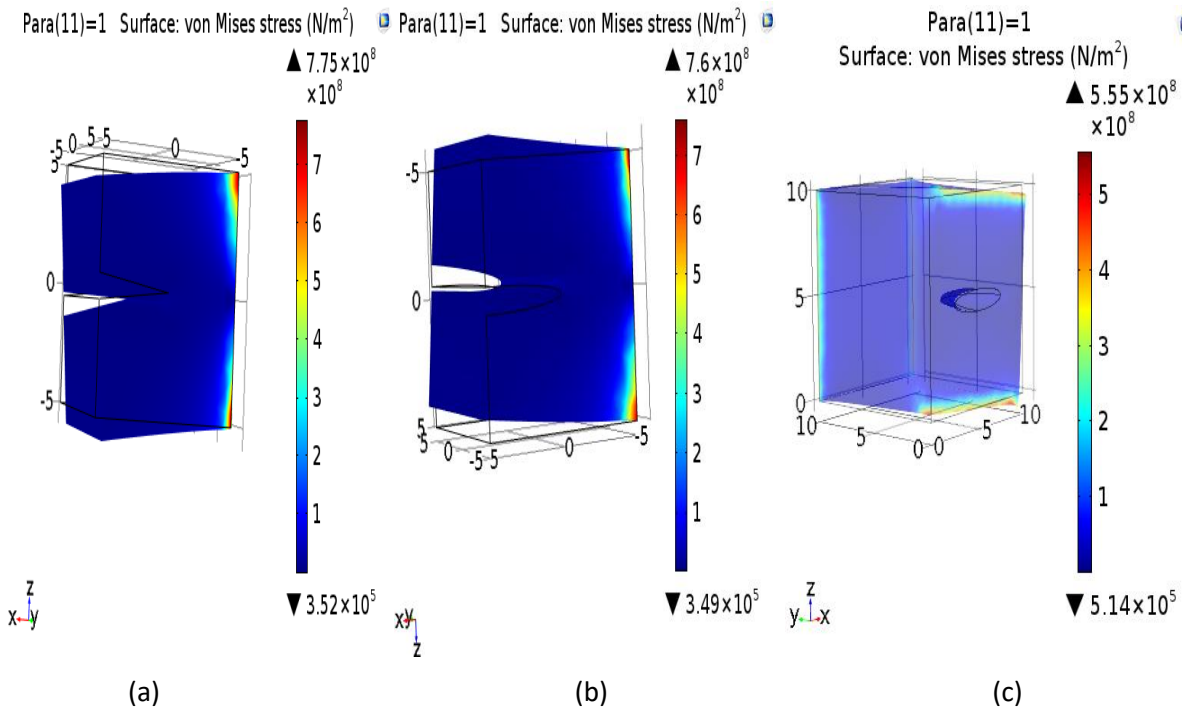

Figure A.D1. 4: Shows the simulated mode-II Von Misses stress distribution for of edge triangular through crack (a), edge elliptical through crack (b), and centred elliptical thickness crack (c).

At maximum load of 500N, defect types a-c has approximate KI values of 70MPa, 70MPa, and 45MPa respectively. Comparing the KI derived from equation 4.4a and 4.14a in Figure A.D1.5 suggests that a variation of no more than 4%. This variation is because of including the influence of crack mouth opening in equation 4.14a. It is also noticed that 5MPa depreciation in KI from the results of defect type a, and that of defect type b is due to the influence of crack tip radius difference between a triangle and an ellipse. While defect type c with similar surface length and mouth opening as the other two defect types is observed to have a lower KI by approximately 25MPa. This variation is because of the difference in depth of about 7.5mm compared to defect type a and b.

At maximum load of 500N defect types a-c have approximate KII values of 57MPa, 19MPa, and 27MPa respectively. Comparing the KII of equations 4.4b and 4.14b, Figure A.D1.6 suggests that including the crack mouth opening increases the value of KII by about 7MPa, 1MPa, and 3MPa for defect type's a-c respectively. The only difference between defect type a and b is the crack tip radius, which is observed to be inversely proportional to amplitude of KII. For defect type's a-c under same stress amplitude, utilising same the load, is most sensitive to fixed constraint physics settings. Defect types a and b show a maximum variation in SIF of 25% when boundary fixed constraints are applied in place of the original edge fixed constraint. An exception is in the case of defect type c, where negligible variation is observed despite changes in fixed constraint settings. Other influential boundary conditions within the FE model include the distribution of meshing elements especially within the vicinity of the crack tip. By means of predefined mesh distribution (100 elements with an element ratio of 25SIF changes of no more than 12% is recorded.
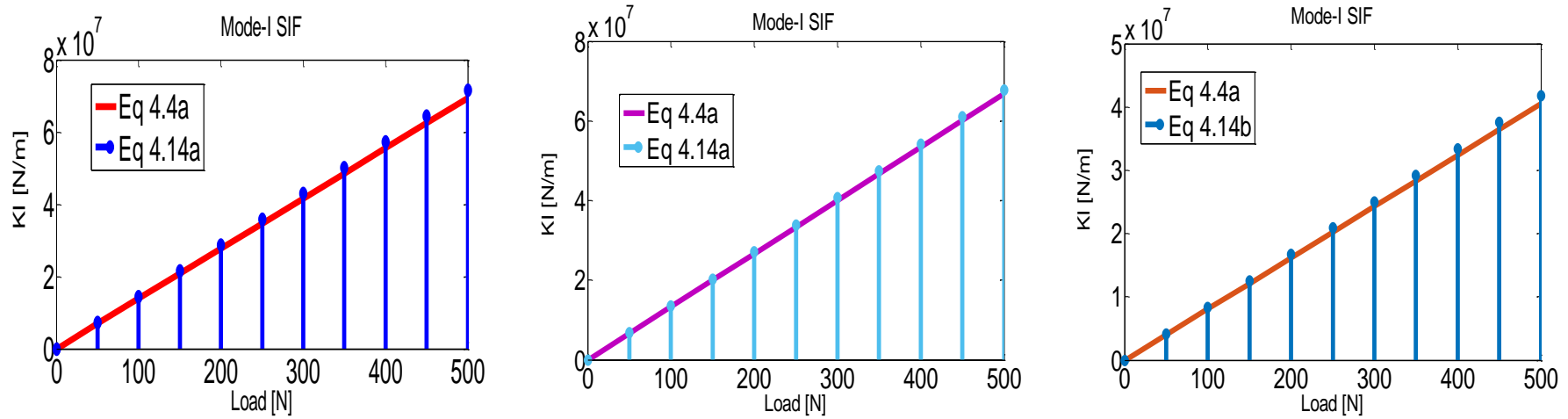
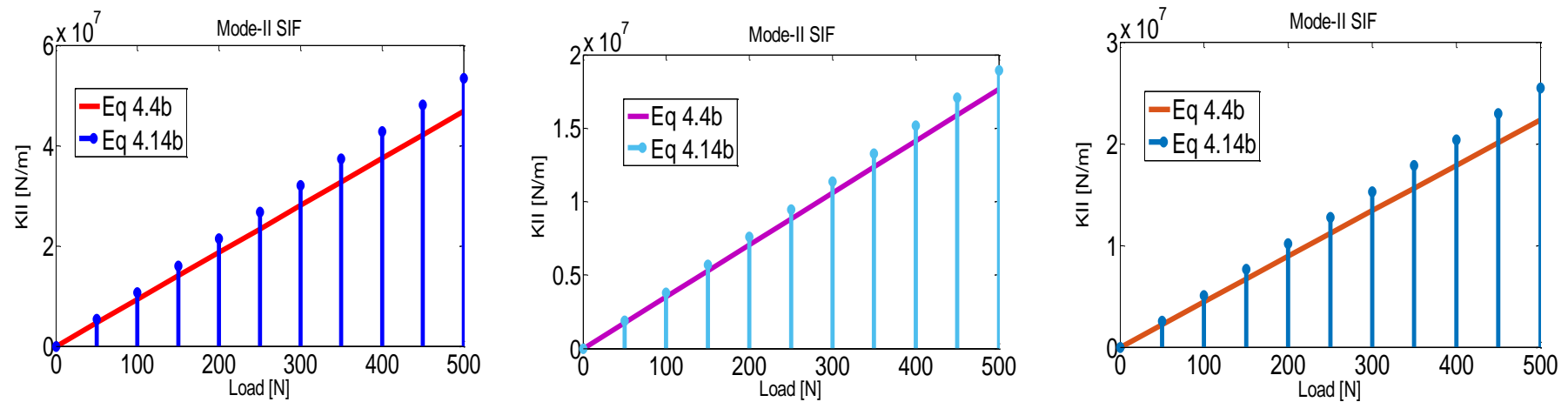**Figure A.D1. 5: Shows the simulated KI for defect type a(left), b(centre), and c(right).**

Figure A.D1. 6: Shows the simulated KII for defect type a (left), b(centre), and c(right).

The Keqv for defect type a-c in Figure A.D1.7 is observed to have a maximum value of about 70MPa, 60.5MPa, and 40MPa respectively. Finer mesh element distribution within the vicinity of the crack tip (100 elements with an element ration of 25), has been observed to result in a variation of no more than 12%. The influence of replacing boundary fixed constraint with that of an equivalent edge constraint corresponds to a maximum variation in the range of a few MPa for all defect types.

In Figure A.D1.8, the growth behaviour observed is synonymous to rapid fracture as crack approaches final failure. A closer observation of the simulated results confirms the least growth rate (of a no more than a 10nm/cycle) within the range 0-150N of applied load for all defect types. However, at applied loads ranging between 200N - 250N a considerable increase of no less than 500nm/cycle is recorded. In this range (200N-250N) a linear-like relationship exists between the applied load and growth rate. Moreover, above 50% maximum applied load, a rapid increase of approximately 1000nm/cycle in growth rate for every 100N loading step is recorded. It can also be deduced that at peak loading amplitude of 500N, a growth rate of 4900nm/cycle 3900nm/cycle and 600nm/cycle is recorded for defect type a-c respectively. These estimates of growth rate mean that defect type a and b propagate faster than defect type c. The difference of 1000nm/cycle between defect type a and b geometry, and crack tip radius (more pronounced defect type b and c, which is observed to have a lower KII. The above observations related to crack growth rate are consistent regardless of using mathematical models related to surface crack length only (equation 4.4) or that which includes the crack mouth opening as elaborated in equation 4.14. Replacing boundary with edge constraint corresponds to a growth rate variation in hundredths of nm/cycle for defect types a and b except in the case of defect type c where negligible variation is observed.
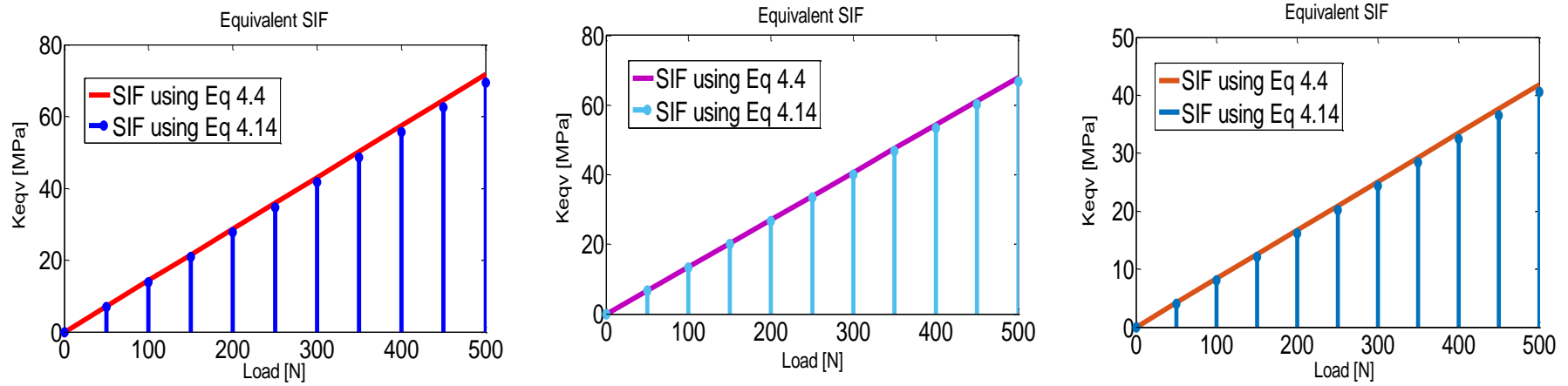
Figure A.D1. 7: Shows the simulated equivalent SIF (Keqv) for defect type a (left), b(centre), and c(right).
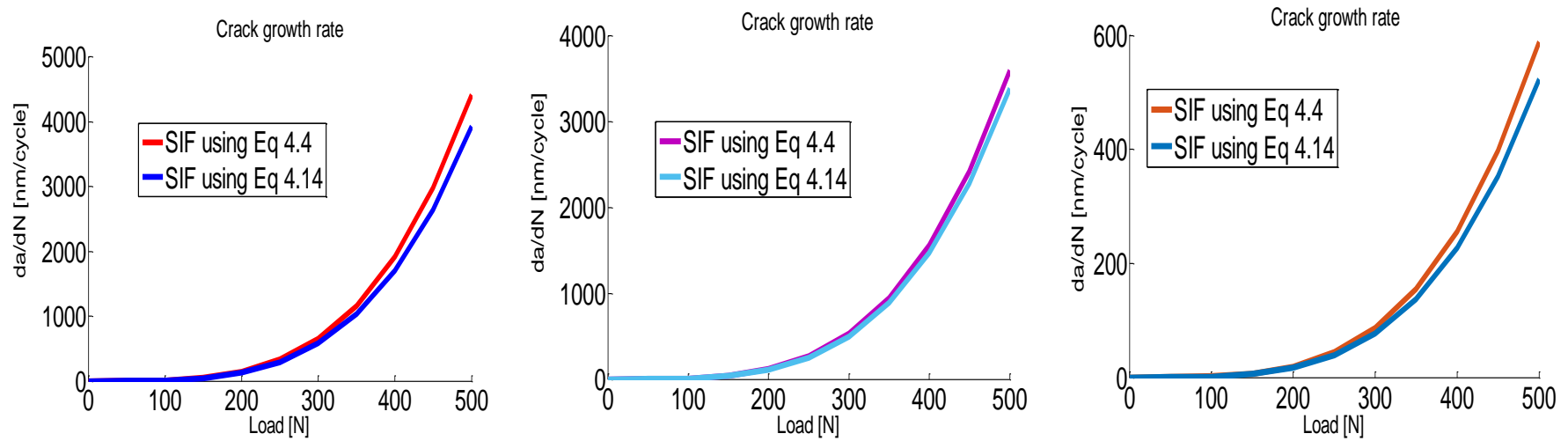

Figure A.D1 8: Shows the simulated crack growth rate for defect type a (left), b(centre), and c(right).

The propagation direction depicted in Figure A.D1.9 below, confirms that defect types a, b, and c have a growth path of 49^0, 52^0, and45^0. It is therefore reasonable to conclude that that under the same loading conditions regardless of the crack configuration, a variation of no more than 14% is attained in crack growth direction for cracks of same length and opening. It is also worth mentioning that a negligible variation (no more than 12%) is observed for predefined mesh distribution within the crack tip vicinity. While the influence of changing fixed constraint settings from edge to boundary condition is observed to decrease propagation direction by up to one fifth. The simulated results also confirm that the inclusion of crack minor semi-axis length in SIF estimation (using equation 4.14) to determine crack growth path results in a decrease in estimated direction of propagation of no more than 3^0. A closer observation of the simulated propagation direction of defect type a, suggests that a linear relationship exists between growth path ranging from 0° - 25° due to an applied load 0-50N. Above this 50N to about 300N of applied load, a constant growth path is maintained. However above 300N it is observed that another proportional increase in branching angle with respect to load occurs with an overall peak value of 49deg attained at 350N. Above this applied load a constant branching direction is maintained up till maximum applied load of 500N is reached. A similar observation is made in the case of defect type b. However, an increase of about 10^0 in the amplitude of the initial linear growth direction (in the range of 0-50N) is observed, while this branch angle totalling to about 35^0 is maintained for the second range of above 50N to 300N. Also for applied loads of above 300N to about 350N an increase in amplitude to about 52^0 is recorded, while above this applied load until maximum value of 500N, the branching angle is kept constant at 52^0. In the case of defect type c, a constant branching angle of less than 0.7^0 is observed despite the increase in applied load. This behaviour is attributed to the depth configuration of the crack. In comparing the behaviour of crack growth path for defect type a-c, it is obvious that defect type a and b propagate approximately at an angle almost 50 times that of defect type c.

In Figure A.D1.10, the simulated number of stress cycles until failure is presented and it can be observed that the number of loading cycle to fracture for an edge through crack regardless of its tip radius is 13 times lower than that of a centred elliptical thickness crack. Defect type a and b are observed to have same maximum residual life span of 6E5 cycles. While defect type c is observed to have a maximum residual life span of 8E6 cycles. The number of cycles for all defect types is further observed to have an inverse relationship with the applied load for a range of 0-200N, while above 50% of the applied load the number of cycles to fracture decays to almost zero, which is indicative of tendency for crack failure to occur.
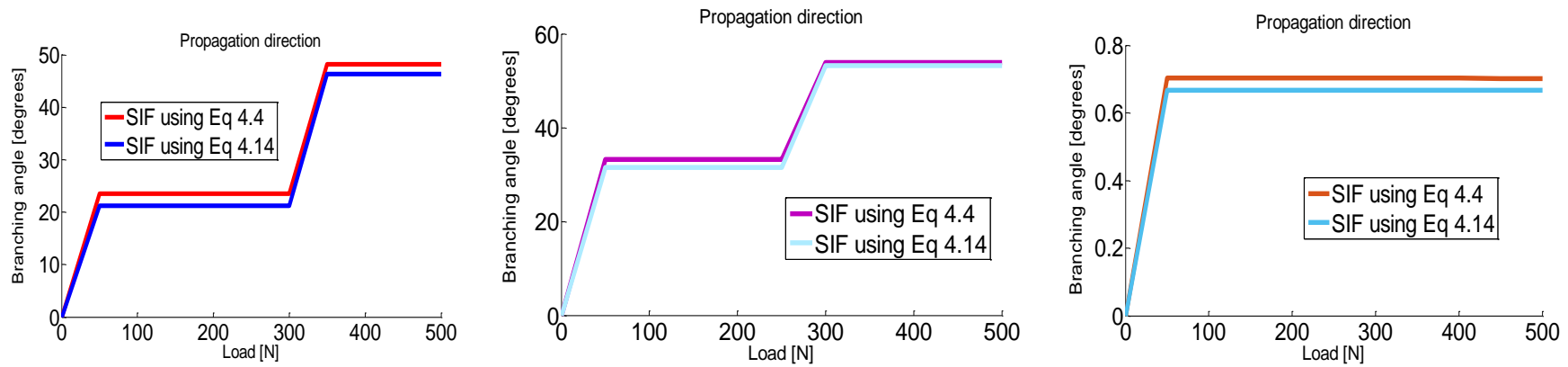
Figure A.D1 9: Shows the simulated crack growth direction for defect type a (left), b(centre), and c(right).
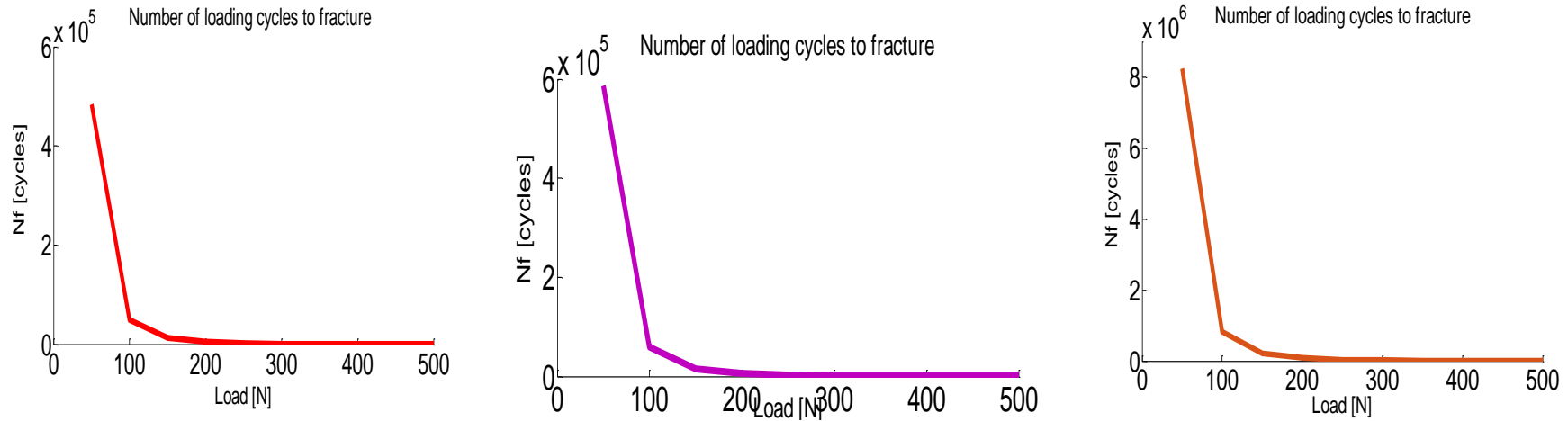

Figure A.D1 10: Shows the simulated number of loading cycles until failure for defect type a (left), b(centre), and c(right).

The influence of inclining defect type 'c' under the influence of uniaxial tension is presented. Under same FE solid mechanics settings as described in section 4.4.1. It is observed in Figure A.D1.11 that pure mode-I fracture mechanism $\beta = 90^0$ and mixed mode loading case of $\beta = 75^0$ have the fastest growth rate of 550nm/cycle and 545nm/cycle respectively. While pure mode-II loading $\beta = 45^0$ and mixed mode case of $\beta = 60^0$ gives the least growth rate of 50nm/cycle and 180nm/cycle respectively.
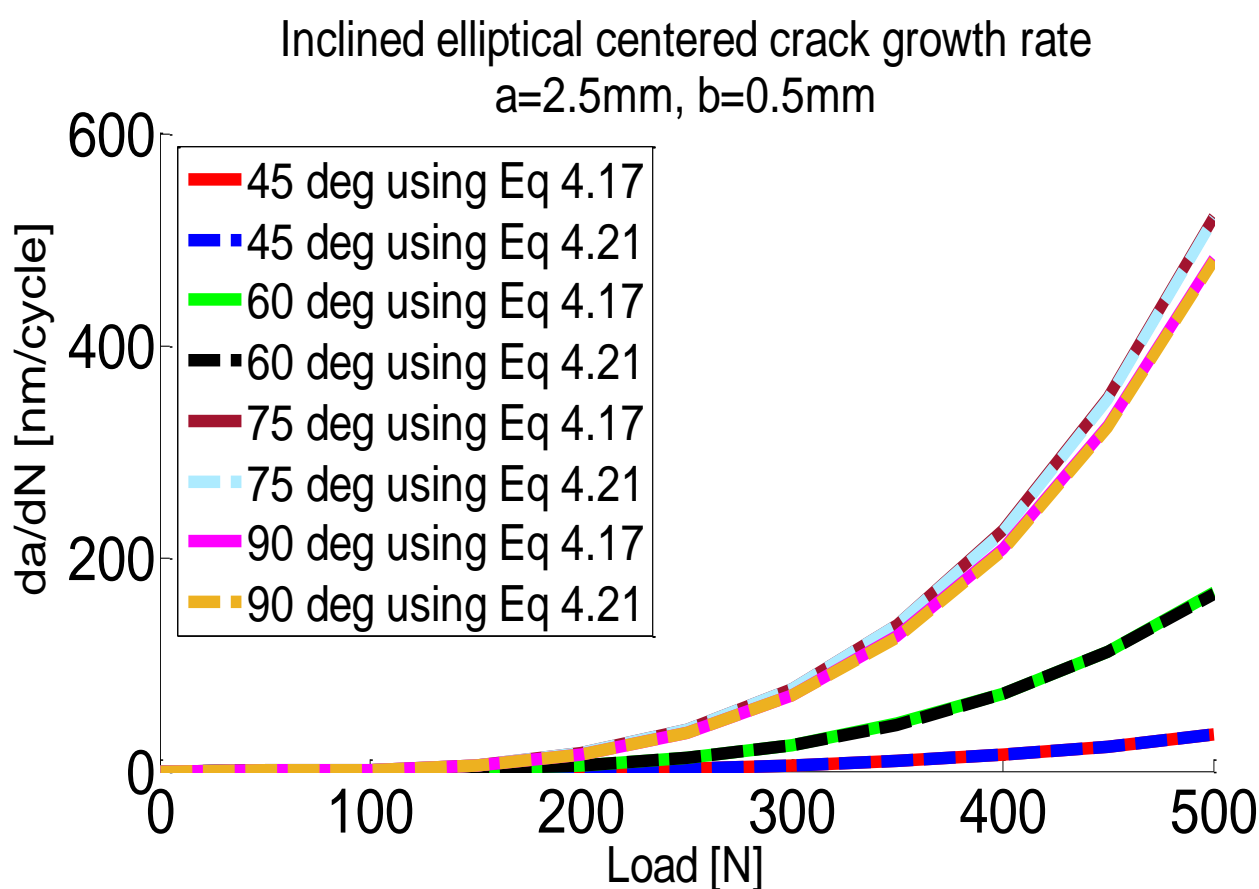
Figure A.D1 11: Influence of changing inclination angle (β) on growth rate for defect type 'c'.

In Figure A.D1.12 below it is observed branching angle is inversely proportional to inclination angle. The maximum growth path of 55^0, $45^0, 30^0$, and 0^0 is coincident with pure mode-II loading $\beta = 45^0$, mixed mode $\beta = 60^0$, $\beta = 75^0$ and pure mode I case of $\beta = 90^0$ respectively.
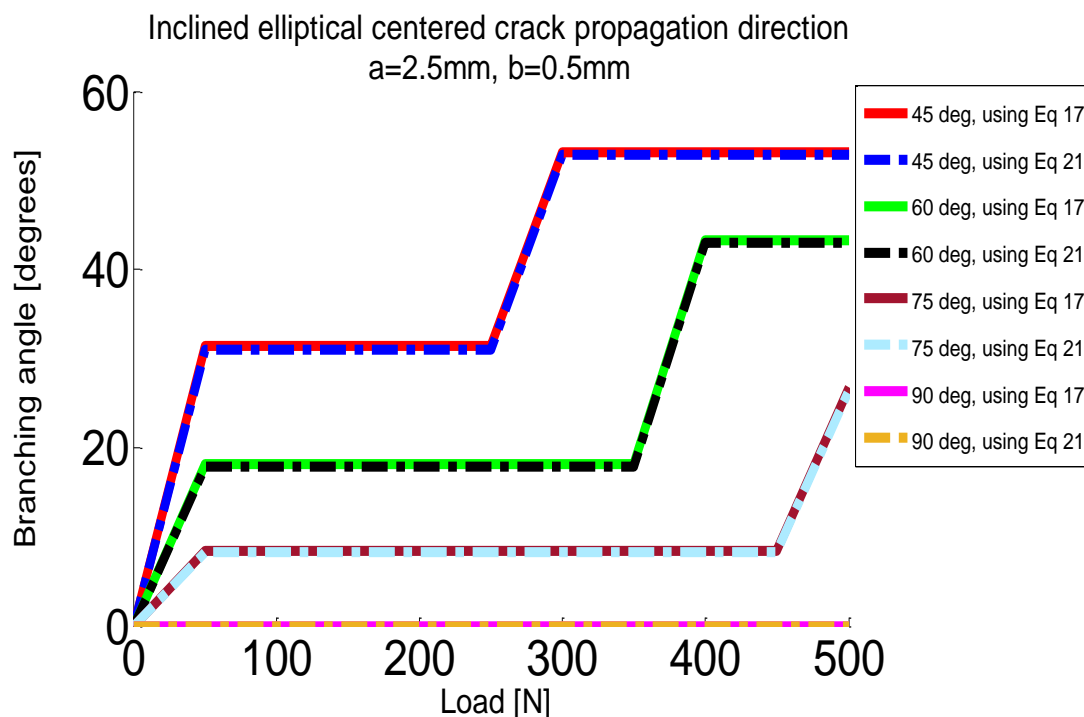
Figure A.D1 12: Influence of changing inclination angle (β) on growth path for defect type 'c'.

```
%%%% Analytical solution of SIF %%%%

a=(1E-3);%length of crack
b=0.5e-3;%opening of crack
c=sqrt(a^(2) - b^(2));%elliptical eccentricity
v=0.33;%Poission ratio
Eo=209E9;%Youngs modulus
E= Eo/(1-(v^2));%Plane stress assumption of Youngs modulus

%Analytical solution for pure KI
A=(2*a^(4))/(c*(a-b)^(2));
B=((c*b)/(a^(2)))+atan(c/b);
C=((4*a^(2)*b*c)/((a-b)^(3)))*atan(c/b);
D=(b*c^(4))/(a-b)^(4);
Gnom3 = Gnom_v_inf;
J_I= (( Gnom3.^(2))./E ).*(A*B-C+D);
KI22=sqrt(abs(J_I).*E); % pure mode I SIF

%Analytical solution of pure KII
EE=(4*c^(5))/((a-b)^(4));
F=(0.5*atan(c/b))-((b*c)/(2*a^(2)));
G=(4*b^(2)*c^(3))/(a-b)^(4);
H=(c/b);
I=((3/2)*atan(c/b));
J=(b*c)/(2*a^(2));
Gnom2 = Gnom3;
J_II=(( Gnom2.^(2))./E ).*(EE*F-G*(H-I+J));
KII22=sqrt(abs(J_II).*E); % pure mode II SIF

Keqv=(1E-6).*sqrt( dk1.^(2) + ( (614/507).*(dk2.^(3.21))
)).^(2/3.74));%Equivalent SIF
```

```matlab
Figure(28),hold on;plot(linspace(0,500,max((size(Keqv)))),Keqv,'y');
title('Equivalent SIF')
xlabel('Load [N]')
ylabel('Keqv [MPa]')
hold off


 for v=1:max(length(Keqv))
R(v)=min(Keqv(1:v))/max(Keqv(1:v));
R(isnan(R))=0;
Del_K_th3(v)=max(Keqv(1:v))*((1-R(v)) ^(0.33)) ;%Threshold SIF range
end


da3=0.000507.*(Keqv.^(3.74) - ((1E-6).*Del_K_th3).^(3.74));%Crack growth
rate
Figure(25),hold on;plot(linspace(0,500,max((size(da3)))),da3,'y');
title('Crack growth rate')
xlabel('Load [N]')
ylabel('da/dN [nm/cycle]')
hold off


Theta22=2.*atan((-2.*KII22)./(KI22+sqrt( KI22.^(2) + 8.*(KII22.^(2))
)));%Crack growth path
Figure(26),hold
on;plot(linspace(0,80,max((size(Theta22)))),abs(rad2deg(rad2deg(Theta22))),
'y');
title('Propagation direction')
xlabel('Load [N]')
ylabel('Branching angle [degrees]')
hold off


af=5e-3;%initial crack length
ai=2.5e-3;%Final crack length
A=4.47*10^-9;%material constant
m=3.3;%Paris constant
Ps=Keqv;%Stress amplitude


Nf2=(1./(A*(1.2^m).*(Ps.^m).*pi^(m/2))).* ((af.^(1-(m/2)) - ai^(1-
(m/2)))/(1-(m/2)));%Number of cycles until failure

Figure(27),hold on;plot(linspace(0,500,max((size(Keqv)))),abs(Nf2),'b');
title('Number of loading cycles to fracture')
xlabel('Load [N]')
ylabel('Nf [cycles]')
```

Appendix D2: Mixed Mode-I and Mode-II SIF formulations

```matlab
a=(1E-3);%length of crack
b=0.5e-3;%opening of crack
c=sqrt(a^(2) - b^(2));%elliptical eccentricity
v=0.33;%Poission ratio
Eo=209E9;%Youngs modulus
E= Eo/(1-(v^2));%Plane stress assumption of Youngs modulus
% %Mode I contribution in mixed mode loading


W_I= (b/a)*( (21/(5*pi))-1 )+1;        %correction functions of ellipse under
                                       %tensile stress
```

```matlab
L_1=Gnom.*sind(Bet)^2;
A=(L_1.^(2)./E);
B=(2*a^(4))/(c*(a-b)^(2));
C=((c*b)/a^(2) )+ atan(c/b);
D=(a^(2)*b*c)/(a-b)^(3);
F=(b*(c^(4)))/(a-b)^(4);
J_I=A.*( B*C - (4*D*atan(c/b)) + F); % Analytical solution of mode-I J
                                     %integral
KI22=sqrt((J_I./W_I).*E); %Mixed mode-I SIF


% %Mode II contribution in mixed mode loading

W_II= (b/a)*( (128/(15*pi))-1 )+1;  %correction functions of ellipse under
                                    %tensile stress
L_2=(-sind(Bet)*cosd(Bet)).*Gnom;
G1=(4*c^(5))/((a-b)^(4));
H=(0.5*atan(c/b))-((b*c)/(2*a^(2)));
I=(4*b^(2)*c^(3))/(a-b)^(4);
J=(c/b);
K=((3/2)*atan(c/b));
L=(b*c)/(2*a^(2));
G=((L_2.^2)./E);
J_II=G.*(G1*H-I*(J-K+L)); % Analytical solution of mode-II J
                          %integral


KII22=sqrt((J_II./W_II).*E); %mixed mode-II SIF

Keqv=(1E-6).*sqrt( dk1.^(2) + ( (614/507).*(dk2.^(3.21)) 
).^(2/3.74));%Equivalent SIF
Figure(28),hold on;plot(linspace(0,500,max((size(Keqv)))),Keqv,'y');
title('Equivalent SIF')
xlabel('Load [N]')
ylabel('Keqv [MPa]')
hold off

 for v=1:max(length(Keqv))
R(v)=min(Keqv(1:v))/max(Keqv(1:v));
R(isnan(R))=0;
Del_K_th3(v)=max(Keqv(1:v))*((1-R(v)) ^(0.33)) ;%Threshold SIF range
end

da3=0.000507.*(Keqv.^(3.74) - ((1E-6).*Del_K_th3).^(3.74));%Crack growth
rate
Figure(25),hold on;plot(linspace(0,500,max((size(da3)))),da3,'y');
title('Crack growth rate')
xlabel('Load [N]')
ylabel('da/dN [nm/cycle]')
hold off

Theta22=2.*atan((-2.*KII22)./(KI22+sqrt( KI22.^(2) + 8.*(KII22.^(2)) 
)));%Crack growth path
Figure(26),hold 
on;plot(linspace(0,80,max((size(Theta22)))),abs(rad2deg(rad2deg(Theta22))),
'y');
title('Propagation direction')
xlabel('Load [N]')
ylabel('Branching angle [degrees]')
hold off
```

```
af=5e-3;%initial crack length
ai=2.5e-3;%Final crack length
A=4.47*10^-9;%material constant
m=3.3;%Paris constant
Ps=Keqv;%Stress amplitude

Nf2=(1./(A*(1.2^m).*(Ps.^m).*pi^(m/2))).* ((af.^(1-(m/2)) - ai^(1-
(m/2)))/(1-(m/2)));%Number of cycles until failure

Figure(27),hold on;plot(linspace(0,500,max((size(Keqv)))),abs(Nf2),'b');
title('Number of loading cycles to fracture')
xlabel('Load [N]')
ylabel('Nf [cycles]')
```

Appendix D3: Biaxial mode-I and mode-II SIF formulations

A case study of equi-tension ($\aleph = 1$) and equi-tension-compression ($\aleph = -1$) bi-axial loading of defect type 'c' is presented in this section of the thesis. The load settings in this section of the investigation is modified from that of section 4.4.1 to include a 500N boundary load as depicted in Figure A.D1.13a and A.D1.13b respectively.

<div align="center">(a)   (b)</div>
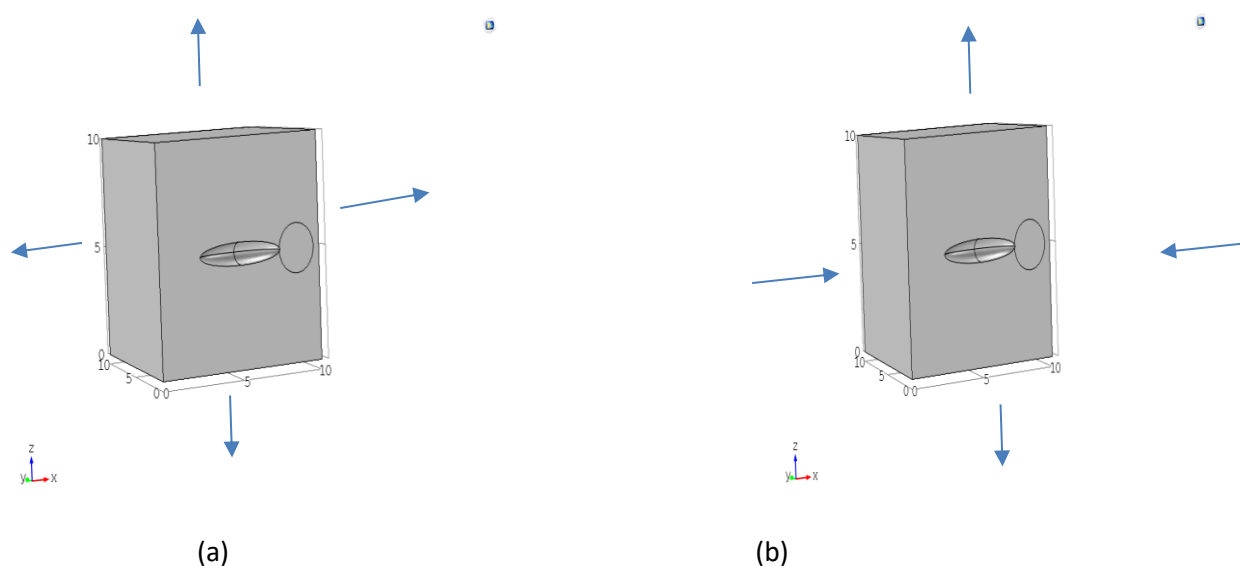
Figure A.D1 13: Bi-axial (500N) equi-tension ($\aleph$=1) and equi-tension-compression ($\aleph$=-1) model.
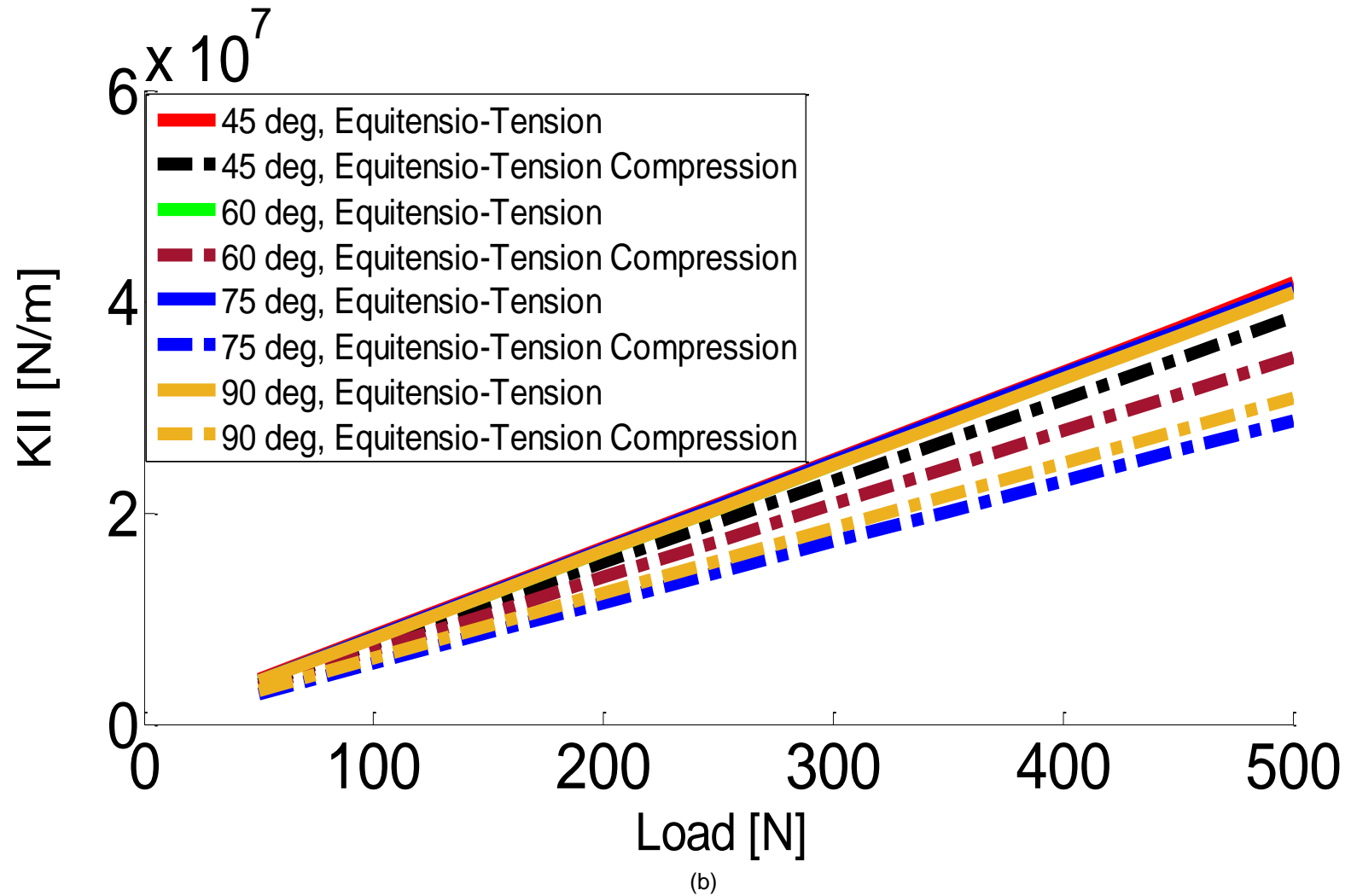
The simulated models are analysed for β ranging from 45^0 to 90^0 and the results in Figure A.D1.14 indicate that KI in the case of $\aleph = 1$ negligibly varies from a maximum value of 42.5MPa for all inclination angles investigated. While in the case of $\aleph = -1$ the peak value of KI ranges 40MPa to 30MPa for β equals 45^0-90^0. In Figure A.D1.14b, simulated KII in the case of $\aleph = 1$ negligibly varies from a maximum value of 42MPa to 5MPa for all inclination angles

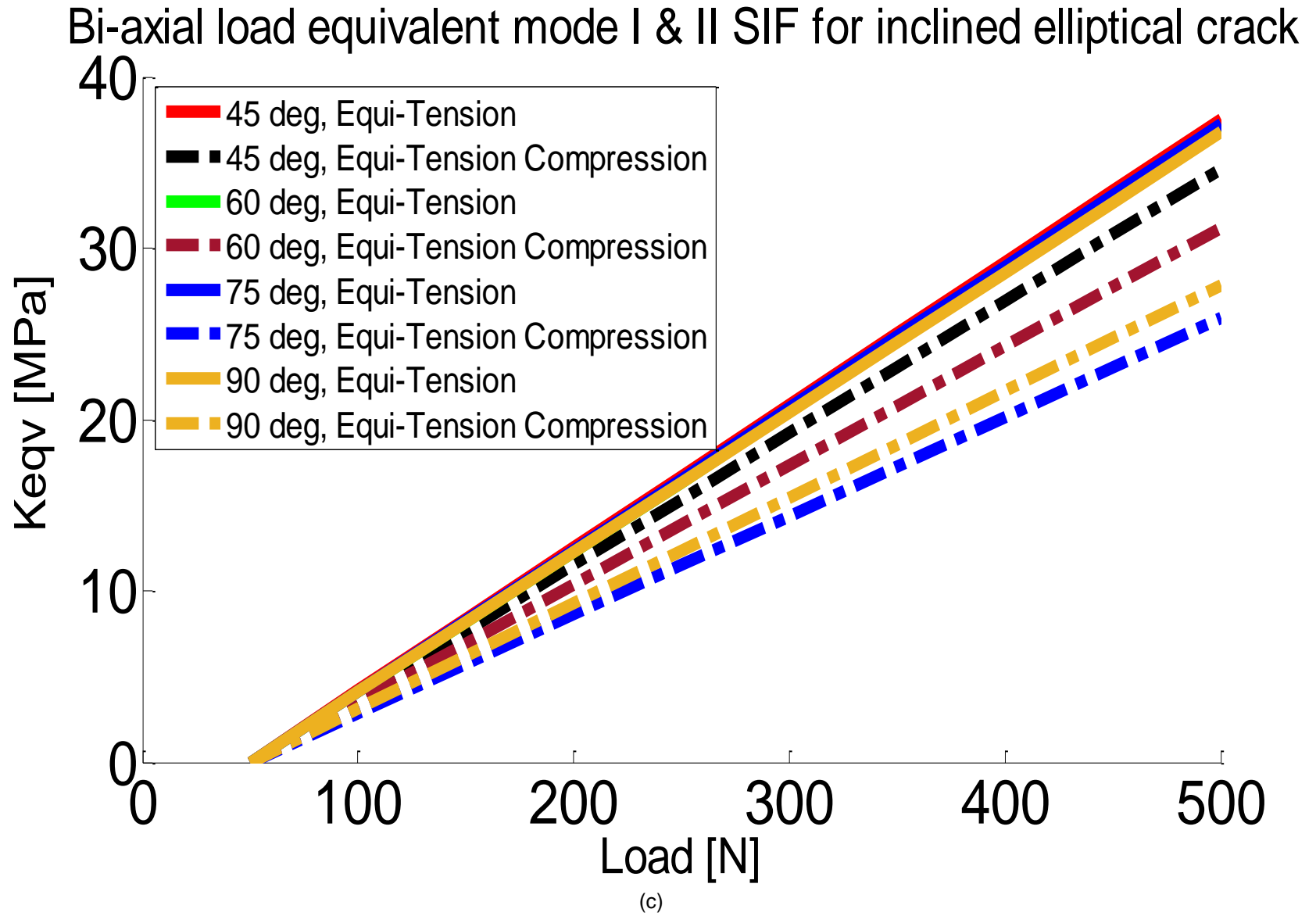investigated. While in the case of $\aleph = -1$ the peak value of KII ranges 40MPa to 30MPa for β equals 45^0-90^0. Keqv in Figure A.D1,14c is observed to negligibly vary from a maximum value of 40MPa. And for $\aleph = -1$ Keqv ranges from 37MPa to 27MPa for β equals 45^0-90^0. The crack growth rate in Figure A.D1.14d is in the case of $\aleph = 1$ negligibly varies from a maximum value of 400nm/cycle. While in the case of $\aleph = -1$ the growth rate ranges from 300nm/cycle to 100nm/cycle. The growth path in Figure A.D1.14e, confirm for $\aleph = 1$ that mixed mode loading corresponds to higher branching angles of 35^0, 32^0 for β equals 60^0,75^0. And in the case of pure fracture mode loading, low branching angles of 15^0, and 0^0 for β equals 45^0, and 90^0 is recorded. However in the case of $\aleph = -1$ simulated results confirm for that mixed mode loading corresponds to lower branching angles of, 18^0, 10^0 for β equals 60^0,75^0. And in the case of pure fracture mode-I fracture mechanism as expected a branching angle of 0^0 is recorded.
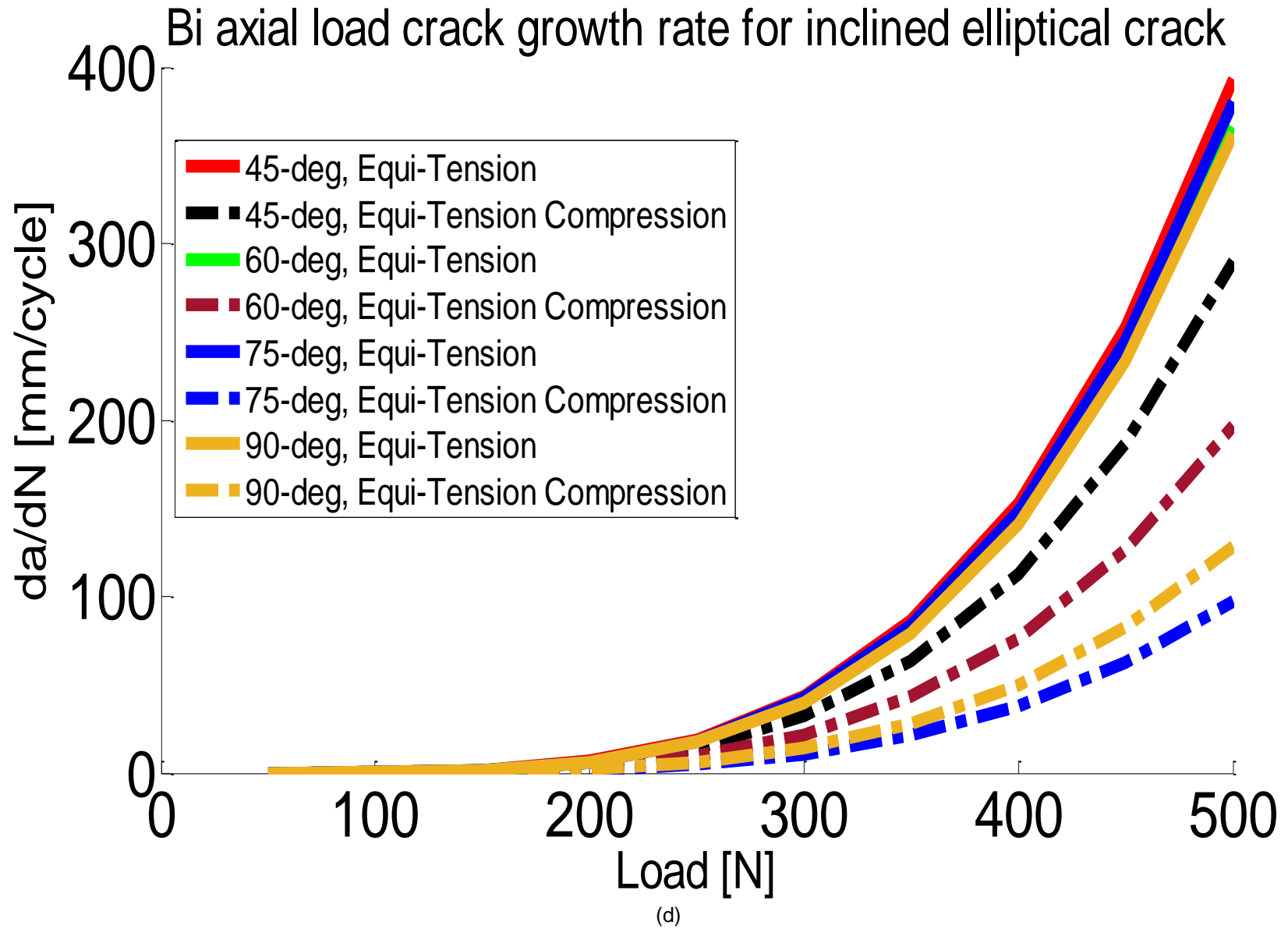
(a)
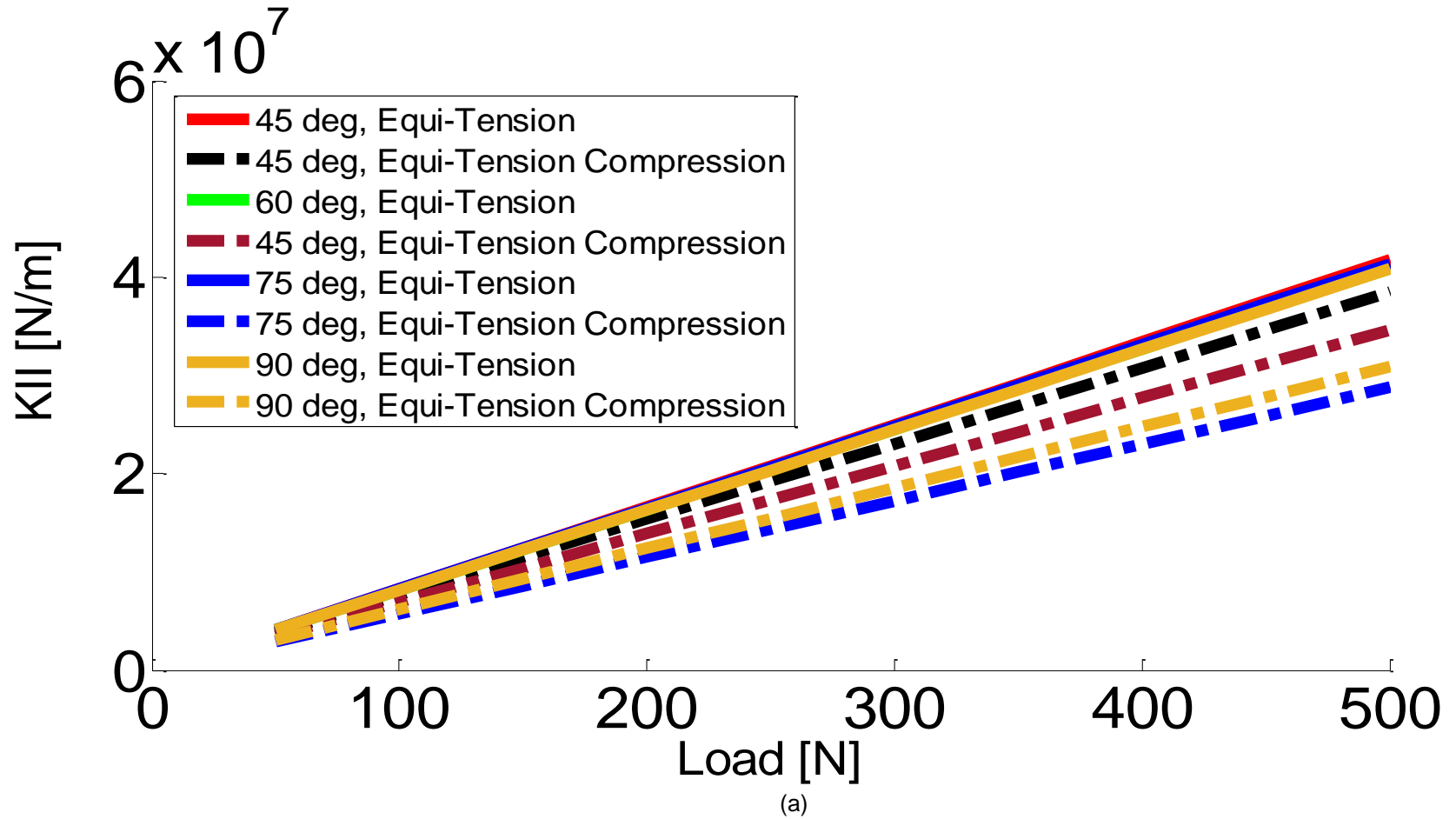
Bi-axial load mode-II SIF of inclined elliptical crack

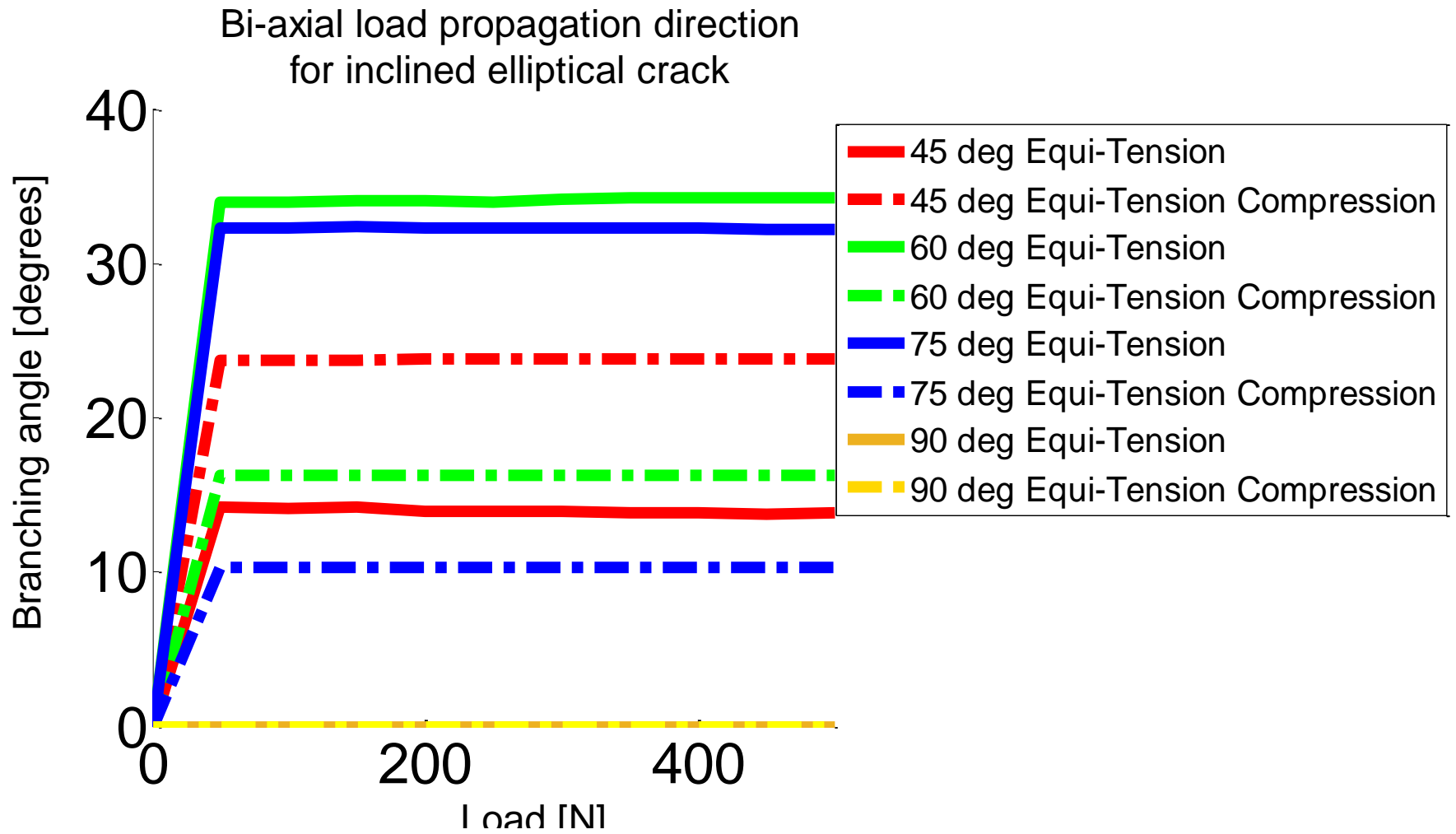

(b)

Bi-axial load equivalent mode I & II SIF for inclined elliptical crack

(c)

Bi axial load crack growth rate for inclined elliptical crack

(d)

(e)

```matlab
%%%%%%% Biaxial loading of elliptical inclined crack %%%%%%%
    b=0.5e-3;% crack opening
    ang=0;
    c1=(5e-3)/2;%length of crack
    a=c1;
    l=1/sqrt( (cosd(ang)/c1)^(2) + (sind(ang)/a)^(2) );%depth of crack
    n=1;% material hardness parameter
    nn=(Gnom_l.^(n))./(Gnom_v.^(n));% biaxiality of load
    Bet=45;%crack inclination angle

SI_inf=(Gnom_v.^(n)).*( (cosd(Bet)^(2)) + (nn.*(sind(Bet)^(2))) );%mode-I
Sstress component
SII_inf=((Gnom_v).^(n)).*((1-nn)./2).*( sind(2*Bet)*cosd(ang) );%mode-II
Sstress component
% SIII_inf=(Gnom_v)^(n).*((1-nn)/4)*( sind(2*Bet)*sind(ang) );%mode-III
Sstress component


KI22=SI_inf.*sqrt(pi*l);%Mode-I SIF
KII22=SII_inf.*sqrt(pi*l);%Mode-II SIF
% KIII=SIII_inf.*sqrt(pi*l);%Mode-III SIF

Keqv=(1E-6).*sqrt( dk1.^(2) + ( (614/507).*(dk2.^(3.21)) )
).^(2/3.74));%Equivalent SIF
Figure(28),hold on;plot(linspace(0,500,max((size(Keqv)))),Keqv,'y');
title('Equivalent SIF')
xlabel('Load [N]')
ylabel('Keqv [MPa]')
hold off

 for v=1:max(length(Keqv))
R(v)=min(Keqv(1:v))/max(Keqv(1:v));
R(isnan(R))=0;
Del_K_th3(v)=max(Keqv(1:v))*((1-R(v)) ^(0.33)) ;%Threshold SIF range
end

da3=0.000507.*(Keqv.^(3.74) - ((1E-6).*Del_K_th3).^(3.74));%Crack growth rate
Figure(25),hold on;plot(linspace(0,500,max((size(da3)))),da3,'y');
title('Crack growth rate')
xlabel('Load [N]')
ylabel('da/dN [nm/cycle]')
hold off

Theta22=2.*atan((-2.*KII22)./(KI22+sqrt( KI22.^(2) + 8.*(KII22.^(2))
)));%Crack growth path
Figure(26),hold
on;plot(linspace(0,80,max((size(Theta22)))),abs(rad2deg(rad2deg(Theta22))),'y
');
title('Propagation direction')
xlabel('Load [N]')
ylabel('Branching angle [degrees]')
hold off

af=5e-3;%initial crack length
ai=2.5e-3;%Final crack length
```

352

```
A=4.47*10^-9;%material constant
m=3.3;%Paris constant
Ps=Keqv;%Stress amplitude

Nf2=(1./(A*(1.2^m).*(Ps.^m).*pi^(m/2))).* ((af.^(1-(m/2)) - ai^(1-(m/2)))/(1-
(m/2)));%Number of cycles until failure

Figure(27),hold on;plot(linspace(0,500,max((size(Keqv)))),abs(Nf2),'b');
title('Number of loading cycles to fracture')
xlabel('Load [N]')
ylabel('Nf [cycles]')
```

Appendix E: Linking MATLAB and COMSOL

Appendix E1: Linking NDE to fracture mechanics in rails

In recent times, several attempts of linking detection with prediction tools have required a combination of experimental testing, followed by numerical analysis, and/or model based prediction methods in order to enhance the performance of structural integrity assessment and online conditioning monitoring not only in the rail industry but in other industries also. Several links between field based detection and model based prediction of rail infrastructure have been discussed in literature of which vibration, modal analysis and magnetic flux leakage systems have received the most attention not only within rail research but in other fields. However, this research focuses on remote, detection and prediction tools for rail damage, the technology is supported by data related to defect geometry, and contact dynamics. The MATLAB image processing toolbox (defect detection platform), in collaboration with the VAMPIRE contact calibration data (defined as a MATLAB look-up matrix) is linked to communicate directly (in a multi-directional manner) with global track model (COMSOL multi physics FE software) by means of the LiveLink™ for MATLAB® add on. As detailed in the user guide hand book of COMSOL Multiphysics, the following command line when executed from MATLAB command line, automatically creates a replica of FE model (referred to as 'Model2') as a variable (called 'model') in MATLAB. The

function also creates a directory for modifying FE model parameters from MATLAB window and vice-versa. The interaction between the linked data is best described by Figure A.E1.1below.
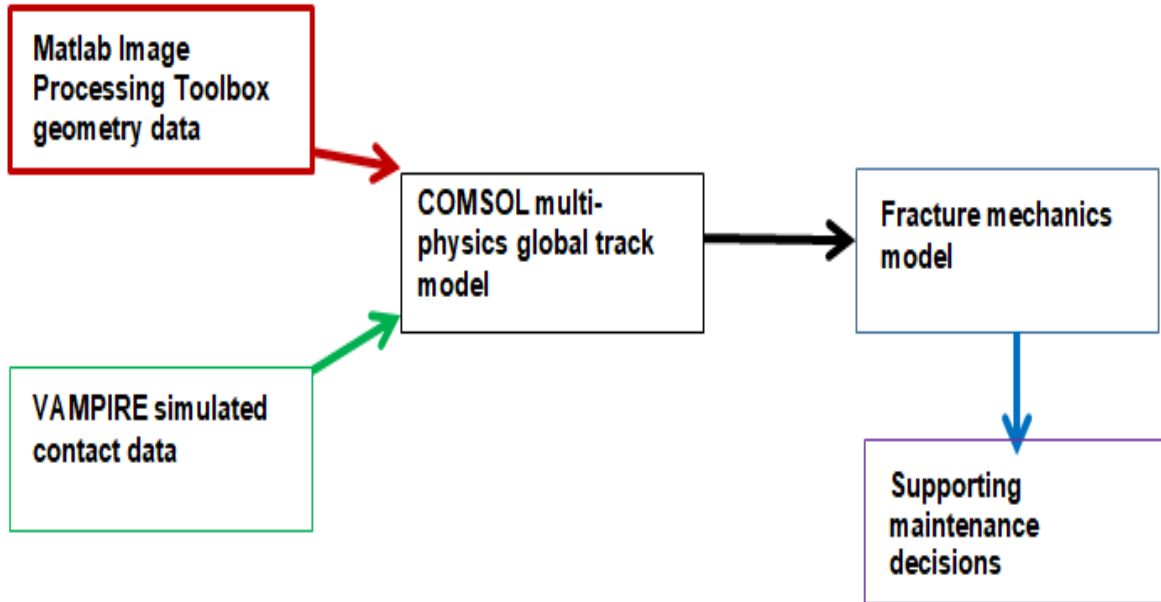

Figure A.E1. 1: Shows the interaction between VAMPIRE contact data, Image processing data, and Fracture mechanics prediction.

To start the linking process between NDE (MATLAB image processing toolbox) and fracture prediction model (COMSOL multi-physics FE software) the icon in Figure A.E1.2 below is selected.

Figure A.E1. 2: Shows the icon for initiating COMSOL with MATLAB.

After successful launching of the linking icon above, the window depicted in Figure A.E1.3 below

appears with the specified listening port for establishing the link (also referred to as the client server) between COMSOL and MATLAB.
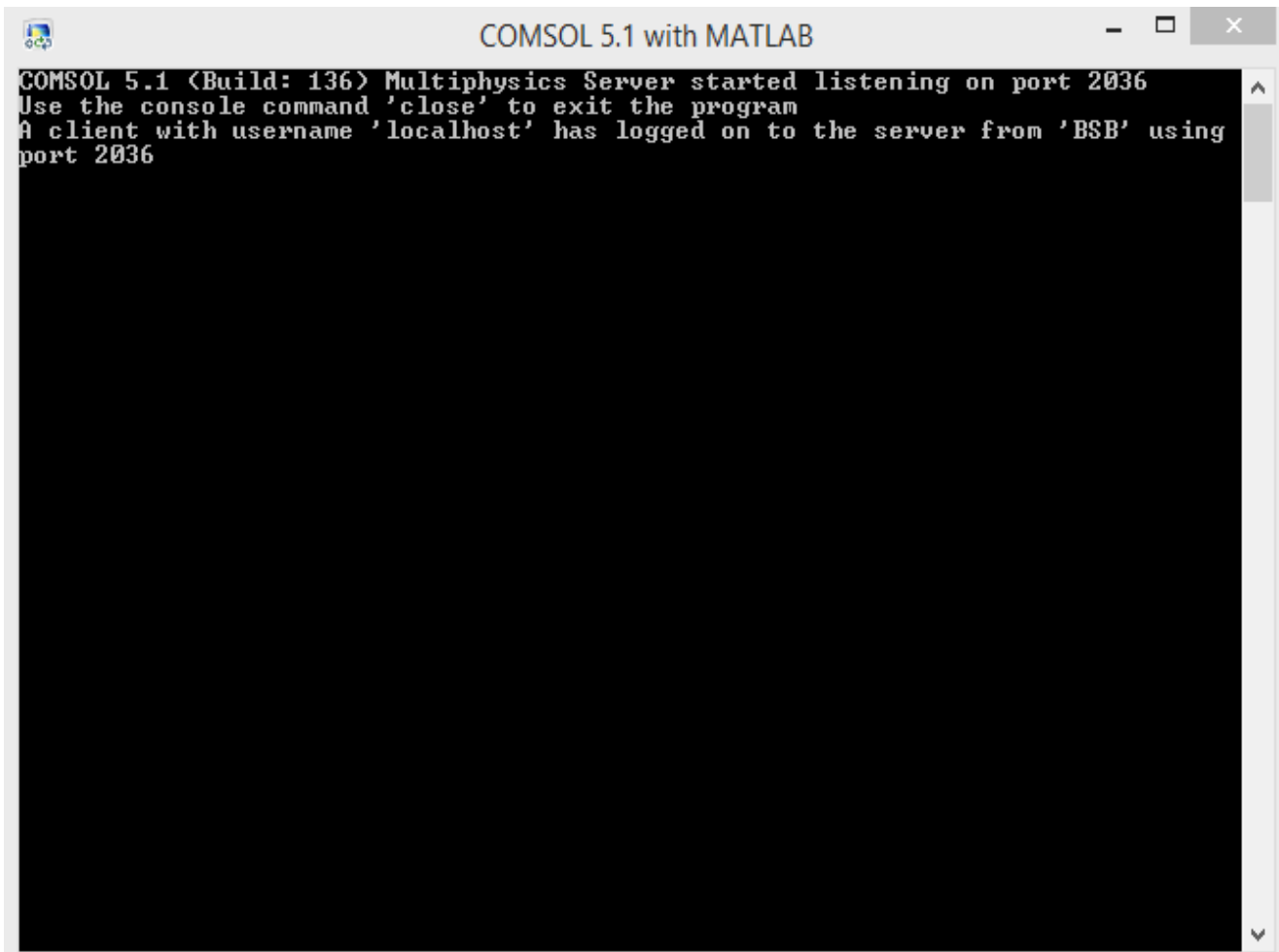

COMSOL 5.1 with MATLAB

```
COMSOL 5.1 (Build: 136) Multiphysics Server started listening on port 2036
Use the console command 'close' to exit the program
A client with username 'localhost' has logged on to the server from 'BSB' using
port 2036
```

Figure A.E1. 3: Shows the client server window with details of listening port for establishing link between COMSOL and MATLAB.

The global track model is connected (via the client server) to MATLAB image processing toolbox by clicking File, on the tool bar of COMSOL Multi-physics, under Client Server, click Connect to Server as depicted in Figure A.E1.4 below.
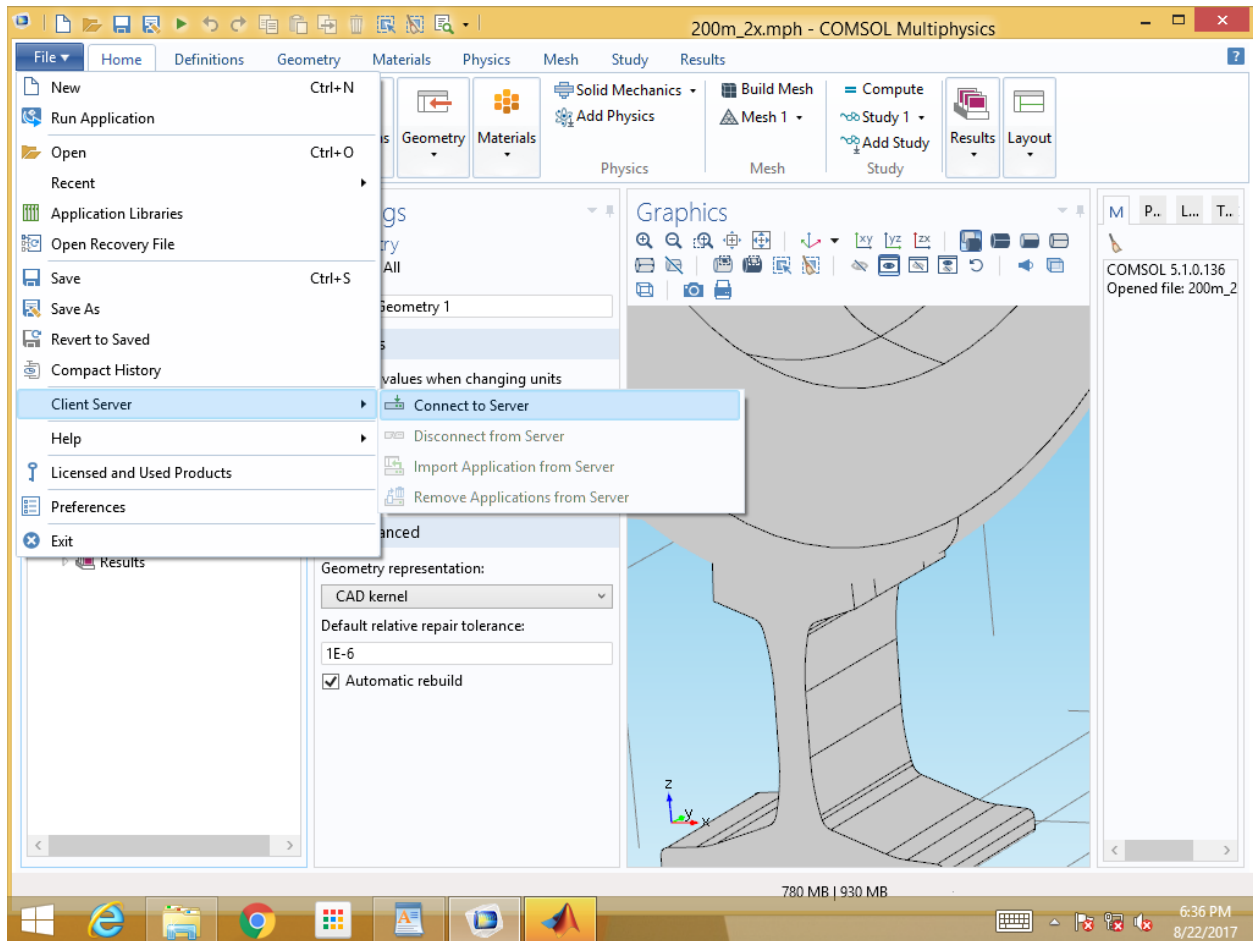
Figure A.E1. 4:Shows the steps for connecting to client server on the COMSOL Multi-physics platform.

This client server connection (referred to as localhost) is further secured by setting a user defined

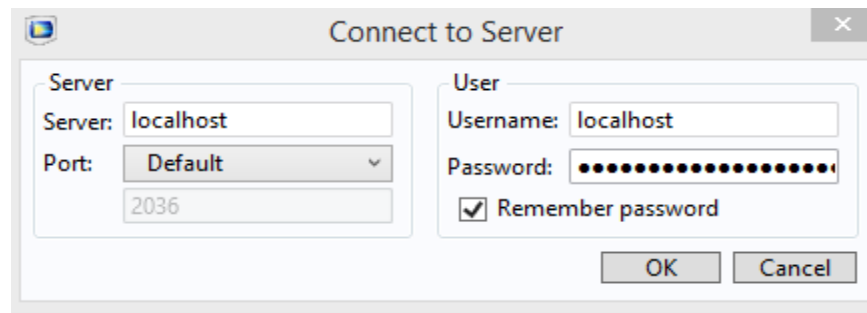username and password and clicking ok, as depicted in Figure A.E1.5 below.

Figure A.E1. 5: Shows the security settings for server connection using COMSOL Multi-physics.

Successful completion of the step in the figure above, would lead to opening a MATLAB window (automatically). It is also noticeable with a variable created in the workspace which has a directory of COMSOL folder in PC program file as shown in Figure A.E1.6 below.
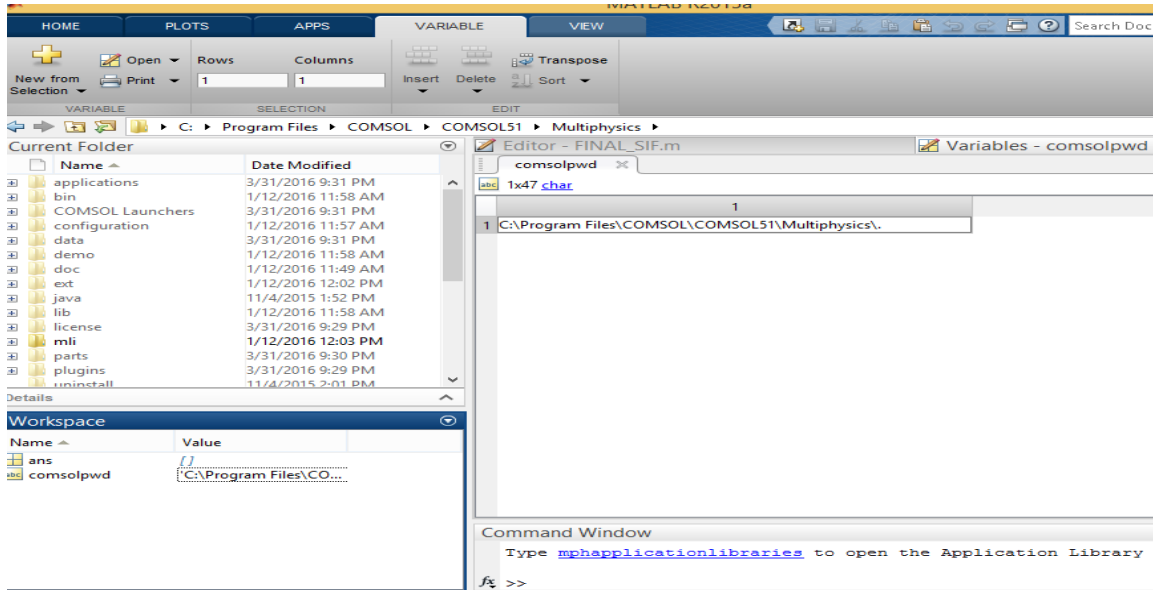

Figure A.E1. 6: Shows the MATLAB explorer program to be adopted for communicating with COMSOL.

On the MATLAB command window of the Figure above, the following syntax is used to export the global track model as a variable in MATLAB workspace. This synchronisation enables the multi-channel communication with image processing toolbox and VAMPIRE contact data (all MATLAB variables) with the global track model.

model=ModelUtil.model('Model2'); %Establish a link between parameters of the Comsol FE
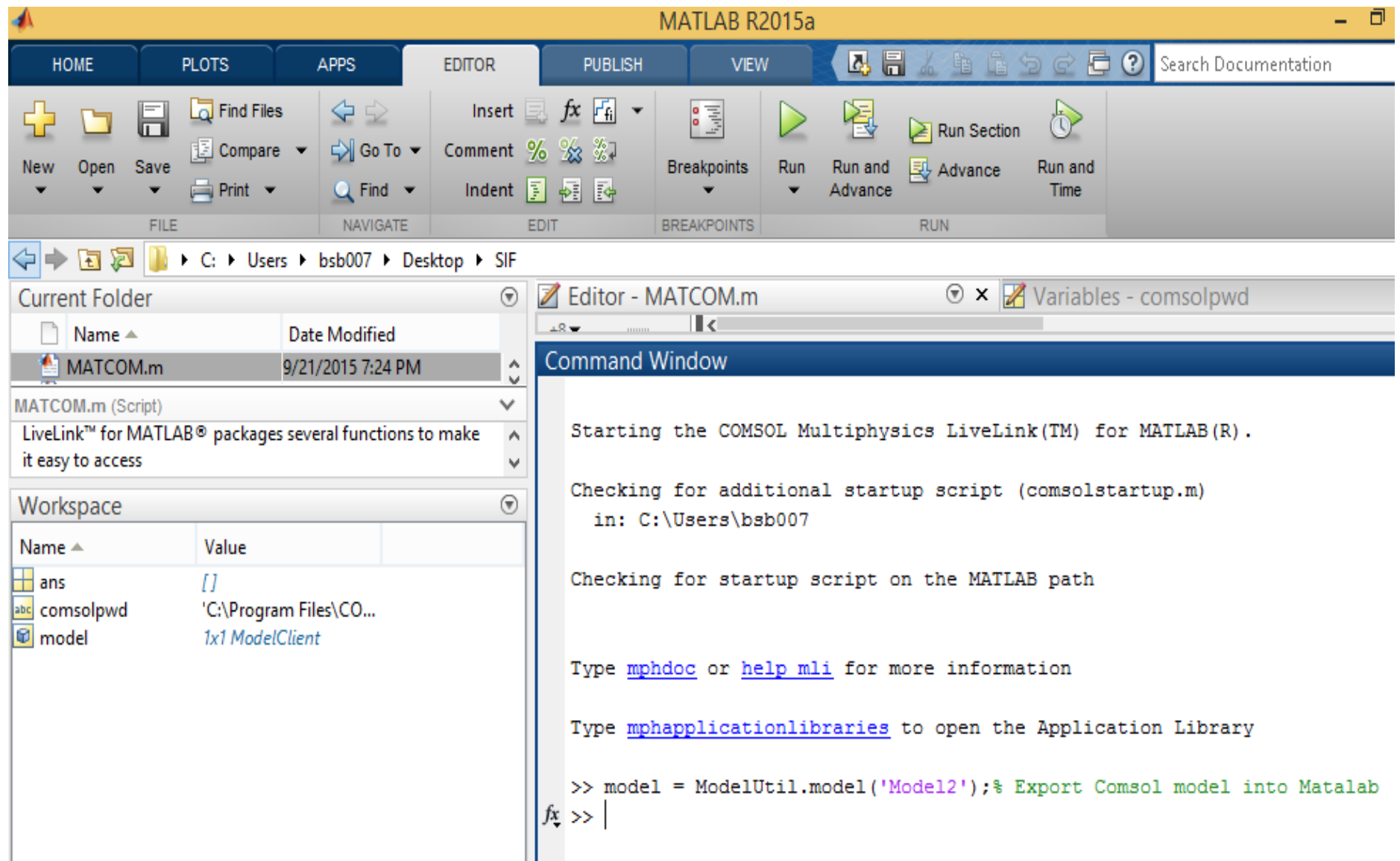%model to variables in MATLAB workspace.

Incorporating automated rail RCF damage detection algorithms with crack growth modelling

Figure A.E1. 7: Shows the exported global track model as a variable in MATLAB workspace.

It can be seen from Figure A.E1.7 above that the global track model referred to as ('Model2') is automatically exported to the MATLAB workspace as 'model'. This enables the automatic calibration of contact and crack settings within the 'Geometry node' of the global track model. The calibration data and defect geometry is concatenated into a loop up table referred to as MATCOM_LUT. The MATCOM_LUT function returns the following data as observed in the MATLAB workspace window depicted in Figure A.E.18 below.
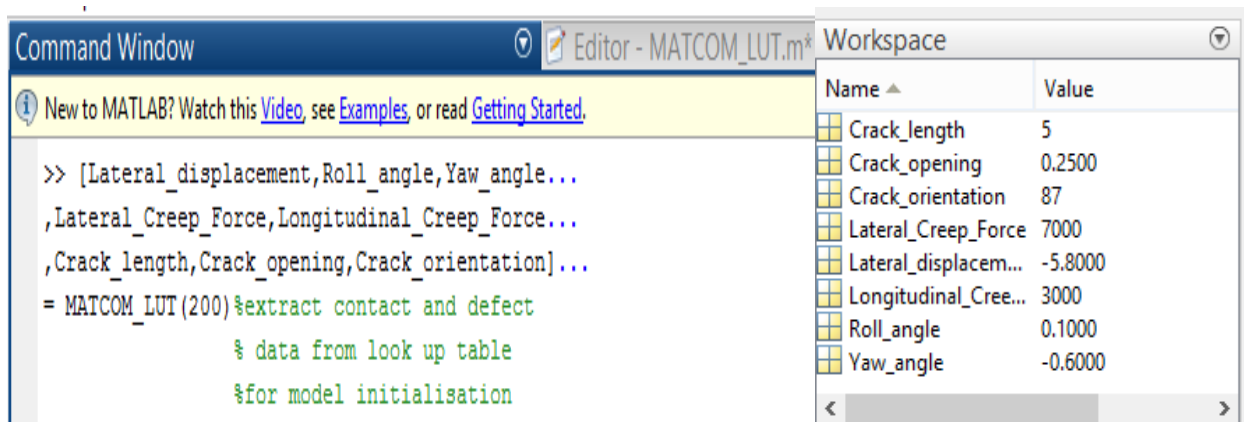
Figure A.E1.8: Shows the syntax for extracting contact data related to 200m curve radius and details of a specified defect geometry.

All the above parameters (with the exception of the boundary coordinates) are set within the global track model using parameter initialisation syntax as described below.

model.param.set('a',Lo);%Change parameter in global track model
%from MATLAB workspace

The above syntax is used as shown in Figure A.E1.9, to change the global track model parameter 'a' to equal the value of MATLAB variable 'Lo'. This approach can be used to export the lateral displacement, roll angle, yaw angle, axle load, maximum surface length, crack orientation, elliptical semi-axis length (major and minor) parameters of the proposed technology.

```
Command Window

>> %Automatic parameter set in global track model
model.param.set('lat_d',Lateral_displacement);%set Global track model parameter
model.param.set('R_a',Roll_angle);%set Global track model parameter
model.param.set('Y_a',Yaw_angle);%set Global track model parameter
model.param.set('lat_cf',Lateral_Creep_Force);%set Global track model parameter
model.param.set('long_cf',Longitudinal_Creep_Force);%set Global track model parameter


model.param.set('c_l',Crack_length);%set Global track model parameter
model.param.set('c_op',Crack_opening);%set Global track model parameter
model.param.set('c_l',Crack_orientation);%set Global track model parameter
model.param.set('c_offset_x',c_x);%set Global track model parameter
model.param.set('c_offset_y',c_x);%set Global track model parameter
fx >>
```

Figure A.E1. 9: Shows the automatic setting of global track model parameters from MATLAB command window.

The result of Figure A.E1.9 above is observed to be the initialisation elliptical crack geometry as depicted in Figure A.E1.10 and also the contact settings of global track model as depicted in Figure A.E1.11 below.
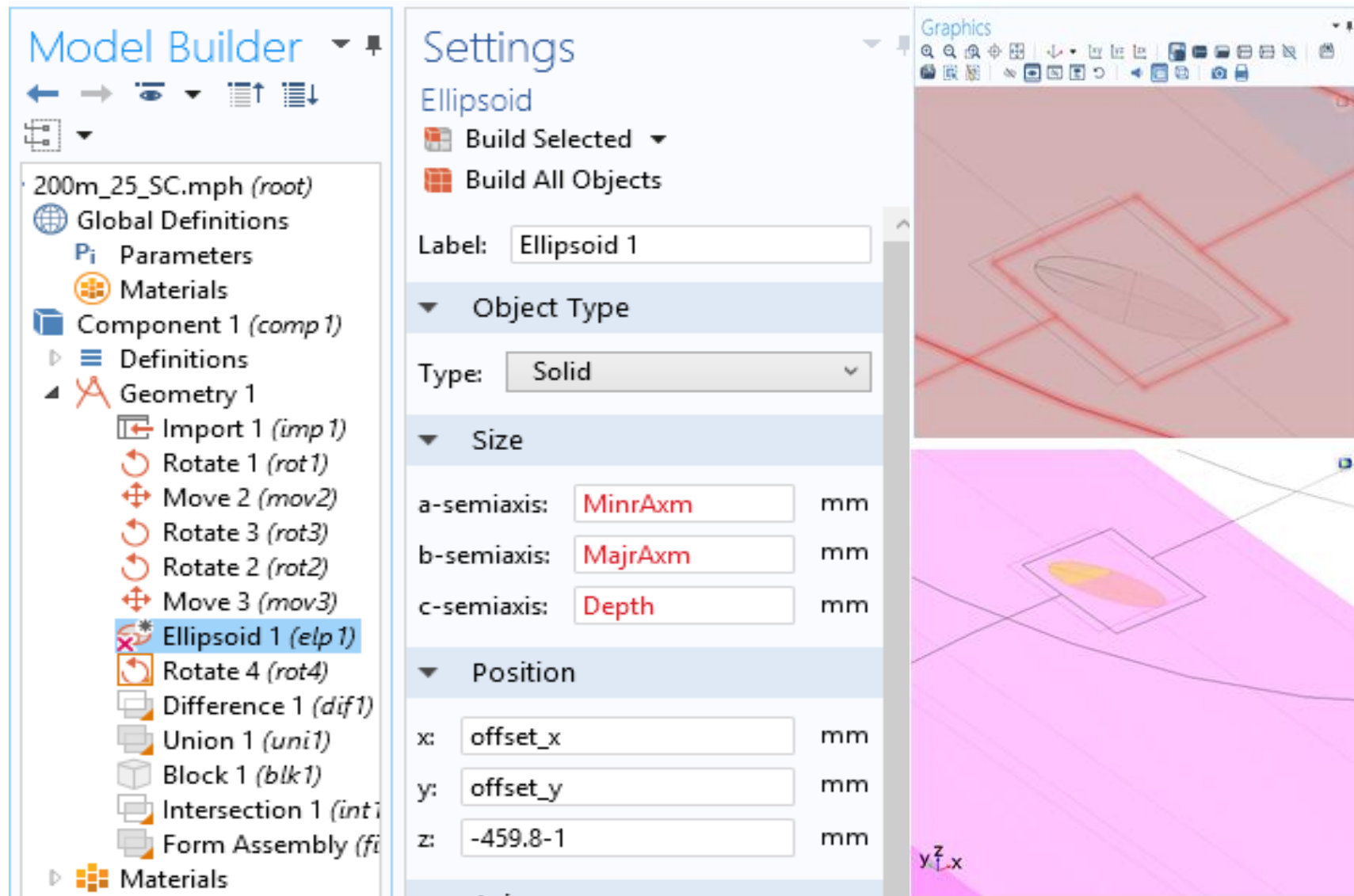
Figure A.E1. 10: Show elliptical crack geometry settings in COMSOL software.

Figure A.E1. 81: Shows the parameters related to contact data and defect geometry initialised from MATLAB command window.

In the case of utilising the actual crack boundary in the geometry node of the global track model, the interpolation curve node is selected and the text file containing the boundary coordinates of the defect under investigation is specified in the file directory as depicted in Figure A.E1.12 below.
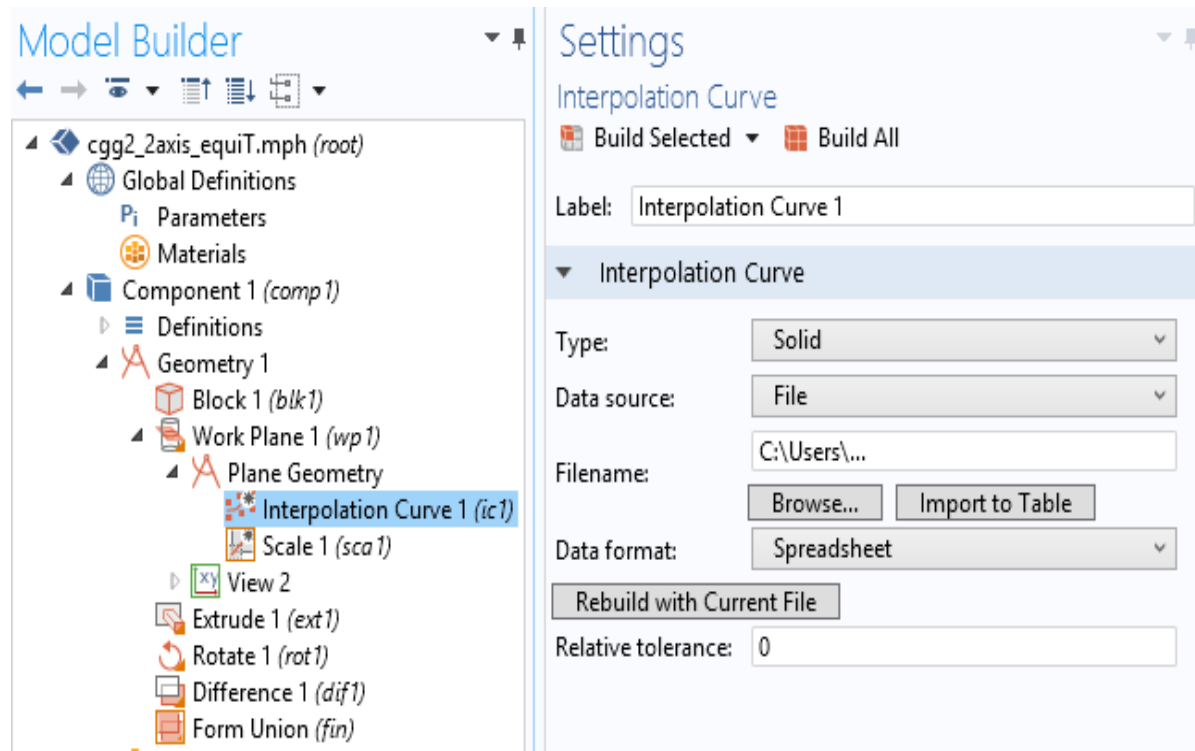
Figure A.E1. 92: Shows the importation of the actual crack boundary coordinates.

The crack geometry is moved to required offset location and further scaled to recalibrate the defect geometry from pixel scale unit to mm dimensions as shown in figure A.E1.13 below.
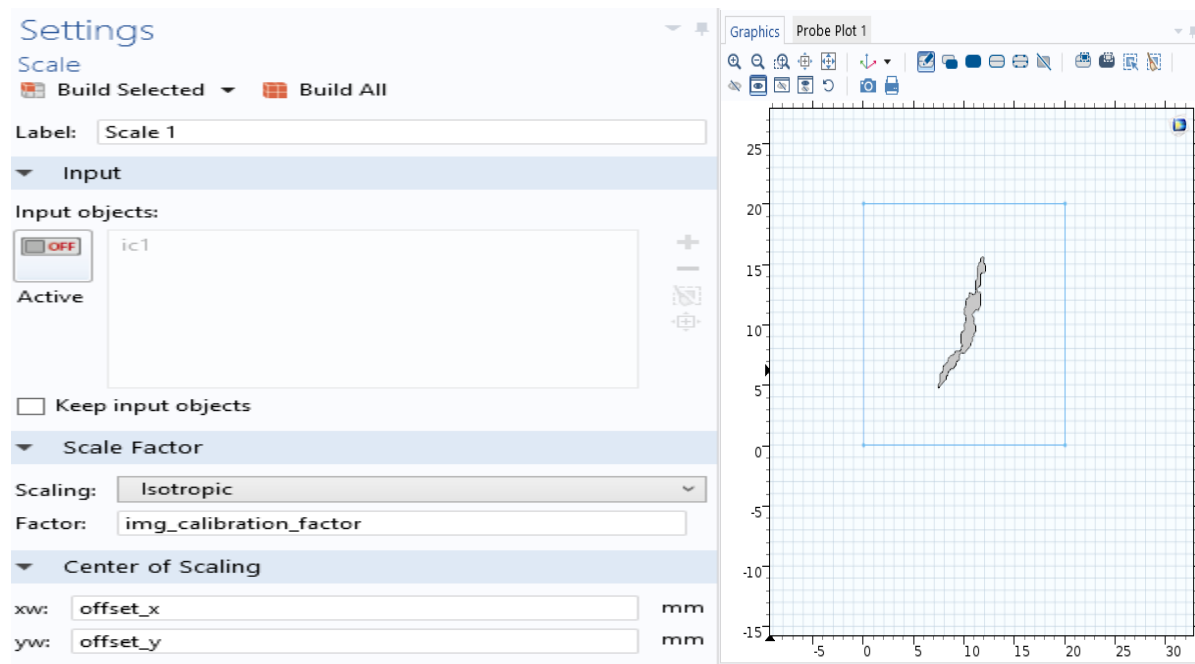

Figure A.E1. 103: Shows the settings for defect offset and scaling to mm dimension.

An extrude function is used to convert the plane representation of the defect boundary into a solid 3D crack geometry within the global track model as shown in Figure A.E1.14 below.
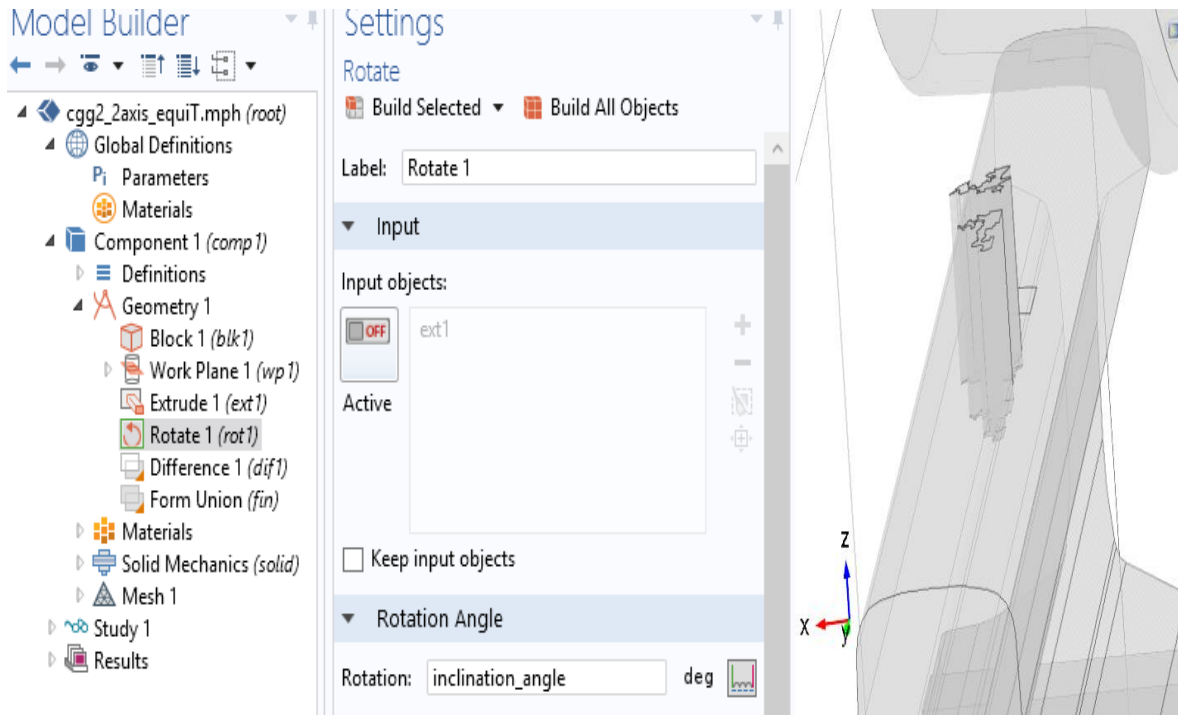

Figure A.E1. 11: Shows the defect extrusion from plane geometry to 3d solid structure.

A difference node is adopted to inscribe the 3D representation of the defect boundary coordinates within the global track model as shown in Figure A.E1.15 below.
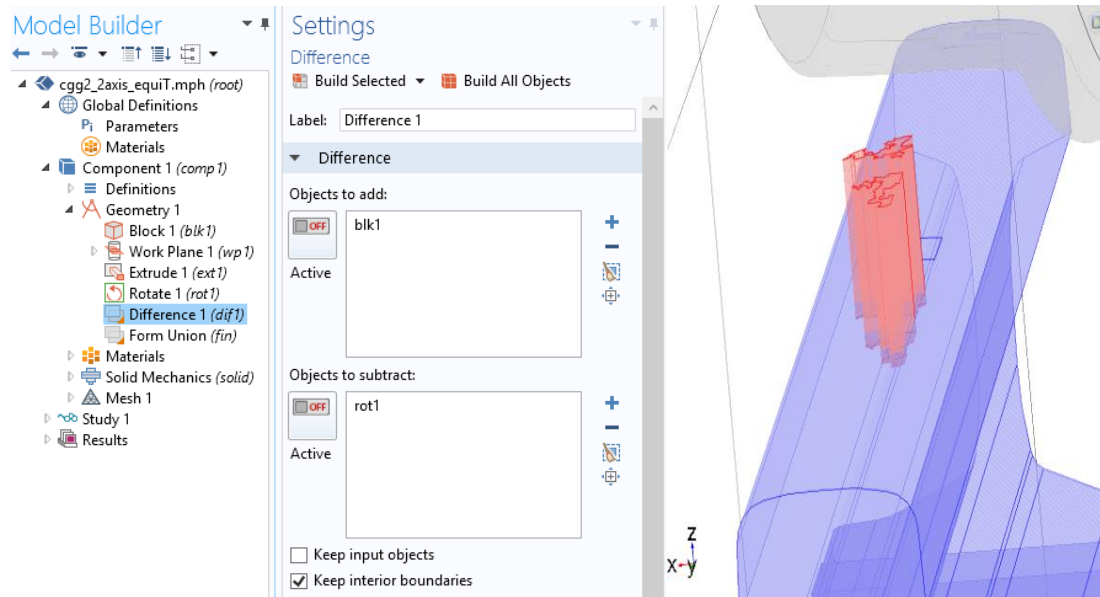
Figure A.E1. 125: Shows the incision for the defect boundary within the global track model using difference function.

Upon complete initialisation of the global track model as depicted in Figure A.E1.16 below, it is envisaged that the actual condition of rail can be analysed automatically.
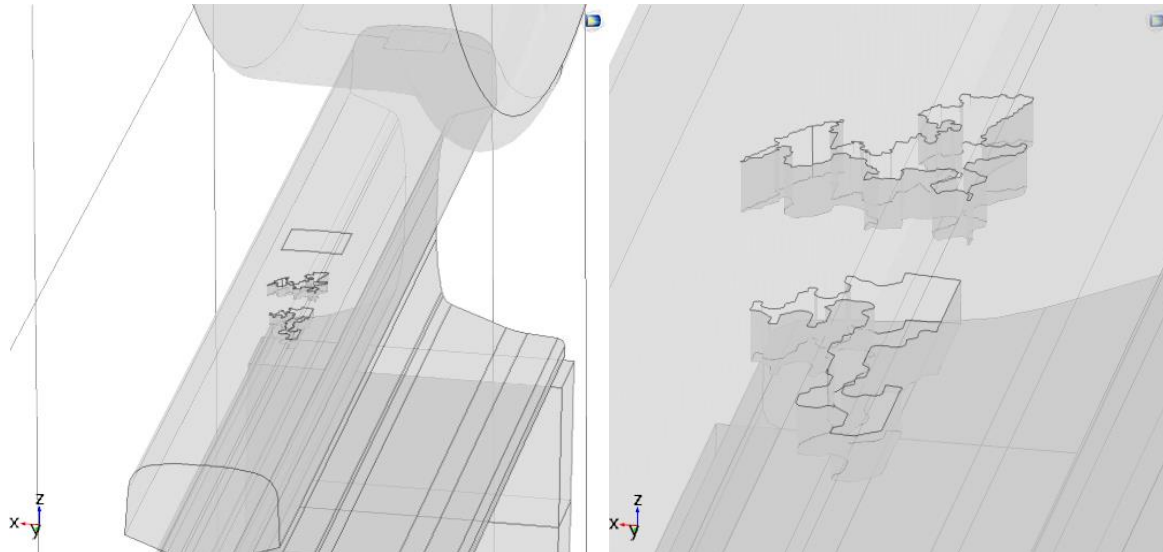
Figure A.E1. 13: Shows the global track model containing defects whose fracture mechanics behaviour is to be predicted.

The model is set to run and the required solid mechanics expressions (COMSOL simulated contact stresses) are extracted for further processing. According to the mathematical methods elaborated in Chapter 4, the solid mechanics expression that is most relevant for the prediction of crack behaviour is the nominal stress (represented in the solution of the global track model as

'solid.PGpzZ', 'solid.PGpyY', 'solid.PGpxX' for all principal directions). Other MATLAB functions that can be adopted for extracting simulated results are described below.

- mpheval, to evaluate solid mechanics expressions on all node points of a specified selection.

- mphinterp, to evaluate expressions at arbitrary location.

- mphglobal, to evaluate global expressions.

- mphint2, to integrate the value of expressions on selected domains.

- mphmax, to evaluate the maximum value of an expression.

- mphmin, to evaluate the minimum value of an expression.

- mphmean, to evaluate the average value of an expression.

For each of the above functions the following syntax can be used to extract the specified expression either at a specific coordinate, boundary, or domain.

```
SP = mphinterp(model,'expression','coord',[x;y;z]);%Extracting the value of
                                                   %an expression at a
                                                   %specified x,y,z
                                                   %coordinate
SP = mphinterp(model,'expression','volume',domain_number); %Extracting the
                                                   %value of
                                                   %an expression
                                                   %of a specified
                                                   %domain
SP = mphinterp(model,'expression','surface',boundary_number);%Extracting
                                                   %the value of an
                                                   %expression at a
                                                   %specified boundary
```

The algorithms of image processing and fracture mechanics have been extensively analysed in Chapters 3 and 4 respectively. As shown in Figure A.E1.17, it is envisaged that a rail image can be automatically incorporated in the global contact model taking into account the actual boundary coordinate of the crack. The benefits of the linking NDE and prediction aspects of rail maintenance

are yet to be explored in rail industry. By using this geometry, more realistic stress distribution is achieved which in turn offer more realistic prediction of rail fracture behaviour. Furthermore, a considerable improvement in flexibility of rail condition monitoring, is attained such that separate maintenance tasks (i.e. detection and prediction) combined within a single package. This technology is considered more intelligent compared to similar automatic visual inspection systems such as patent applications of Bar-Am, M., & Yaakov, Z. (2007) which are only capable of forwarding detected defects to maintenance engineer stations for further processing.

After successful implementation of an incorporated detection and prediction model, the LiveLink connection is disconnected from the COMSOL multi-physics tool bar by clicking File, Client Server, and Disconnect from Server as shown in the Figure A.E1.17 below.
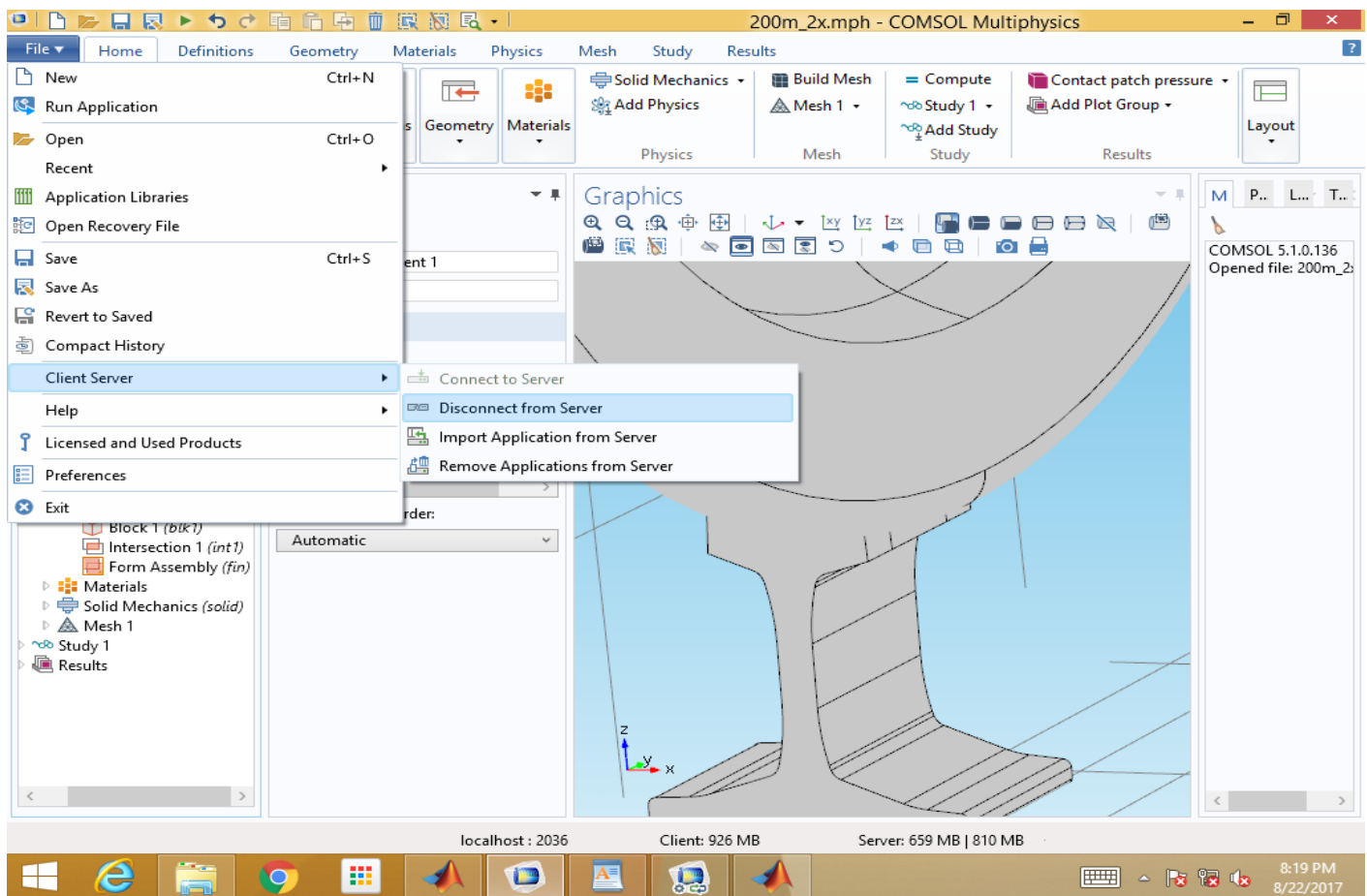

Figure A.E1. 147: Shows the procedure for disconnecting the COMSOL Multiphysics software from client server.

When successfully disconnected from the client server the confirmation is prompted as shown in
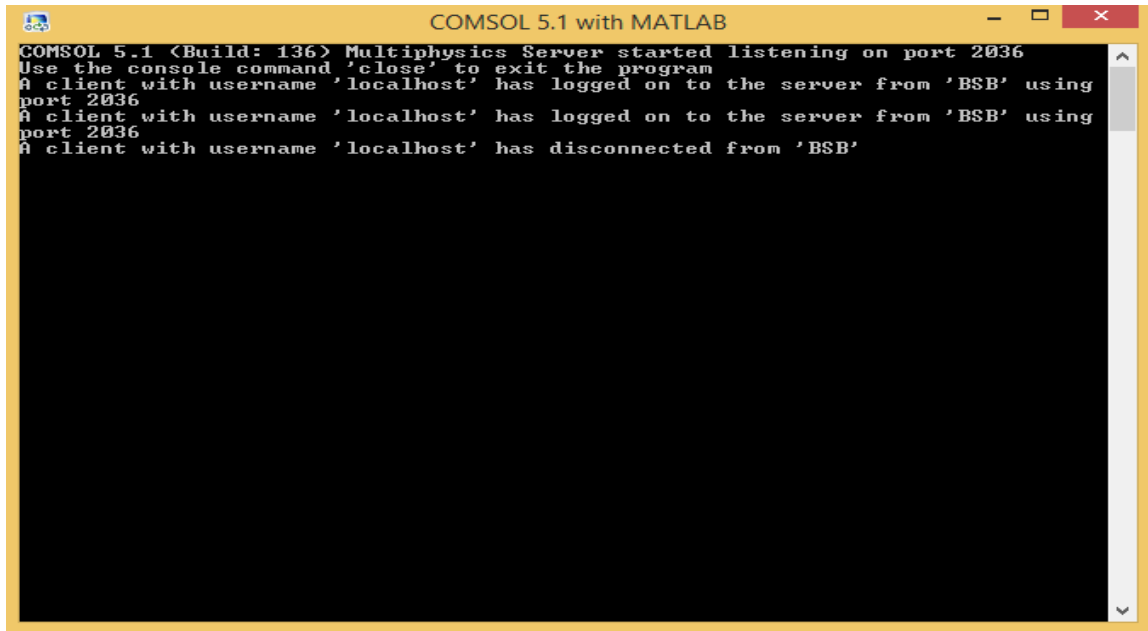
Figure A.E1.18 below.

Figure A.E1.18 : Shows the procedure for disconnecting the COMSOL Multi-physics software from client server.

Appendix E2: MATLAB function for establishing link between defect detection and

prediction

```matlab
%%%%%%%%%%% MATCOM_LUT %%%%%%%%%%%%%%%%
function [Lateral_displacement,Roll_angle,Yaw_angle...
,Lateral_Creep_Force,Longitudinal_Creep_Force...
,Crack_length,Crack_opening,Crack_orientation]...
= MATCOM_LUT(Curve_radius)%extract contact and defect
                          % data from look up Table
                          %for calibration purpose

Calibration_data=load('V.txt') %Load vampire contact file .txt format

C_R==Curve_radius;%record curve radius

for i=1:length(Calibration_data(1,:))%loop all curve radius values to find
                                     %specific choice
Lateral_displacement=Calibration_data(i==C_R,2);%select value of lateral
                                                %displacement at specified
                                                %curve radius
Roll_angle=Calibration_data(i==C_R,3);%select value of roll angle at
                                      %specified curve radius
Yaw_angle=Calibration_data(i==C_R,5);%select value of yaw angle at specified
```

```matlab
                                              %curve radius
Lateral_Creep_Force=Calibration_data(i==C_R,7);%select value of lateral creep
                                                %force at specified curve %
                                                %radius
Longitudinal_Creep_Force=Calibration_data(i==C_R,9);%select value of
                                                    %longitudinal creep
                                                    %force at specified
                                                    %curve radius

end % for loop



%Image defect statistical data
Crack_length=DL;%set crack length from image processing algorithm
Crack_opening=MnrAxm;%set crack opening from image processing algorithm
Crack_orientation=Om;%set crack orientation from image processing algorithm
c_x=Centriod.x;%set crack centre from image processing algorithm
c_y=Centriod.y;%set crack centre from image processing algorithm

%Automatic parameter set in global track model
model.param.set('lat_d',Lateral_displacement);%set Global track model
parameter
model.param.set('R_a',Roll_angle);%set Global track model parameter
model.param.set('Y_a',Yaw_angle);%set Global track model parameter
model.param.set('lat_cf',Lateral_Creep_Force);%set Global track model
parameter
model.param.set('long_cf',Longitudinal_Creep_Force);%set Global track model
parameter



model.param.set('c_l',Crack_length);%set Global track model parameter
model.param.set('c_op',Crack_opening);%set Global track model parameter
model.param.set('c_l',Crack_orientation);%set Global track model parameter
model.param.set('c_offset_x',c_x);%set Global track model parameter
model.param.set('c_offset_y',c_x);%set Global track model parameter
end % function
```