Development of Advanced Techniques For Gear Wear Monitoring and Prediction

Haichuan Chang

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Mechanical and Manufacturing Engineering

Faculty of Engineering

The University of New South Wales

January 2022

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Chang**

First name: **Haichuan** Other name/s:

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Mechanical and Manufacturing Engineering** Faculty: **Faculty of Engineering**

Title: Development of Advanced Techniques For Gear Wear Monitoring and Prediction

Abstract

Gears are widely used in industrial machinery, and gear failure is a main cause of machine failure. Since gear wear is often the initial stage of gear failure, its monitoring and prediction are key to minimising machine downtime, maintenance costs, and safety risks. However, existing gear wear monitoring and prediction techniques face some ongoing technical challenges, including providing direct wear information and wear assessment and prediction in a cost-effective and efficient manner. To tackle the challenges, this research aims to develop a set of advanced techniques for gear wear monitoring and prediction. The four objectives of the research and their corresponding methodologies and outcomes are summarised as follow.

(a) To develop a method to obtain direct and comprehensive wear information without disassembling the gearbox. This objective was realised by combining surface replication with image analysis, allowing easy acquisition of high-resolution mould images showing wear evolution on a tooth flank. (b) To investigate the relationship between the features of worn gear surfaces and those of wear debris. To further understand the role of wear debris analysis in wear assessment, a study on various features of macropits and wear particles in the same fatigue process was conducted and provided new insights into gear pitting and its monitoring. (c) To develop an automated system for gear wear assessment. Deep learning models were developed to identify wear mechanisms and severities using gear mould images and wear debris images. High classification accuracies were achieved, and comparisons between the two image sources were made. (d) To develop a gear wear prediction model using direct wear information. A deep generative model was developed and trained on time series of gear mould images. Tests showed that the model using the state-of-the-art AI technology can generate realistic and accurate predictions.

Overall, this research addressed the main limitations of existing methods and provided a direct and evidence-based tool for monitoring and predicting gear wear. Its specific contributions include a new moulding-imaging method for monitoring gear wear evolution, a detailed comparison between worn gear surfaces and wear debris in a wear process, and AI and image-based gear wear assessment and prediction models for the first time. The techniques could be performed during regular inspections of machines and used with online methods for increased robustness.

FOR OFFICE USE ONLY Date of completion of requirements for Award

Originality Statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Haichuan Chang
7 January, 2022

Copyright Statement

Haichuan Chang
7 January, 2022

Authenticity Statement

I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.

Haichuan Chang
7 January, 2022

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

☑ The candidate has declared that **some of the work described in their thesis has been published and has been documented in the relevant Chapters with acknowledgement**.

A short statement on where this work appears in the thesis and how this work is acknowledged within chapter/s:

> In my thesis, Chapter 4 in my thesis is a modified version of the published journal article below.
>
> [1] H. Chang, P. Borghesani, W. A. Smith, and Z. Peng, "Application of surface replication combined with image analysis to investigate wear evolution on gear teeth – A case study," Wear, vol. 430–431, pp. 355–368, 7 2019.
>
> Chapter 5 is a modified version of the published journal article below.
>
> [2] H. Chang, P. Borghesani, and Z. Peng, "Investigation on the relationship between macropits and wear particles in a gear fatigue process,"Wear, vol. 484-485, p.203724, 11 2021.
>
> Chapter 6 expands on the published journal article below.
>
> [3] H. Chang, P. Borghesani, and Z. Peng, "Automated assessment of gear wear mechanism and severity using mould images and convolutional neural networks," Tribology International, vol. 147, p.106280, 2 2020.
>
> Chapter 7 is a modified version of the manuscript below which will be submitted to the journal of Wear in Jan/Feb 2022.
>
> [4] H. Chang and Z. Peng, "Image-based gear wear prediction: A deep learning approach," Wear, 2022.
>
> The works above are documented and acknowledged at the beginning of each chapter.

Candidate's Declaration

I declare that I have complied with the Thesis Examination Procedure.

Abstract

Gears are widely used in industrial machinery, and gear failure is a main cause of machine failure. Since gear wear is often the initial stage of gear failure, its monitoring and prediction are key to proactive maintenance for critical machines, which could minimise downtime, maintenance costs, and safety risks. However, existing gear wear monitoring and prediction techniques face some ongoing technical challenges, including how to provide direct wear information and wear assessment and prediction in a cost-effective and efficient manner. To tackle the challenges, this research aims to develop a set of advanced techniques for gear wear monitoring and prediction. The four objectives of the research and their corresponding methodologies and outcomes are summarised as follow.

(a) To develop a method to obtain direct and comprehensive gear wear information without disassembling the gearbox. This objective was realised by combining surface replication with image analysis, allowing easy acquisition of high-resolution gear mould images showing wear evolution on a tooth flank. (b) To investigate the relationship between the features of worn gear surfaces and those of wear debris. To further understand the role of wear debris analysis in wear assessment, a study on various features of macropits and wear particles in the same fatigue process was conducted and provided new insights of gear pitting and its monitoring. (c) To develop an automated system for gear wear assessment. Deep learning models were developed to identify wear mechanisms and severities using gear mould images and wear debris images. High classification accuracies were achieved, and comparisons between the two image sources were made. (d) To develop a gear wear prediction model using direct wear information. A deep generative model was developed and trained on time series of gear mould images. Tests showed that the model using the state-of-the-art AI technology can generate realistic and accurate predictions.

Overall, this research addressed the main limitations of existing methods and provided a direct and evidence-based tool for monitoring and predicting gear wear. Its specific contributions include (i) a new moulding-imaging method for monitoring gear wear evolution, (ii) a detailed comparison between worn gear surfaces and wear debris in a wear process, (iii) and AI and image-based gear wear assessment and prediction models for the first time. The techniques could be performed during regular inspections of machines and used with online methods for increased robustness.

Acknowledgement

This research was supported by the Australian Government through the Australian Research Council Discovery Project DP160103501 and the Australian Government Research Training Program Scholarship.

I would like to express my deepest gratitude to my supervisor, Prof. Zhongxiao Peng, for all the support from encouraging me to pursue a PhD, to guiding me through various difficulties during my candidature, to helping me to produce this thesis. The consistent efforts you put in providing insightful and timely advice for all my questions and requests are extraordinary. You have set a great example for me in my next journey in numerous ways, and you are a remarkable mentor who I will not forget.

I would like to extend my full appreciation to my joint supervisor, Dr Pietro Borghesani, for your exceptional guidance. I have always been impressed by your critical thinking, problem-solving, articulating abilities, etc, and you have always been an inspiration to me.

I am also very grateful to Dr Wade Smith, who provided crucial support for me, especially during the drafting of my first paper. It was truly delightful to work with you, who are always kind and dedicated.

I was honoured to work with Em/Prof. Robert (Bob) Randall and obtain his recognition.

I cannot be thankful enough for my fellows: Dr Dikang Peng, Dr Hongkun Wu, Dr Ke Feng, Dr Peipei Feng, Dr Yifeng Fu, Jacky Chin, Dr Hengcheng Zhang, Siyuan Zhuang, Yuanning Mao, and Runyu Lu. The moments we shared across our daily life will never be forgotten and your friendships will be cherished.

Finally, special thanks to the ones I love, my parents, my wife, Yalun, my dog, Johnny, and others. Your unconditional supports gave me the strength to tackle every challenge during this journey. All I want to give is my heart, and all I want to say is 'I Love You'.

Publications and Presentations

List of publications during the PhD candidature

- **H. Chang**, P. Borghesani, and Z. Peng, "Investigation on the relationship between macropits and wear particles in a gear fatigue process," *Wear*, vol. 484-485, p.203724, 11 2021.
- **H. Chang**, P. Borghesani, and Z. Peng, "Automated assessment of gear wear mechanism and severity using mould images and convolutional neural networks," *Tribology International*, vol. 147, p.106280, 2 2020.
- P. Feng, P. Borghesani, **H. Chang**, W. A. Smith, R. B. Randall, and Z. Peng, "Monitoring gear surface degradation using cyclostationarity of acoustic emission," *Mechanical Systems and Signal Processing*, vol. 131, pp. 199–221, 9 2019.
- **H. Chang**, P. Borghesani, W. A. Smith, and Z. Peng, "Application of surface replication combined with image analysis to investigate wear evolution on gear teeth – A case study," *Wear*, vol. 430–431, pp. 355–368, 7 2019.

List of presentations

- **H. Chang**, "New techniques for gear wear monitoring and prediction," *23rd International Conference on Wear of Materials*, Online, 4 2021.

Manuscript to be submitted

- **H. Chang** and Z. Peng, "Image-based gear wear prediction: A deep learning approach," *Wear*, 2022.

Abbreviations

AE	Acoustic emission
CBM	Condition-based maintenance
CM	Condition monitoring
CNN	Convolutional neural network
ConvLSTM	Convolutional long short-term memory
EAP	End of active profile
EHL	Elastohydrodynamic lubrication
GAN	Generative adversarial network
HOG	Histogram of oriented gradients
IRT	Infrared thermography
LSCM	Laser scanning confocal microscopy
LSTM	Long short-term memory
MS-SSIM	Multi-scale structural similarity index measure
OA	Oil analysis
OLVF	Online visual ferrograph
PAE	Pitted area error
PRE	Pitted region error
PCA	Principal component analysis
RCR	Relative closeness ratio
RUL	Remaining useful life
SAP	Start of active profile
SEM	Scanning electron microscopy
SVM	Support vector machine
TBM	Time-based maintenance
VAE	Variational autoencoder
WDA	Wear debris analysis

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Research Background

Gears are crucial mechanical components for power transmission in many industry sectors such as wind power, metal & mining, and power generation. The global market for industrial gearbox reportedly reached US $38.4 billion in 2020 [1]. Gear failures caused by gear wear and/or faults are one of the leading causes of machine failures, which often lead to significant downtime, maintenance costs, and threats to human safety. For instance, surveys conducted in [2, 3] found that gears used in onshore wind farms usually have a moderate failure rate among all critical components of a wind turbine but cause the longest downtime (up to 33% of the total downtime). For helicopters, it was found that drive train (gearbox is a main component) failures made up nearly 20% of their accidents in the US from 1963 to 1997 [4].

In order to minimise the impacts brought by gear faults and machine failures, the maintenance strategies for machines and their critical components have evolved from run-to-failure maintenance, to (time-based) preventive maintenance and then condition-based maintenance (CBM) [5]. The run-to-failure maintenance is only suitable for low-cost and easy-to-replace machines, and time-based preventive maintenance that is performed at reg-

ular intervals often leads to high but unnecessary maintenance costs and potential damage during the maintenance. Recognised as the most efficient maintenance strategy [5], CBM allows maintenance to be conducted at the optimal time by monitoring a machine's condition and carrying out essential maintenance based on the fault severity. As a result, CBM can minimise maintenance costs and maximise the machine's remaining useful life (RUL).

Many condition monitoring (CM) methods, including vibration, acoustic emission (AE), and oil/wear debris analysis, have been developed for CBM. Among the existing techniques, vibration and AE are predominantly used for gear fault detection and diagnostics instead of gear wear monitoring and prediction. Gear wear monitoring and prediction are crucial because the initial stage of most gear faults is often gear wear, which is an inevitable degradation process of the gear tooth flank [6]. Gear wear is still difficult to be detected using most CM methods before it is noticeably severe (e.g., visible dimensional change of the tooth profile), except visual inspection and wear debris analysis (WDA) [5,7,8]. They have been routinely used for gear wear monitoring in many applications [9], and they have shown stronger capabilities than other methods in revealing gear wear in the early stage. This is because they examine either the direct indications of gear wear (surface appearance by visual inspection) or its by-products (wear debris by WDA) [7].

However, existing visual inspection and WDA approaches have some limitations in gear wear monitoring as follows.

- Wear inspection and assessment of gears often require dismantling of the gearbox to perform visual inspection by trained personnel. For detailed fault diagnosis, tooth samples are often obtained using a destructive method to allow further examinations (e.g., examine micro-cracks on gear surfaces) using a microscope. To avoid causing damage to gears and reducing their useful lives, visual inspections of gear teeth are usually carried out using naked eyes or tools with limited magnification. As a consequence, micro-scale surface information that is critical for early-stage wear monitoring is often unavailable when using the commonly practiced visual inspection method.

- WDA faces challenges in efficiently acquiring accurate gear wear information. Firstly, wear debris may undergo deformation or fracture due to their entrapment between gear contacts [10–12], resulting in potentially different sizes, shapes, and morphology from their original states. Secondly, existing techniques and sensors for online imaging of wear debris still experience some difficulties in acquiring high-quality wear debris images (e.g., lubricant contamination) and processing large batches of images in an efficient manner [13].

- The assessment of gear wear mechanism and severity still heavily relies on trained analysts for both methods. The main reason is that the assessment criteria are primarily qualitative, and practical experience is needed for a proper judgement. Employing trained analysts is increasingly unattractive due to its low turnaround, high cost, and potential subjectivity and human error.

As mentioned before, the existing techniques (visual inspection and WDA) were developed and are commonly applied to wear analysis. Wear prediction based on the current wear condition (wear mechanism and severity) and its propagation rate is crucial because proactive maintenance actions can be planned in advance. For gear wear prediction, physics-based methods (using a bottom-up approach) and data-driven methods (solely relying on CM data) are the two primary groups [14, 15]. Physics-based methods aim to model the gear wear process based on related physical laws. However, the modelling work becomes extremely complex and challenging when including many impacting factors, including material properties, operating conditions (load and speed), lubrication conditions, environmental conditions (temperature, humidity), and potentially the co-occurrence of multiple wear mechanisms [16]. Data-driven methods aim to predict gear wear using CM data solely and they are becoming increasingly popular thanks to the fast development of machine learning technology. Although many traditional problems of data-driven methods, such as data scarcity and feature engineering, are being solved, data quality remains a critical issue. Specifically, commonly used CM data such as vibration or AE signals contain limited gear wear information. In addition, the data could be easily contaminated by noises from inherent imperfections of the gear, the signal transmission path, and the vary-

ing operating conditions of the gearbox, etc [5, 7, 17]. Although images of worn gear teeth and images of wear debris contain valuable gear wear information, limited work has been conducted and reported on utilising the information and machine learning approaches to monitor and predict gear wear.

1.2 Research Objectives

To address the above-mentioned limitations of existing techniques, **this research aims to develop advanced techniques for gear wear monitoring and prediction** using direct gear wear data generated experimentally. To achieve the aim, four objectives are defined in this research.

1. **To develop a method to obtain direct and comprehensive gear wear information without disassembling the gearbox.** Gear teeth are usually difficult to access by high-resolution imaging facilities when they are in the gearbox (space issue), and therefore direct gear wear information is largely forsaken for gear wear monitoring and prediction. This objective will be realised by combining surface replication (moulding) and image analysis to easily acquire high-resolution and comprehensive gear wear information (tooth appearance, surface morphology, tooth profile, surface roughness, etc.). This development will provide a useful tool for obtaining high-resolution, direct wear information on gear surfaces.

2. **To investigate the relationship between the features of worn gear surfaces and those of wear debris.** This task is to deepen the understanding of to what extent collected wear debris has similar tribological features to those of the gear surfaces where the wear debris are generated. An experimental investigation that both qualitatively and quantitatively compares various features of macropits and wear debris in the same gear pitting process will be conducted for the first time. The investigation will provide new insights into the gear pitting process and experimental evidence for improving WDA in gear wear monitoring.

3. **To develop an automated system for gear wear mechanism and severity assessment.** This task is to reduce the cost and improve the efficiency of gear wear assessment, which is predominantly performed by trained analysts. This objective will be achieved by developing deep learning models (convolutional neural networks (CNNs)) to identify gear wear mechanisms and severities using gear mould images and wear debris images. This system will utilise the great capability of CNNs in image classification to achieve gear wear assessment with high accuracy and efficiency.

4. **To develop a gear wear prediction model using direct wear information.** Direct records of gear wear evolution (e.g., gear mould images) are highly valuable data for the understanding of gear wear dynamics, but they have not been used for gear wear prediction. Therefore, a deep generative model, which uses a limited number of history gear mould images as inputs and generates a future image as the output, will be developed for the first time. The wear prediction model will employ advanced deep learning techniques to generates realistic and accurate gear wear predictions.

1.3 Thesis Structure

In addition to this introduction chapter, this thesis consists of 8 chapters, including a literature review chapter, a methodology chapter, and 4 technical chapters (Chapters 4 to 7) corresponding to the four objectives, followed by Discussion and Conclusions. Chapters 4 to 6 are modified versions of three journal articles published during the PhD candidature, and Chapter 7 is a modified journal article that will be submitted to *Wear* in 2022. Details about the articles and modifications will be presented at the beginning of each of the chapters. The brief introduction and organisation of each chapter is presented below.

Chapter 1 - This chapter introduces the general background and objectives of this research and the structure of this thesis.

Chapter 2 - This chapter reviews important literature related to gear wear monitoring and prediction. It covers topics including common gear wear mechanisms and their tribological features, existing techniques for gear monitoring, image-based gear wear assessment, and existing approaches and developments for wear predication. This chapter ends with a summary of the key developments and research challenges/gaps in the area.

Chapter 3 - This chapter outlines the methodology designed to achieve the four objectives in this research. Experimental tests are first planned so that the moulding technique can be developed and applied to obtain gear moulds, and wear particles can be collected during the tests. Then, the techniques and procedures for collection and analysis of mould and wear particle samples are presented. Finally, the development of image-based gear wear assessment and prediction using deep learning approaches is presented.

Chapter 4 - This chapter presents the development of a novel method that combines surface replication and image analysis to obtain gear moulds and their images for investigating the tribological features and their changes of the gear surfaces in a wear process. Its application in a gear wear test demonstrates that comprehensive and high-resolution wear information of worn gear can be obtained using this method without disassembling the gearbox.

Chapter 5 - This chapter details the investigation on the relationship between gear macropits and wear debris in a surface fatigue process. The investigation compares macropits and wear debris qualitatively (via their images at different times in the same gear test) and quantitatively (via numerical features regarding their concentration, size, shape, and edge details), and provides new insights into the evolution of the gear surfaces and their wear particles in the gear wear process.

Chapter 6 - This chapter details the development of automated assessment systems for gear wear mechanism and severity assessment. Deep CNNs are developed to identify gear wear mechanisms and severities using gear mould images and wear debris images. Key procedures, including data preparation, model construction, and model training, are detailed, and the models' performance in tests are systematically evaluated.

Chapter 7 - This chapter details the development of an image-based gear wear prediction model using a deep learning approach. A time series of gear mould images is similar to a short video, and therefore, deep learning techniques developed for video prediction are adapted to gear wear prediction. The preparation of mould image sequences and other datasets is detailed, and the construction and training of the generative adversarial network (GAN) are introduced. Lastly, qualitative and quantitative evaluations of the realism and accuracy of predicted images are presented.

Chapter 8 - This chapter discusses how the moulding-imaging techniques developed in this research can be used in real-world applications. It also presents how the further insights into the evolution of the tribological features of wear particles generated in the wear process and their relationship with those of worn surfaces can be helpful for further development of wear monitoring techniques based on wear debris. Reflections on the AI-based wear classification and prediction are also included, followed by discussions on the strengths and limitations of the works in the thesis.

Chapter 9 - This chapter summaries key findings obtained in this research and recommends future work.

Chapter 2

Literature Review

Due to the increasing demand for reducing downtime and maintenance costs of critical machines, where gears are widely used to transmit power, gear wear monitoring and prediction has received growing research interests in recent years. This chapter reviews common gear wear mechanisms and their surface features and main methods for gear wear monitoring and prediction, with a focus on their strengths and weaknesses, and highlights urgent challenges that need to be tackled for a higher reliability and efficiency. This research is conducted within the context of condition-based maintenance (CBM), and this topic is briefly introduced below.

CBM, or predictive maintenance, is "a maintenance program that recommends maintenance actions based on the information collected through condition monitoring" [18]. A CBM program usually starts from the acquisition of machine health data (data acquisition), followed by processing the raw data to estimate the machine condition (data processing), and ends by recommending appropriate maintenance tasks (decision-making) [18]. The core of CBM is condition monitoring (CM) that covers the process of data acquisition and data processing. Diagnostics and prognostics, which provide crucial support to the decision-making process, are two critical procedures of CBM. In recent decades, the rapid development of various CM techniques with enhanced diagnostics and prognostics capability has given rise to the expansion of CBM in many industries [5, 19]. Diagnostics refers

to the process of determining the location and severity of a fault as well as identifying its potential cause(s). Prognostics refers to the process of predicting the remaining useful life (RUL) of a component based on the current condition of its degradation [18]. Diagnostics is a reactive process when a fault occurs, while prognostics is a proactive process to predict and prevent a failure caused by the fault [14], which assists in condition-based maintenance scheduling. Machine condition monitoring (MCM) with prognostics capability allows maintenance to be planned in advance and facilitates the optimisation of maintenance schedule, which could minimise machine downtime and maximise its availability [18].

In this chapter, section 2.1 overviews gear wear and focuses on the main mechanisms and different severity levels of gear wear and their surface features. Section 2.2 reviews the main techniques for gear wear monitoring. Section 2.3 reviews image-based gear wear assessment, focusing on image acquisition techniques and conventional and deep learning based image analysis techniques. Section 2.4 reviews the main methods for gear wear prediction. Section 2.5 summarises this chapter and identifies key research gaps in gear wear monitoring and prediction.

2.1 Gear wear: mechanisms, severities, and surface features

In general, gear wear refers to the material removal process occurring on the gear tooth flank during gear operation, and sometimes it also includes material transfer on the tooth flank (e.g., the plastic flow of material) [20, 21]. Gear wear is one of the most common forms of gear degradation, and it affects gear performance. Severe gear wear can induce excessive transmission error, vibration and noise, and could even lead to tooth breakage.

In tribology, it is widely accepted that there are mainly four types of wear mechanisms, including adhesive wear, abrasive wear, fatigue wear, and corrosive wear [22–25], and all of them can occur in gear contacts. The term "wear mechanism" is sometimes used interchangeably with "wear mode" [25, 26]. As an extremely complex process, gear wear can be

impacted by numerous factors, including load, speed, gear material, lubrication, tooth profile, surface roughness, etc. As the impacting factors vary, gear wear's mechanism, mode, and severity can be drastically different, resulting in different tooth surface appearances. The ability to link the surface appearance of a worn gear to its wear mechanism and severity levels is critical in gear wear monitoring. Therefore, this section provides an overview of common gear wear mechanisms and distinct surface features for wear mechanism and severity assessment.

2.1.1 Adhesive wear

Adhesive wear refers to material being transferred from one surface to the other, due to welding and tearing between two contact surfaces [21,22]. The wielding results from solid-solid adhesion without intervening medium such as water and oil, and adhesion between metals is particularly strong due to the formation of metallic bonding [27]. Metal surfaces, except for noble metals, are usually covered by a thin layer of oxidation film that prevents the formation of metallic bonding. In lubricated contacts, the oil film between two contact tooth surfaces has similar separation effects. However, when the separation is lost, metallic bonding forms in the real contact area when two opposing surfaces come into contact, as shown in Figure 2.1. Then, when the two surfaces move away from each other, plastic deformation will occur in the weakest region in the softer material, initiating internal cracks. The cracks propagate until a piece of material is pulled out and transferred to the other material.

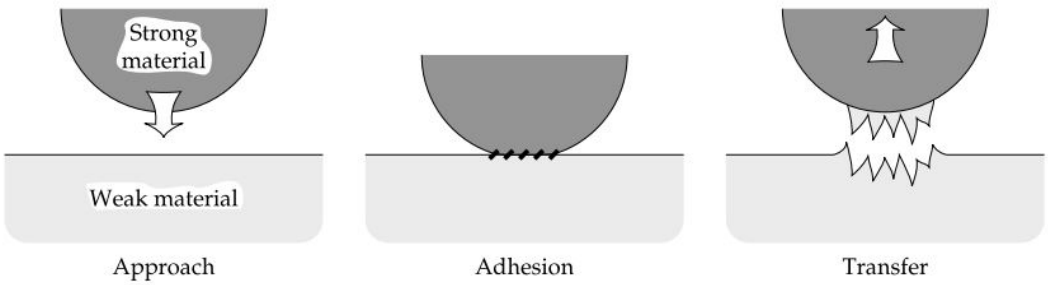


Figure 2.1: A schematic of the adhesive wear process [27].

Adhesive wear is mild or moderate when the welding is microscopic and only thin surface films are transferred from one gear to the other [21]. It becomes severe when metal-metal adhesion is formed and large pieces of material are transferred. Mild adhesive wear causes nearly no alteration of the tooth surface, and it usually stops after running-in. Moderate adhesive wear partially or entirely eliminates original manufacturing marks, but the tooth surface is not noticeably rough [21].

Severe adhesive wear on gear teeth, also known as scuffing, has significant impacts on gear operation and is often more concerning in gear CBM. Due to repetitive wielding and tearing, the surfaces are often very rough and have a matte texture. Scuffed areas are often manifested as bands separated by the pitch line, as shown in Figure 2.2 [28, 29]. At the micro scale, scuffing features rough and torn surfaces with signs of plastic deformation, as shown in Figure 2.3. In practice, the severity of scuffing can be classified into three levels, namely mild, moderate, and severe, according to the scuffed area compared to the area of the active tooth flank [21].

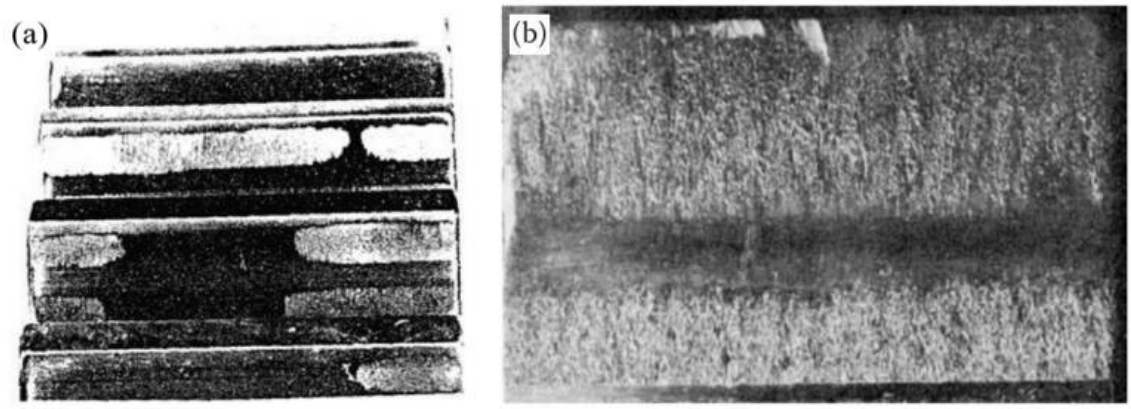


Figure 2.2: Examples of scuffing on gear teeth at the macro scale [28, 29].

Adhesive wear often occurs in dry contacts of metal gears where the oxidation films on gear teeth can be easily broken. In lubricated gear systems, the occurrence of adhesive wear often indicates lubrication failure, which is often caused by low oil viscosity, high oil temperature, high load, etc [29]. The wear rate of adhesive wear is usually high, and it often signals a rapid deterioration of contact surfaces. Gear scuffing can be controlled

by using the lubricating oil with a higher viscosity, improving oil cooling, using extreme pressure (EP) additives, etc [21, 29].

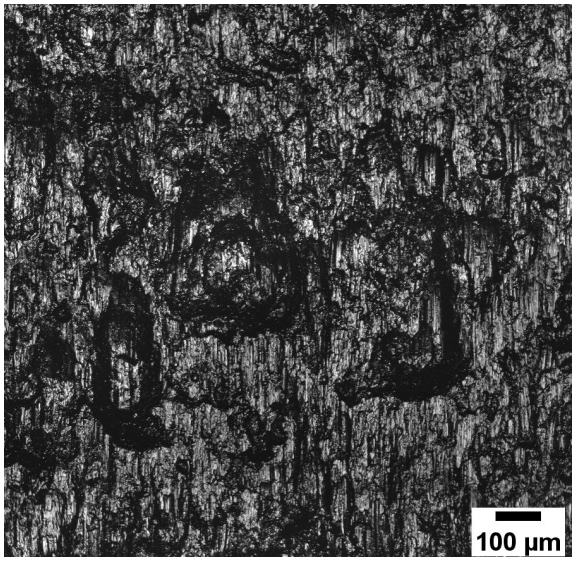


Figure 2.3: An example of scuffing on gear teeth at the micro scale.

2.1.2 Abrasive wear

Abrasive wear is a prevalent wear mechanism, and it refers to material removal via cutting or ploughing by hard particles or protuberances [7]. It is usually classified into two-body and three-body abrasive wear. When the abrading bodies are fixed on one contact surface, such as asperities or embedded particles, and directly abrade the other surface, the process is two-body abrasive wear [27]. When the abrading bodies are loose particles that roll and slide between two contact surfaces, the process is three-body abrasive wear [27]. Abrasive wear on gears can usually be treated as the combination of two-body and three-body wear, since surface asperities on gear teeth and entrapped wear particles can all act as abrasives.

Abrasive wear on gear teeth can be distinguished by scratches along the sliding direction, as shown in Figure 2.4 (a) [28]. At the micro scale, parallel furrows along the sliding direction indicate abrasive wear, as shown in Figure 2.4 (b). The severity of abrasive wear is usually determined by the number of abrasive scratches and remaining manufacturing marks [21].

Mild abrasive wear features a small number of scratches, and most of the manufacturing marks are visible. When most manufacturing marks are removed, the wear severity is considered moderate. Severe abrasive wear eliminates manufacturing marks, and the tooth profile may be noticeably altered. Geometric deviations of gear teeth caused by abrasive wear usually increase transmission error, dynamic loads, vibration, and noise [30].

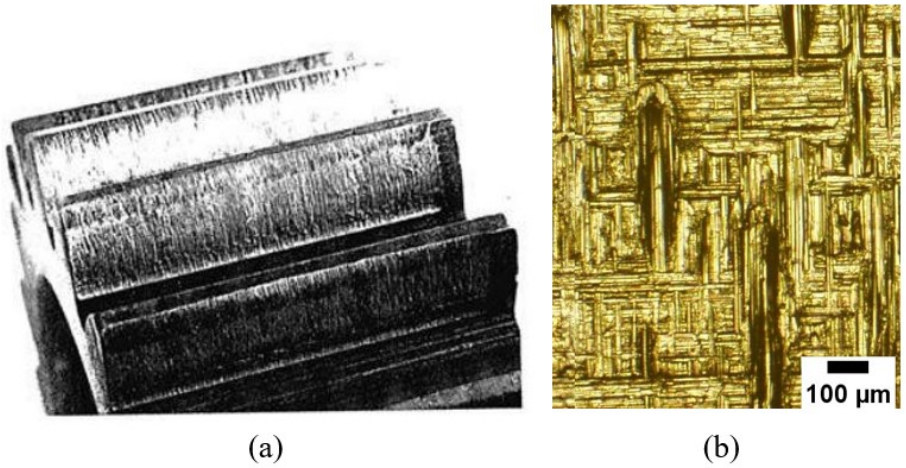


Figure 2.4: Abrasive wear on gear teeth. (a) A macroscopic view [28]. (b) Microscopic furrows in parallel. Horizontal lines are manufacturing marks.

A major impacting factor on abrasive wear is the material hardness, and the relative abrasive wear resistance of a material is proportional to its hardness [31]. Therefore, using harder materials, applying hard surface coatings, or surface heat treatment are widely used to reduce abrasive wear. Surface roughness also has an important effect on abrasive wear in that high surface roughness enhances asperity contacts and wear particles generation. In lubricated contacts, the specific oil film thickness, which is the ratio of the oil film thickness to the composite surface roughness of two surfaces, determines the extent of separation between the two surfaces and negatively impacts the rate of abrasive wear [32]. The Archard's law is widely used to calculate the wear rate for sliding contacts, and further information can be found in section 2.4.

2.1.3 Fatigue wear

Fatigue wear is caused by surface fatigue accumulated under cyclic rolling and/or sliding [33]. Unlike common fatigue that could lead to structural fractures such as bending fatigue of gear teeth, fatigue wear only affects surface and near-surface materials. However, they share similar development stages, namely, crack initiation and propagation. For fatigue wear, cracks initiate due to the accumulation of plastic deformation at stress concentrations, which could be at the contact surface such as surface asperities [34], or the sub-surface such as inclusions and voids [35]. After initiation, cracks propagate within the sub-surface under cyclic loading until the detachment of material fragments. Figure 2.5 [27] illustrates the process of surface-initiated fatigue wear.

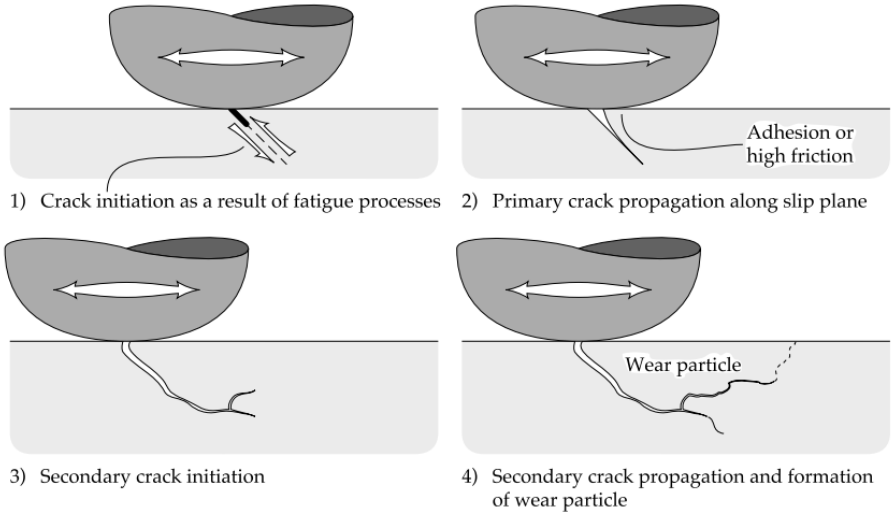


Figure 2.5: A schematic of surface-initiated fatigue wear [27].

There are several types of fatigue wear for gears, and the most common ones are macropitting and micropitting. As suggested by their names, macropitting and micropitting are distinguished by the size and depth of pits. Macropits are usually larger than 100 µm and as deep as 20 to 30 µm. Macropitting [36–38] is widely regarded as related to macroscopic stress distribution in Hertzian contacts where the maximum shear stress reaches at Hertzian depth (below the contact surface) [39]. Micropits are usually smaller than 20 µm and as shallow as a few microns. Micropitting is closely related to localised stress

distribution at the asperity scale [21,40]. Figure 2.6 [41–43] shows images of macropitting and micropitting on gear teeth at the macro and micro scales.

The severity of macropitting is usually evaluated by the ratio of pitted area to the area of the active tooth flank [21]. Mild macropitting occurs when a small area on the tooth flank is covered by macropits, and moderate macropitting corresponds to a large surface area being pitted. Macropitting is considered severe when a significant portion of the tooth flank is pitted, such as the entire dedendum. Severe macropitting substantially increases the risk of tooth flank breakage and usually signals the end of a gear's useful life. Compared to macropitting that could notably alter the tooth profile and impact the normal operation of gears, micropitting is less harmful and less likely to cause catastrophic failure [21]. However, micropitting could promote abrasive wear and macropitting due to the roughened surface with a high crack density [43].

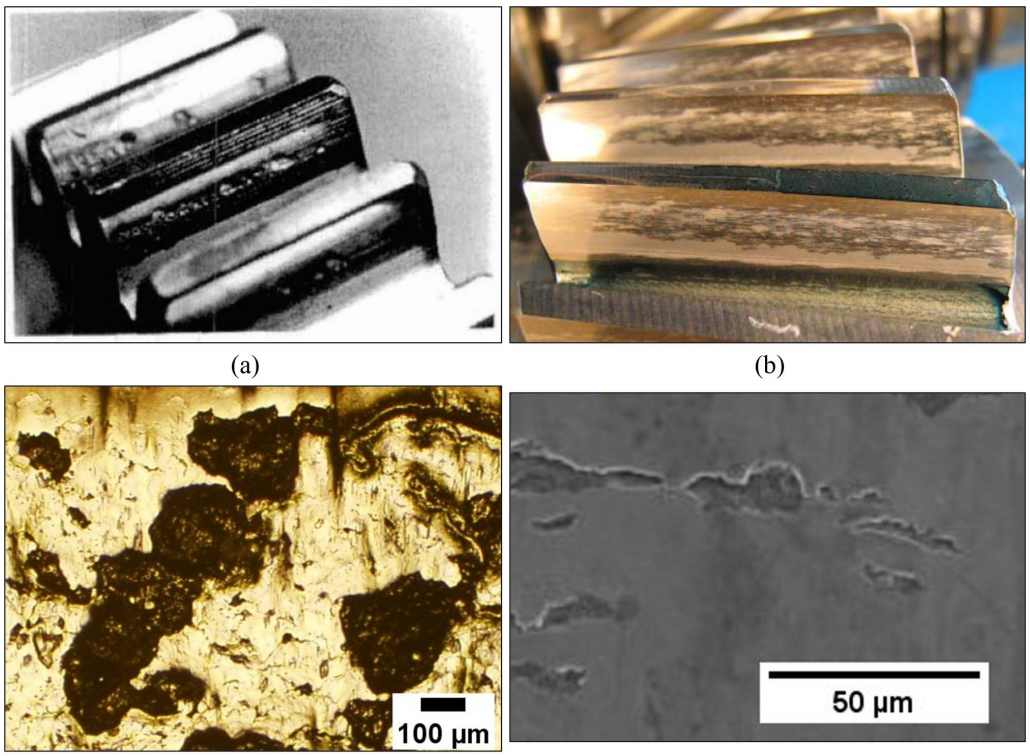


Figure 2.6: Examples of macropitting and micropitting on gear teeth. (a) Macropitting on a spur gear [41], (b) micropitting on a helical gear [42], (c) macropits on a gear tooth mould at the micro scale, (d) micropits at the micro scale [43].

The useful lives of many lubricated gears are often governed by fatigue wear since other types of wear that can be effectively minimised with careful selection of material, lubricating oil, and other parameters [44]. Common approaches to reduce or control fatigue wear focus on alleviating stress concentrations, such as lowering loads, selecting materials with fewer impurities, increasing surface hardness, lowering surface roughness, and increasing the oil viscosity [21, 27].

2.1.4 Corrosive wear

Corrosive wear occurs due to chemical or electrochemical reactions between a surface and its environment [22, 24]. For contact surfaces of mechanical components, corrosive wear often co-occurs with mechanical damage caused by other types of wear, and thus a more precise term, tribocorrosion, is frequently used in those cases [7, 27]. Gear tribocorrosion can be identified by stained or rusty gear tooth or etch pits across the tooth flank, as shown in Figure 2.7 [45].

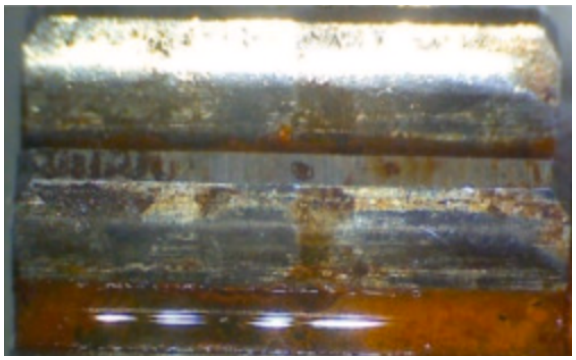


Figure 2.7: Rusty gear teeth due to tribocorrosion [45].

Tribocorrosion often leads to a higher wear rate than either of the two acts alone. For instance, reaction products that loosely adhere to contact surfaces can be easily removed by abrasive wear, and cleaned surfaces can easily react with corrosive reagents, thus boosting the wear rate. In this case, severe tribocorrosion may lead to a considerable wear depth and increase the risk of tooth breakage. Nevertheless, tribocorrosion can also reduce the wear rate in some cases. For instance, extreme pressure (EP) additives [46] are commonly used

for lubricated gears under high pressure to avoid scuffing because the additives react with metals under high pressure, forming a strong film on the gear tooth separating two contact surfaces. Gear tribocorrosion is often caused by oxygen, water, and elevated acidity in the lubricating oil [21], and therefore, gear tribocorrosion can be controlled by oil changing and reducing the chance of contaminants (e.g., water) entering the oil. Monitoring oil properties such as water content and acidity is also very beneficial to minimise the risk of tribocorrosion [47].

2.2 Gear wear monitoring techniques

Gear wear is the most common cause of gear degradation, and through monitoring its development and taking proper maintenance actions, the risk of gear failure could be effectively minimised. Gear wear usually shows various indications and they can be classified into direct indications and indirect indications [7, 48], as shown in Figure 2.8 [7]. As a result, various techniques have been developed to acquire and analyse different types of data related to those indications. The following subsections concisely review main techniques for gear wear monitoring, including an overview of common techniques (section 2.2.1), visual inspection and surface imaging (section 2.2.2), and oil and wear debris analysis (section 2.2.3).

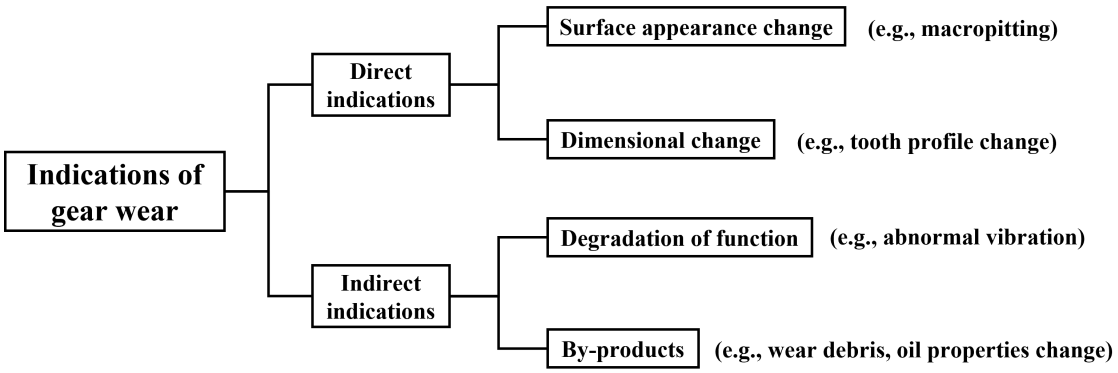


Figure 2.8: Categories of gear wear indications with examples [7].

2.2.1 Overview of common techniques

Currently, oil analysis (OA) and wear debris analysis (WDA) are the most popular techniques for gear wear monitoring in industrial applications [49], while visual inspection and surface imaging are still commonly adopted. Further information about the two methods can be found in section 2.2.2 (for visual inspection and surface imaging) and section 2.2.3 (for OA and WDA). Other techniques that have been predominantly used for gear fault detection and diagnosis, such as vibration analysis, are also being applied to gear wear monitoring in a few recent studies [50,51]. Compared to using direct gear wear data, they have main advantages of non-intrusive and real-time monitoring of health condition of rotating components. However, the existing vibration-based techniques also have inevitable limitations for wear analysis. The main reason is that vibration signals of a gearbox system are usually collected using sensors located on the casing and thus contain the dynamic responses from the gears and other components, such as shafts, bearings, couplings. Thus, the signals are complex for gear wear analysis and often have low sensitivity to early-stage gear wear [5]. This section briefly introduces three types of techniques that have shown some capabilities in gear wear monitoring, including thermography (section 2.2.1.1), vibration analysis (section 2.2.1.2), and acoustic emission analysis (section 2.2.1.3).

2.2.1.1 Infrared thermography

Infrared thermography (IRT) refers to the process of using thermal cameras to capture infrared radiation emitted by an object, and converting the radiation to thermal images showing its temperature distribution [52]. The use of IRT is based on the fact that any object emits electromagnetic radiation when its temperature is over absolute zero (0 K). The radiation can be related to the object's temperature using Stefan-Boltzmann law [53]. IRT is a non-contact and non-invasive technique, and it allows real-time visualisation of the temperature distribution of an object [52]. Due to the advantages, IRT is extensively used for medical applications, and it has also been used in the condition monitoring for civil structures, electrical and electronic equipment, and machinery [52, 53]. IRT can

be classified as passive IRT and active IRT according to the adoption of external thermal stimulation [52]. For the condition monitoring for gearboxes, passive IRT is often preferred because it does not require the gearbox to be stopped and opened [54], and the temperature distribution can be monitored remotely using an infrared camera. Existing studies that used IRT for the condition monitoring of gears primarily focus on fault detection, and studies on gear wear monitoring are scarce [54]. Resendiz-Ochoa et al. [54] used an artificial neural network to classify the infrared thermal images of a gearbox with four different gear wear depths as shown in Figure 2.9 [54] (healthy, 25%, 50%, and 75% of uniform wear, with 100% of uniform wear referring to the total removal of gear teeth). The results showed that the thermal images could be used to distinguish the four gear wear depths with 100% accuracy using the neural network. However, the differences between the four wear depths were macroscopic (up to several millimetres, see the upper row in Figure 2.9), and the capability of IRT in monitoring gear wear with a higher resolution (e.g., $<$ 100 µm) remains to be investigated.

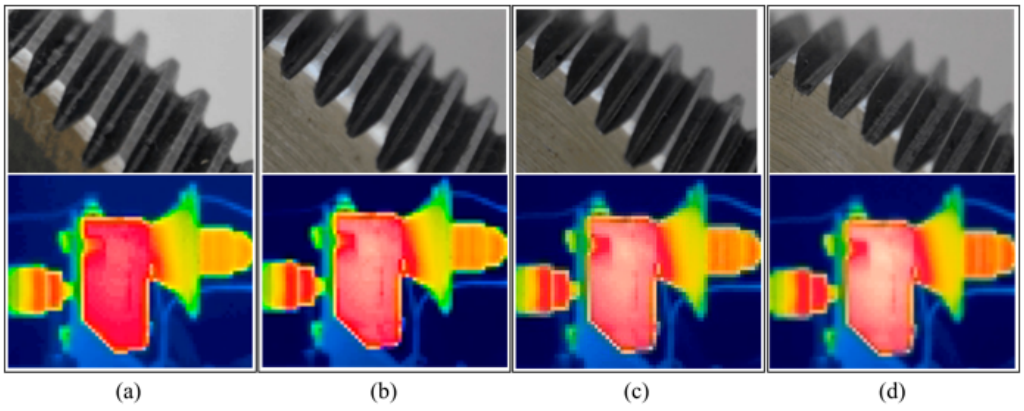


Figure 2.9: Photos of different gear wear depths and the corresponding thermal images of the gearbox captured by infrared thermography [54]. (a) Healthy gear, (b) 25%, (c) 50%, (d) 75% of uniform gear wear.

2.2.1.2 Vibration analysis

Vibration analysis is the most popular CM method with decades of successful industrial applications. It uses vibration sensors (e.g., accelerometers) to obtain vibration signals

and employs signal processing techniques to estimate the machine condition. It is a powerful method for fault detection and diagnosis because the vibration signature of a machine deviates from its healthy one when a fault occurs, and the source of the fault is physically related to the deviation [5]. Applying mature signal processing techniques, in the time domain, frequency domain, and time-frequency domain, to vibration signals greatly enhance its capability to separate fault indications from other masking signals. The capability of vibration analysis in monitoring gear wear is less studied than that in fault detection and diagnosis, partly due to the coupling effect between gear wear and gear dynamics [51]. Hu et al. [55] proposed two vibration-based indicators for gear wear monitoring, namely ALR and mALR, which measured the generalised distance between two gear state vectors extracted from time synchronous averaged vibration signals. Two gear tests demonstrated the effectiveness of the two indicators in detecting notable changes to the gear state. Feng et al. [51] considered the differences between abrasion- and pitting-induced gear vibration features and proposed an indicator of second-order cyclostationarity (ICS2) to separate fatigue pitting and abrasive wear. Moreover, the proposed indicator showed micro-level wear tracking capability in a lubricated and a dry gear test. Nevertheless, further developments of vibration-based gear wear monitoring are needed to tackle several important issues. For example, the sensitivity of vibration signals to gear wear could be impacted by the inherent deviation of a gear tooth profile gear from its ideal geometry that is comparable to wear depth, and noise or masking signals from different sources of excitation ('blind source separation') [5]. In addition, varying operating conditions of gearboxes cause more difficulties for the carefully designed indicators to track gear wear evolution [6].

2.2.1.3 Acoustic emission analysis

Acoustic emission (AE) analysis is a relatively new CM method, which uses similar procedures and signal processing techniques to vibration analysis [56]. Acoustic emissions are solid-borne acoustic waves, e.g., generated during the interaction of two surfaces in relative motion, due to damages to a material. Piezoelectric sensors are usually used to capture AE signals [56]. Compared to vibration analysis, AE has shown greater sensitivity

to faults, and can provide earlier warnings of faults [56–58]. However, AE signals can be easily impacted by noises from other components, and the AE sensor needs to be close to the target component to reduce the impact [56].

AE has shown effectiveness in gear wear monitoring due to its high sensitivity to surface degradation [50, 57, 59, 60]. Tan et al. [57] conducted fatigue gear tests that ran at the same speed but three different loads and compared the pitting detection capability of three CM techniques, including AE, vibration and spectrometric oil analysis (SOA). The results showed that the RMS values of AE were linearly correlated to pitting growth at all tests and suggested that AE has good potential in gear pitting monitoring. Loutas et al. [59] applied AE, vibration and oil debris monitoring to a long-running gear test and extracted various features from vibration and AE signals to monitor the gear condition. Results showed that AE had a similar capability to vibration analysis in monitoring normal gear wear but demonstrated superiority in detecting cracked tooth. Sharma et al. [60] developed a theoretical model to correlate gear pitting (more specifically, pitting size) and other operating parameters with AE and experimental tests showed that the model could estimate the pit size with reasonable accuracy. Feng et al. [50] considered the cyclostationarity and non-Gaussianity of AE signals generated in lubricated gear contacts and proposed two indicators, namely cyclic kurtosis and cyclic scale parameter, to monitor gear wear evolution in a gear wear test. Direct measurement of gear wear in the test showed a close correlation between the two indicators and gear wear where cyclic kurtosis was sensitive to surface roughness evolution, and cyclic scale parameter was sensitive to pitting propagation.

2.2.2 Visual inspection and surface imaging

Visual inspection and surface imaging are vision-based techniques that examine direct indications of gear wear, i.e., changes in tooth surface appearance and tooth profile, to monitor gear wear by capturing and analysing images of gear teeth. Traditional visual inspection by naked eyes is a typical example of such techniques, and it is still among

the most common ways to assess the condition of a machine. Surface imaging using enhanced magnification and image processing techniques often provide more detailed surface information than visual inspection. Wear analysis using high-resolution images have the following main advantages.

- Image-based techniques provide higher monitoring accuracy and reliability than other CM techniques due to their direct nature [61, 62].
- Comprehensive wear information, such as surface appearance, surface roughness, profile change, and 3D shapes of defects, can be obtained [7].
- An automated process, from imaging to diagnosis and prognosis, can be easier to achieve, thanks to the fast development of deep learning techniques for image analysis in recent years [63].

However, the most significant disadvantage of vision-based techniques that limits their research development and industrial applications, is the difficulty accessing gear teeth by imaging facilities due to the limited space within the gearbox casing. For detailed examinations, the gearbox often needs to be dismantled, and worn or damaged gears are taken out so that they can be examined using high-resolution imaging facilities such as an optical microscope. For gear wear tests conducted in laboratories, gear teeth sometimes are cut off in order to use advanced imaging facilities such as scanning electron microscopy (SEM) for high-precision surface characterisation [64]. For industrial machines, disassembly of gearboxes is usually prohibited unless in major repairs. Therefore, the frequency and quality of visual inspection and surface imaging are often limited.

Different approaches have been adopted to address this issue, and surface replication has emerged as a feasible one that enables high-resolution imaging without disassembling the gearbox. Surface replication, or moulding, is the process to obtain negative replicas of active gear tooth flanks. It usually involves applying replication compounds to the valleys between certain teeth when the gearbox is stopped and opened, curing of the moulds, and taking off cured moulds [65]. After replication, the moulds can be transferred to imaging

facilities to capture 2D and 3D images, with which various surface information can be extracted. The procedures of surface replication and imaging of gear teeth are illustrated in Figure 2.10.

Surface replication has been successfully applied in surface manufacturing (with specific micro or nano geometries) and surface characterisation [65]. However, there are limited research works that applied surface replication to monitor gear wear. Andersson [66, 67] in 1974 proposed using acetyl cellulose films and acetone as the replication compounds for gears and observed the replica films using optical microscopy and SEM. However, the cured replicas were flexible films, and thus they were unable to retain the tooth profile information (only micro-scale surface features were impressed). The sample preparation was time-consuming, which involved pre-softening the acetyl cellulose film by acetone, gluing the films to glass slides, and coating a conductive layer on the samples for SEM examination. In addition, acetone might enter and contaminate lubricants in the process.

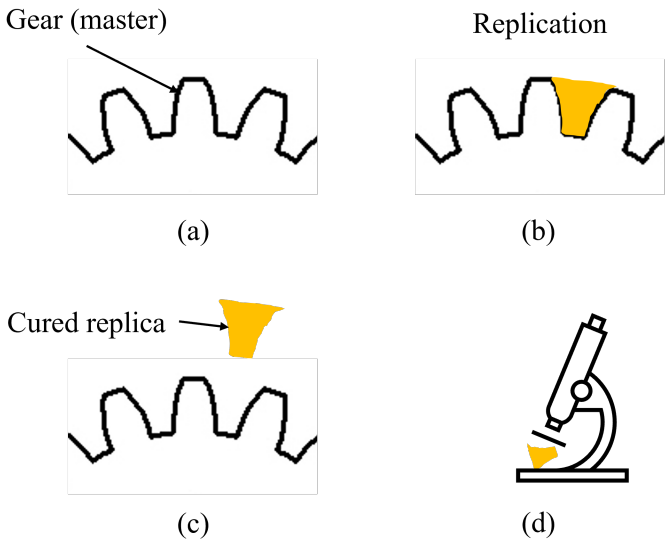


Figure 2.10: A schematic of surface replication and imaging for gears. The procedure usually has four steps: (a) surface preparation (cleaning), (b) applying replication compounds, (c) separating cured replica from the gear, and (d) replica imaging.

Then, Flodin [64,68] used a methyl methacrylate-based resin as the moulding compounds, which were supplied as liquid and powder. A clamp was specially designed to enclose the

compounds between gear teeth, and putty was also used for sealing, as shown in Figure 2.11 [68]. Rigid replicas were obtained, and a 3D stylus instrument was employed to measure their waviness and surface roughness. The author reported satisfactory performance of the replication method in measuring surface roughness but found that the replicas had serious deformation issues due to the significant internal stress, which led to the loss of actual tooth profile information [68]. More recently, Moorthy and Shaw [69, 70] used silicone polymers for gear surface replication. The replicas were then used to observe tooth surface damage under an optical microscope and measure tooth profile change using a coordinate measuring machine. Images of replicas showed many important wear features that were difficult to capture during normal visual inspections, and the use of silicone polymers prevented the loss of tooth profile. However, the replicas may deform when being in contact with the probe of the coordinate measuring machine, which might cause significant measurement errors. Moreover, due to potential issues in imaging, the image quality remained relatively low, and micro surface features were not captured, as shown in Figure 2.12 [69]. Similarly, Ames [71] reported using silicone polymers to make replicas of gear teeth and used both phase-shift interferometry and confocal microscopy to measure surface roughness and local wear depth of spiral bevels gears by superfinishing. The replicated surface appearances were captured using low-resolution imaging devices, and micro surface features were not examined.

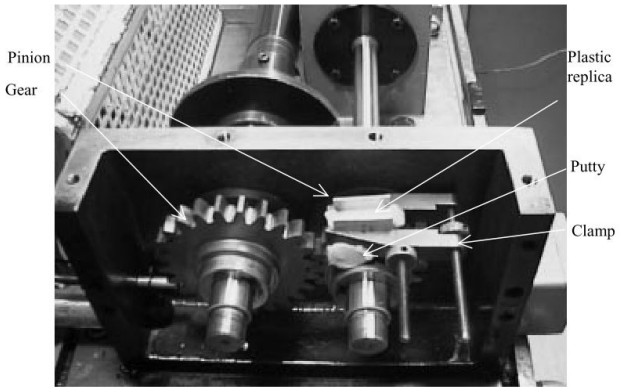


Figure 2.11: The specially designed clamp in [68] for surface replication on a FZG test rig.

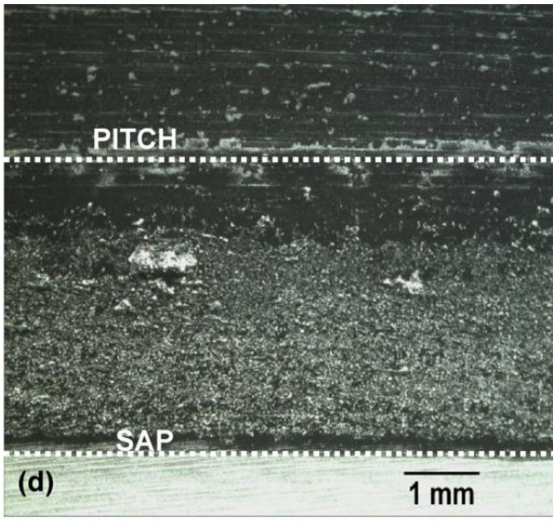


Figure 2.12: An optical stereo image of micropitting on a gear tooth replica [69].

2.2.3 Oil and wear debris analysis

Oil and wear debris analysis (OA/WDA) are commonly used CM methods that have shown effectiveness in providing early warnings of faults for lubricated components [26,47]. They are used to examine various features or properties of the lubricating oil and wear debris to estimate the gear wear rate, severity and mechanism. OA mainly focuses on key physical and chemical properties of the lubricating oil related to gear wear, such as viscosity, water content, and acidity [47]. WDA mainly examines wear debris features, such as concentration, size, shape, surface morphology [26].

2.2.3.1 Detection techniques for key wear debris features

The techniques to examine the key properties of the lubricating oil and wear debris have been developed for decades and many of them, particularly offline techniques, are mature and standardised [72–74]. Routine offline analyses such as spectrometric oil analysis (SOA) and particle counting are used to estimate the oil condition and wear debris's concentration, size, and chemical composition [9]. When routine analyses report abnormal

results, microscopic characterisation of wear particles is usually needed for further wear information. The characterisation is usually carried out in three steps: (i) periodic collection of oil samples, (ii) particle separation and specimen preparation, which can be done via methods such as analytical ferrography [75] and filter patch analysis [72], and (iii) specimen examination, usually under an optical microscope, to identify the concentration, size, and types of wear particles [73]. Offline OA and WDA systems that integrate analyses of multiple oil and wear debris properties are commercially available [76] and they improve the overall efficiency of offline analyses.

Despite the fact that offline techniques can provide detailed and accurate wear information, the demand for real-time monitoring has been driving the development of online oil and wear debris sensors for the last two decades. Online wear debris sensors [77–79] have drawn more attention since wear debris carries more direct information about gear wear than lubricating oil [49]. There are mainly five types of online wear debris sensors with different particle detection principles: inductive, capacitive, electrostatic, acoustic, and optical [47, 49]. These sensors have different capabilities in measuring wear debris properties, as summarised in Table 2.1 [47, 49, 80], which is determined by their working principles.

Table 2.1: Main types of online wear debris sensors, their detectable wear debris properties, and example commercial products [47, 49, 80].

Sensor type	Detectable wear debris properties	Example products
Inductive	Size, quantity, material	[81, 82]
Capacitive	Size, quantity	None
Electrostatic	Size, quantity	None
Acoustic	Size, quantity	None
Optical (photoelectric)	Size, quantity	[83]
Optical (imaging)	Size, quantity, morphology, material	None

Table 2.1 shows that most online wear debris sensors are designed to detect basic wear debris features, i.e., size and concentration/quantity, related to gear wear rate and severity. Inductive sensors usually introduce an external magnetic field (via coils) in the oil flow, and the magnetic flux increases when ferromagnetic particles (e.g., cast iron and steel) pass through the sensing zone [47, 49]. This is due to the high relative permeability of

ferromagnetic materials, despite the counteracting field generated by eddy currents. For other materials with low relative permeability (e.g., copper and titanium), the magnetic flux decreases, and thus inductive sensors can distinguish ferrous debris from non-ferrous ones. Moreover, inductive sensors are insensitive to air bubbles and water droplets in the oil because of their low relative permeability (close to oil) and negligible eddy currents. The microfluidic inductive sensors developed by Zhe et al. [79, 84–87] are representative research outcomes in recent years. The use of microfluidic channels restricted the flow path of wear debris to the sensing area with the highest magnetic flux density and reduced the chance of multiple particle passing simultaneously, thus increasing the sensor's sensitivity and accuracy. The low throughput problem caused by the small size of microfluidic channels were also addressed by employing multiple sensing channels in parallel. The difficulty of processing multi-channel data was also tackled by synchronised sampling. The sensors achieved accurate detection of small wear debris (11 µm in 1 mm diameter pipes) and larger wear debris with high throughput (50 µm particles at a flow rate of 750 ml/min) [79, 87]. Research on other types of sensors for monitoring the size and concentration of wear debris has reported less successful results than those on inductive sensors. Capacitive sensors utilise the change of capacitance between two electrodes in the lubricating oil when metallic particles pass to measure their size and quantity [49]. The main disadvantages of capacitive sensors are that (i) the capacitance of oil varies when oil properties such as viscosity and acidity change, and (ii) water droplets have a significant impact on capacitance measurement due to the high permittivity of water [47, 49]. Electrostatic sensors detect wear particles based on the phenomenon that adhesive wear particles are positively charged after their release. Several prototypes were designed in laboratory settings for scuffing wear monitoring for gears and bearings [88–91]. However, the mechanisms of debris charging have not been well understood, and thus, electrostatic sensors have not seen their industrial applications for gear wear monitoring [49,91]. Acoustic sensors detect wear particles by measuring the attenuation or the scattering of acoustic waves (usually ultrasonic waves) caused by the passing of particles [47]. An important feature of acoustic sensors is that air bubbles can be distinguished from wear debris due to their different reflection coefficients of acoustic waves [92]. Nevertheless, the lubricat-

ing oils absorb a notable amount of acoustic energy, and the absorption varies with the change of oil viscosity, leading to errors in size detection of acoustic sensors [49]. Lastly, photoelectric sensors, a group of optical sensors, detect wear particles by measuring the change of light transmission caused by passing particles, either light extinction or light scattering [49]. Many offline wear particle counters, such as [93], are designed based on this principle. There are also online extensions of these devices, such as [83]. However, online applications of photoelectric sensors for gear wear monitoring need to address the impacts of low oil transparency, air bubbles, particle overlap, etc [49].

Optical sensors that capture wear debris images, or imaging sensors, are the only ones that can capture comprehensive information of wear debris, including size, quantity, morphology, and material [49,78,94]. The morphology and material information are necessary to determine the type and source of wear debris, and they can be used to infer the wear mechanism and source. Therefore, numerous research on imaging sensors and related techniques have been conducted in recent years. Wu et al. [95] designed an online visual ferrograph (OLVF) sensor that used magnetic deposition to fix ferromagnetic particles in the oil flow and an imaging sensor with both reflected and transmitted light to capture wear debris images. Various tests [94–97] showed that images captured by OLVF could be used to extract the size, quantity, shape, and colour of wear debris. The issue of oil transparency that impacts particle detection was also recently addressed by Li et al. [98], who improved the OLVF sensor using dark-field imaging. However, due to the use of magnetic deposition, wear particles in OLVF tend to form chain-like clusters with particle overlap, causing significant difficulty in accurately extracting size and shape information of individual particles from the images, despite the relief from using advanced image segmentation techniques [49, 99]. A microfluidic approach was adopted by Peng et al. [100] to avoid the issue of particle overlap, and it also allowed multiple views of wear particles to be captured, which is vital for their 3D characterisation. Combining microfluidic channels with imaging sensors is a promising approach for online WDA, but several challenges need to be tackled before its industrial applications. Firstly, the small sizes of microfluidic channels are necessary to capture images of individual particles but seriously limit the oil flow rate [49]. Secondly, usually limited space is allowed for installing online sensors,

and it prohibits the use of high-resolution imaging sensors, resulting in low-quality wear debris images, as shown in Figure 2.13 [96]. Last but not least, for all imaging sensors, efficient processing of a large number of wear debris images in an industrial environment, from image enhancement and segmentation to particle classification, remains a critical challenge [47, 49].

Figure 2.13: A typical wear debris (generated in a four-ball wear tester) image captured by online imaging sensors [96].

2.2.3.2 The relationship between worn gear teeth and oil and wear debris features

It is commonly acknowledged that many oil and wear debris features have close relationships with gear wear, and they are the foundation of using OA and WDA for gear wear monitoring [25, 47, 80, 101]. For oil features, viscosity directly affects the oil film thickness and separation between two contact surfaces, which are closely related to the occurrence or rate of adhesive, abrasive, and fatigue wear [22, 29]. Water content, which often comes from condensation of ambient moisture and coolant leakage, is closely related to the oxidation of lubricants and tribocorrosion of gears [47]. Acidity, which is usually measured using total acid number (TAN), indicates the oxidation level of the oil, has a major impact on tribocorrosion of gears [47]. For wear debris features, the concentration of wear debris indicates the gear wear rate, and the combination of concentration and size is closely related to gear wear severity, as shown in Figure 2.14 [47].

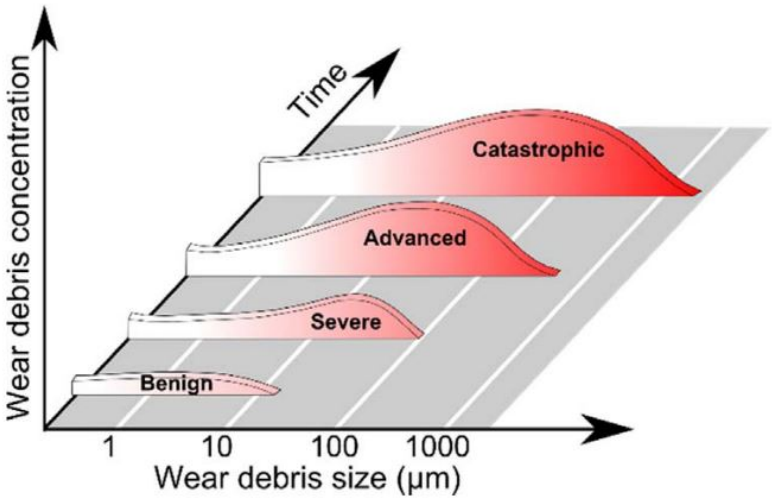


Figure 2.14: The empirical relationships between wear debris concentration, size, and wear severity [47]. The ranges of wear debris size can be machine- and operating-condition-specific.

As for wear mechanism assessment, the size, shape, and surface morphology of wear debris, which are closely related to their generation process, are crucial properties in WDA [25]. Based on the above three properties, wear debris are usually classified into six different types, namely, rubbing, cutting, laminar, fatigue, sphere, and severe sliding particles, and the types of wear debris can be used to identify the wear mechanism [25,102]. For instance, cutting particles, which are curled metal strips, usually indicate the occurrence of abrasive wear, and fatigue particles, which are chunky and have rough surfaces, usually indicate the occurrence of fatigue wear [102].

However, the relationship between worn gear teeth and wear debris features could be impacted by the modifications of wear debris, e.g., deformation or fracture caused by entrapment of wear particles in gear contacts after their detachment from gear teeth [10, 25]. The extent to which wear debris keep their original morphology or even quantity after their detachment has not been well understood [25]. Experimental tests [10, 12, 103] have shown that the particle concentration between gear contacts is significantly (ten to thousands of times) higher than that in bulk lubricating oil. Theoretical analyses and experimental tests [11, 12, 104] on particle entrapment showed that larger particles are

easier to be entrained into contacts than small ones. In general, entrapped particles may be compressed into platelets (ductile material) or broken into fragments (brittle material), as shown in Figure 2.15 [104].

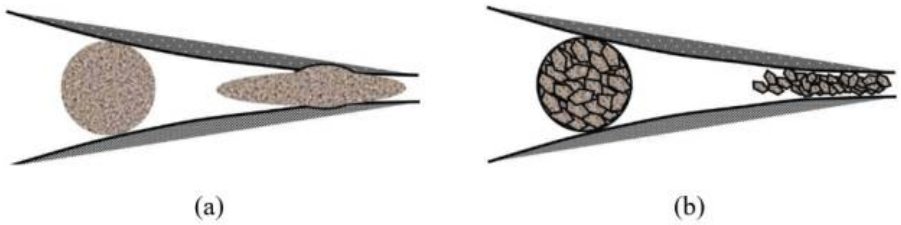


Figure 2.15: Modifications of entrapped particles in contacts [104]. (a) A ductile particle can be compressed into a platelet. (b) A brittle particle can be broken into fragments.

There are few studies [105–108] that investigated the relationship between features of worn gear teeth and those of wear debris in the same gear wear process, although characterisations on either wear debris [109–113] or worn gear teeth [64, 110, 114–116] are numerous. Reda et al. [105] conducted a series of sliding wear tests of steels where a cylindrical sleeve rotated against various fixed samples (cylinders, balls, and a flat plate) and examined worn surfaces and generated wear particles using optical microscopy and SEM. The worn surfaces and wear particles were classified into six wear regimens by the surface appearance of the worn surfaces and the size and colour of wear particles. The classification showed that wear particles became larger (from less than 15 µm to up to 150 µm) as the wear rate rose (from low to high) and worn surfaces were ploughed. Umeda et al. [106] conducted a set of pin-on-disc steel sliding tests and analysed optical microscopic images of worn surfaces and wear particles. Two textural features, namely, angular second moment and contrast, were calculated from the co-occurrence matrices of the images of worn surfaces and wear particles. By qualitative comparisons, the authors found evident differences between the features of wear particles and those of worn surfaces. Yuan et al. [107, 108] conducted rolling and sliding wear tests on a pin-on-disc test and a ball-disc tester, respectively, and measured surface roughness of wear particles and worn surfaces at different wear stages (e.g., running-in, steady, and severe wear). Results showed that the surface roughness (S_a) values of wear particles were generally smaller than those of worn surfaces and suggested

the modifications of wear particles after their generation. Moreover, the surface roughness of wear particles and worn surfaces shared a similar trend across different wear stages. To sum up, existing studies mainly based on pin-on-disc test rigs reveal some changes in the size and surface roughness of wear particles and their difference to those of worn surfaces. Study of the gear wear evolution, in particular, the relationship of wear particles and gear surfaces in a wear process is yet to be conducted.

In summary, the strengths and limitations of OA and WDA in gear wear monitoring are as follow.

- Many features of oil and wear debris have close relationships with gear wear, providing OA and WDA with advantages in gear wear monitoring. On the other hand, OA and WDA monitor indirect indications of gear wear, and their accuracy and reliability could be lower than direct monitoring methods. The morphology of captured wear particles could be significantly different from their original morphology due to the impact of particle entrapment in gear contacts. Further studies on their relationship in a wear process are needed.
- Offline OA and WDA are established techniques widely used to determine gear wear rate, severity, and mechanism, but often require expensive facilities and cause inevitable delays in gear wear monitoring.
- Online OA and WDA can provide gear wear rate and severity information with comparable accuracy to offline techniques, and most importantly, real-time monitoring capabilities. Current online WDA sensors cannot efficiently provide the morphology information of wear debris, and imaging sensors face challenges such as sensor improvement (e.g., improve the quality of wear debris images) and image processing (efficient processing of large batches of images).

2.3 Image-based gear wear assessment

As reviewed in section 2.2, images are commonly used data in gear wear monitoring and crucial data for detailed gear wear assessment. Specifically, images are the major data used in visual inspection and surface imaging, and images of worn gear teeth containing direct wear information are highly valuable for assessing gear wear mechanism and severity. For WDA, wear debris images are necessary data to determine the particle type and then infer the gear wear mechanism. Images of worn gear teeth and wear debris might look completely different, but their acquisition and analysis techniques often share many similarities. Therefore, this section reviews important image acquisition techniques (section 2.3.1) and conventional (section 2.3.2) and deep learning based (section 2.3.3) image analysis techniques for the assessment of gear wear mechanism and severity.

2.3.1 Image acquisition techniques

The process of image acquisition plays a significant role in image-based gear wear assessment as the quality of gear wear information contained in the images can vary greatly depending on the chosen imaging techniques [7,13]. Various types of microscopy and other imaging methods have been employed to meet different demands of image quality. For on-site gear wear analysis, portable devices such as regular digital cameras and hand-held microscopes are often seen. For laboratory gear wear analysis, optical microscopy [70,117], laser scanning confocal microscopy (LSCM) [26, 118], and scanning electron microscopy (SEM) [112,119] are frequently used.

Portable tools are often used for gear teeth imaging due to the tight space between gear teeth and the obstruction of the gearbox housing. Images captured using these tools, such as those in Figures 2.2, 2.4,2.7, and 2.9, are often used to show the overall appearance of worn gear teeth. Due to the lack of precise positioning and appropriate illumination, portable imaging tools are still mainly used for a rough or preliminary gear wear assessment.

Optical microscopy is one of the most popular imaging methods for gear wear analysis, and it is often low-cost and easy to use [7]. By changing different objective lens, an optical microscope can provide a broad range of magnification (e.g., five times to hundreds of times magnification). Optical microscopy is usually used to capture 2D images of gear wear samples(e.g., wear debris slides), but it is not suitable for the direct imaging of gear tooth flank due to the tight space between gear teeth, which also applies to other types of microscopy. As reviewed in section 2.2.1, this issue could be addressed by making replicas of gear teeth using appropriate replicating compounds, and in some cases, destructive measures such as tooth cutting are adopted for detailed examinations [43,119]. The quality of images captured by optical microscopy can be notably affected by the type and the adjustment of illumination [7]. For example, gear wear samples are usually non-transparent and therefore reflected light is often chosen. However, when the feature of interest is the shape of an object, such as tooth profile and edges of wear particles, transmitted light could provide a better contrast between the background and the object (see wear particle images in [77, 95]). For very reflective surfaces, normal reflected illumination might cause glare and the loss of clear surface features and reducing the intensity of illumination or using polarised illumination could be helpful [7].

SEM is another prevalent imaging method for gear wear analysis, and the sample surface is scanned by a high-energy electron beam to generate its image [7]. SEM has the features of high magnification and high spatial resolution, and it can be used to capture 3D surface information and examine the chemical composition of wear samples [120]. However, SEM is usually more expensive than the other two types of microscopes, and the preparation for some samples can be time-consuming [120]. It is usually used for very detailed surface characterisation of wear samples, and example applications in gear wear analysis can be found for cut-off gear tooth in [119] and wear particles in [112].

LSCM is usually used for 3D imaging of gear wear samples such as gear tooth moulds [121] and wear particles [122]. It utilises a pinhole to exclude out-of-focus light, leading to a small depth-of-field. By capturing images at different heights with a small and controlled step size, LSCM can construct the 3D shape of an object [7]. Compared to SEM, the cost

of LSCM is usually lower, and requires much fewer sample preparations.

2.3.2 Conventional image analysis for gear wear assessment

After image acquisition, raw images of gear wear samples usually require different levels of image processing so that meaningful wear information can be extracted and utilised for gear wear assessment. Human analysts are capable of analysing many raw images to assess gear wear states, but such tasks are usually very time-consuming and costly [111, 112, 122]. Therefore, various image processing techniques have been developed to assist human analysts for gear wear assessment. Conventional image analysis for wear assessment often consists of preprocessing raw wear images and extracting meaningful features from the images [61]. Then, the extracted features are used for gear wear assessment via human assessment or computer algorithms such as machine learning models.

The preprocessing of raw images is to improve their quality, and it usually involves reducing noises in images, enhancing their contrast, sharpening images, etc [61]. Both the inputs and outputs of this process are images. Image filtering is often used to reduce noise (e.g., median filter and Gaussian filter) and sharpen images (e.g., Laplacian filter) [61, 123]. Contrast stretching and histogram equalisation are often used for contrast enhancement [123].

Image segmentation is often employed to separate the objects of interest from the background [123]. For images of worn gear teeth, certain surface defects such as pits need to be separated from the tooth surface so that their size, shape, and other features can be extracted. For wear particle images, particle segmentation is a critical step for further analysis, especially for particle chains in ferrograph images [124]. Thresholding is a simple and common method for image segmentation, and different methods have been developed for automatic threshold selection, such as Otsu's method [125]. More difficult segmentation tasks such as separating particles from particle chains usually require more advanced techniques. Wang et al. [124] combined watershed transformation and an improved ant colony clustering algorithm to segment ferrograph images. The proposed method was

shown to effectively minimise the over-segmentation of large abnormal particles. Wu et al. [97] combined a marker-controlled watershed approach and multi-scale morphological operations to improve the segmentation reliability for online ferrograph images.

After image segmentation, various numerical features of the separated objects are extracted for their mathematical descriptions, and many studies have been devoted to this topic for gear wear assessment [111]. In order to obtain a comprehensive description of the separated wear-related objects, multiple types of features are usually extracted, and multi-variant statistics are output as a result. For wear particles, size features (e.g., diameter, perimeter, Weibull parameters of size distribution), shape features (e.g., aspect ratio, roundness, fractal dimension), surface topography features (e.g., S_a and S_q), and colours are commonly used in their characterisation [94, 111, 112, 126, 127]. Those features can also describe certain surface defects on worn gear teeth such as pits. Features for surface topography are of particular importance in classifying worn gear teeth and wear particles, but the development of such features has experienced slow progress due to the extreme complexity of those 3D surfaces [110]. Stachowiak et al. [110] pointed out that many features or methods to describe surface topography were scale-variant and those scale-invariant features obtained via fractal methods were not suitable for anisotropic surfaces. To address the problem, they proposed a modified partitioned iterated function system (PIFS) to encode 3D images of surfaces for the full description of their surface topography, but its application is not seen in other gear wear studies.

The final step of conventional image analysis is to process and interpret the extracted features to obtain results of gear wear assessment, which is often the classification of gear wear mechanism and wear severity levels. Many studies have adopted conventional machine learning methods to efficiently process the multi-variant features, such as support vector machine (SVM), expert system, and neural network. Peng and Kirk [109] extracted nine boundary- and surface topography-related features from LSCM images of wear particles, and input them into a fuzzy Kohonen neural network and a multi-layer perceptron (MLP) network for automatic particle classification. Later they also employed a grey system with fuzzy logic to achieve automatic particle classification with similar input data.

Stachowiak et al. [112] used a linear SVM to classify wear particles using three features extracted from SEM images, which were reduced from 44 features by principal component analysis (PCA) and a nonlinear Fisher reduction, and the SVM achieved an error rate of 3% among abrasive, adhesive and fatigue particles. Wu et al. [94] extracted two representative particle features from online ferrograph images, one of which described the concentration of wear particles while the other described the sizes of large particles. A method of support vector data description (SVDD) was employed for wear state identification. Gonzalez-Arias et al. [114] extracted histograms of oriented gradients (HOG) from SEM images of worn surfaces and compared three classifiers, i.e., Gaussian naive Bayes, decision tree, and random forest, to distinguish mild from severe abrasive wear.

Conventional image analysis has shown effectiveness in gear wear assessment, and many numerical features (e.g., size, aspect ratio, S_a) have been used in routine wear analysis. However, such processes can be very time-consuming and technically demanding. Specifically, a complete analysis procedure that often involves multiple steps of image processing, and for different images, the method of processing and related parameters usually require careful selection [97]. Then, the feature extraction process can be challenging (e.g., crafting surface topography features), and the selected features might not generally be suitable for other surfaces [128, 129].

2.3.3 Deep learning based image analysis for gear wear assessment

The fast development and successful applications of deep learning since around 2012 have shown its superiority in discovering intricate structures in large datasets over other methods [63]. One of its most successful applications is in image processing and analysis, covering tasks such as image segmentation [130], object detection [131], image classification [132, 133]. A key characteristic of deep learning is that little feature engineering is needed compared to conventional machine learning approaches, making it possible to process raw data using deep learning algorithms [63].

In recent years, the intriguing feature and successful applications of deep learning have

driven researchers to adopt deep learning based image analysis for gear wear assessment. Wang et al. [128] employed a deep convolutional neural network (CNN) to identify fatigue and severe sliding particles and achieved 85.7% identification accuracy for fatigue particles and 80% for severe sliding particles. Compared to previous studies [112,122] that required 3D surface information of particles and careful feature engineering, their study demonstrated that efficient particle classification can be realised using only 2D images of wear particles via a deep learning approach. Peng and Wang [129] proposed FECNN, a CNN for ferrography image classification, and they introduced 1D convolution and dropout to increase the learning robustness of the model. Compared to previous particle classification works, they achieved a high classification accuracy (93.89%) using a relatively small number of training samples. The extracted particle features by the deep learning model were visualised.

Apart from particle classification, the deep learning approach was also used to restore blurry particle images. Wu et al. [134] used a CNN with an autoencoder structure and a large kernel size to deblur wear particle images, and they generated training pairs (a sharp and a blurry image of the same particle) by capturing particle images with different focus degrees. Their sample results, including blurry inputs, restored outputs and their ground truth, are shown in Figure 2.16 [134]. The application of deep learning in processing worn gear teeth images is scarce, and it could be caused by the difficulty in obtaining sufficient training samples of worn gear teeth, as mentioned in section 2.2.2.

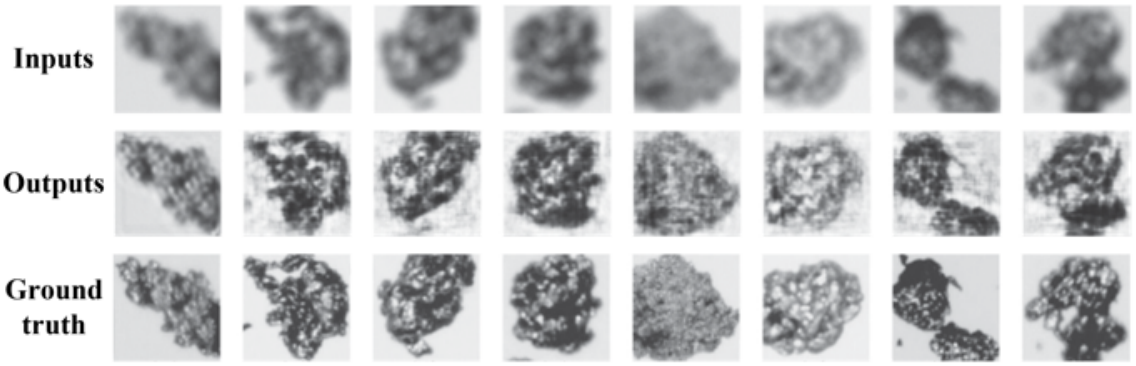


Figure 2.16: Wear particle image restoration using a deep CNN [134].

2.4 Gear wear prediction

Gear wear prediction is a critical task of gear prognosis that enables proactive maintenance actions to be taken to prevent gear fault from happening. Extensive research on gear wear prediction have been conducted, and physics-based methods and data-driven methods have been considered the two primary approaches [48, 135–137]. Physics-based methods attempt to establish explicit mathematical models to relate gear wear with the impacting physical parameters [136]. On the contrary, data-driven methods attempt to establish algorithms that implicitly model the gear wear dynamics based on CM data, without reference to physical laws [135]. In recent years, machine learning methods can be distinguished from statistical data-driven methods, due to their rapid development in recent years [138]. Therefore, this section reviews three main methods for gear wear prediction, namely physics-based methods (section 2.4.1), statistical data-driven methods (section 2.4.2), and machine learning methods (section 2.4.3), and also highlights the opportunities and challenges brought by deep learning.

2.4.1 Physics-based methods

Many physics-based models have been developed to relate gear wear rate (usually for adhesive, abrasive, and corrosive wear) or surface fatigue life of gears (for fatigue wear) to impact factors such as load and speed. Most of the models were built upon simple wear laws or relationships, and significant research efforts have been devoted to their improvements by considering more impacting factors.

The Archard's wear equation [139], or Archard's law, is one of the most frequently used wear equations for sliding contacts, which relates the wear rate, i.e., wear depth per unit sliding distance, to the contact pressure. A commonly used expression [140] is as

$$\frac{dh}{ds} = kP \tag{2.1}$$

where h is the wear depth (m), s is the sliding distance (m), k is the wear coefficient (Pa^{-1}), and P is the normal contact pressure (Pa). The effects of many other parameters, such as gear material, lubrication, and surface roughness of gear teeth, are incorporated in the wear coefficient k, which significantly simplifies the equation but limits its accuracy. The Archard's wear equation was originally proposed to estimate adhesive wear, but its flexibility provided by k allowed it to be adapted for abrasive wear and corrosive wear [141]. Based on Archard's wear equation, Wu and Cheng [142, 143], Flodin and Andersson [144,145], and Kahraman et al. [140,146–148] have improved the original model for gear wear prediction, with a focus on detailed modelling of the contact pressure, considering the varying contact geometry during gear meshing, elastohydrodynamic lubrication (EHL), modified tooth profile, etc. The models of Flodin and Andersson and Kahraman et al. were validated by experimental tests, and results showed their models achieved reasonable prediction accuracy for maximum wear depth. In the tests by Flodin [64], the wear coefficient was empirically chosen, and simulated results for wear distribution and maximum wear depth near the tooth root that experienced the most severe wear showed notable errors compared to experimental results. In comparison, Kahraman et al. [146] conducted a gear test first to determine the wear coefficient and used the k value in wear prediction for other tests with the same configuration (e.g., material, operating condition, and lubrication). The simulation results for maximum wear depth showed excellent agreement with measured values from tests. The comparison suggested that access to the appropriate wear coefficient is critical for the wear prediction models to work well. When the wear coefficient can be experimentally obtained, and the actual running conditions of gears are consistent and close to the experiments, the above models are usually suitable for wear prediction [15, 147]. However, it is not practical for the majority of industrial gearboxes mainly due to the complexity of gear wear, in which the impacting factors could be different from machine to machine, and could vary significantly through the process.

For fatigue wear prediction (e.g., macropitting and micropitting), gear surface fatigue lives that often refer to the number of running cycles before fatigue wear occurs are usually predicted instead of the wear rate. Extensive research on fatigue wear under rolling/sliding contacts have been conducted, especially for bearings, and there are also many gear fatigue

models [136]. Among the gear fatigue models, two groups can be roughly distinguished. The first group usually adopts multi-axial fatigue criteria that relate stress or strain to the fatigue life of material, while the second group usually adopts a fracture mechanics method to model crack initiation and propagation. The models [40, 42, 149–152] based on multi-axial fatigue criteria, such as the critical plane criterion and the Dang Van criterion, usually have detailed modelling of the stress/strain field under gear contacts. The state-of-the-art models [40,153] of this group usually has a mixed EHL model that considers the evolving tooth profile and surface roughness to improve the accuracy of stress simulation. Due to detailed modelling of localised stress variation, they have shown effectiveness in micropitting prediction [40, 150, 152]. Figure 2.17 [152] shows the micropitting prediction results in [152], compared to experimental results.

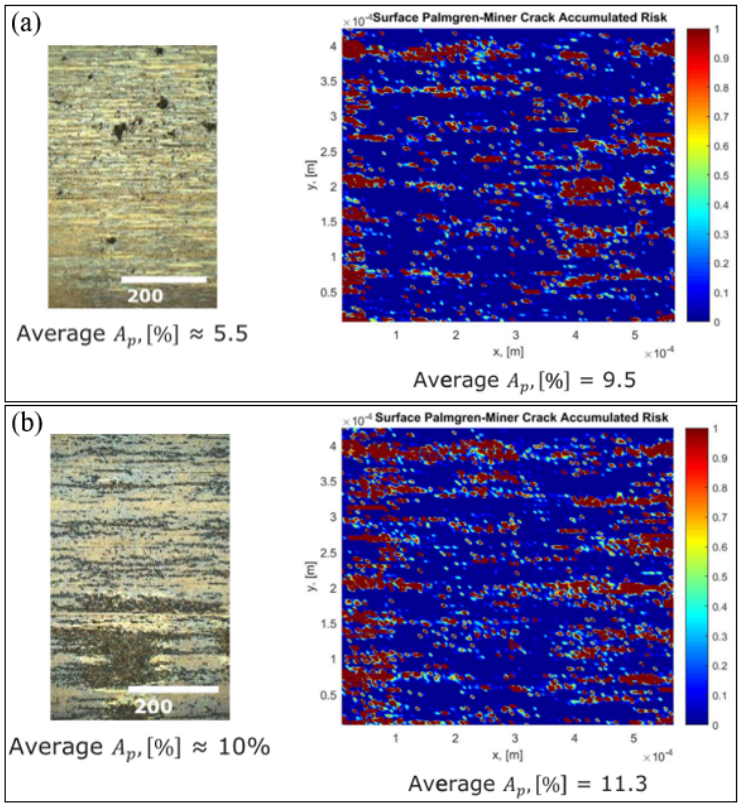


Figure 2.17: The comparison between experimental (left) and simulation (right) results of micropitting prediction by the model developed by Morales-Espejel et al. [152]. The pitted area, A_p, for different sliding/rolling ratio S: (a) $S = -0.15$, (b) $S = -0.3$, are indicated. Predicted pitted areas are in dark red.

The gear fatigue models based on fracture mechanics follow the mechanism of rolling/sliding contact fatigue and focus on the iterative simulation of crack growth [154]. A commonly used crack growth equation is Paris' law [155], or the Paris-Erdogan equation, which can be expressed as

$$\frac{da}{dN} = C(\Delta K)^m \tag{2.2}$$

where $\frac{da}{dN}$ is the crack growth rate (m/cycle), ΔK is the range of stress intensity factor (Pa $\mathrm{m}^{0.5}$), and m (unitless) and C ($\mathrm{Pa}^{-m}\mathrm{m}^{1-0.5m}$, note the difference between the unit m and the variable m) are material constants that need to be experimentally determined. With the crack growth equation and the stress field near the contact surface that is usually calculated using the finite element method (FEM), the crack growth towards the direction with the highest stress intensity during each running cycle can be simulated. With iterative updating of fatigue cracks and stress intensity, the formation of pits could be visualised, as shown in Figure 2.18 [38]. Compared to the first group of models, fracture mechanics models usually adopt simpler lubrication models (e.g., without EHL) and gear contact models (e.g., Hertzian contacts). The predicted fatigue lives of gears (macropitting) in [38, 154, 156, 157] have shown reasonable agreement with experimental results, and the shapes of pits were also well predicted, as shown in Figure 2.19 [158].

Many gear fatigue models in either group made an important hypothesis that a gear's service life ends when pitting initiates. It is largely true for ball bearings of which the rolling motion can be immediately disturbed when pits occur on the raceway, but not necessarily true for gears that undergo line contacts instead of point contacts. As mentioned in section 2.2.3, macropitting is considered severe when a significant portion of gear tooth is pitted and micropitting is less likely to cause severe fault than macropitting. Therefore, the propagation of pitting, i.e., the generation of multiple pits, should also be considered for gear fatigue prediction.

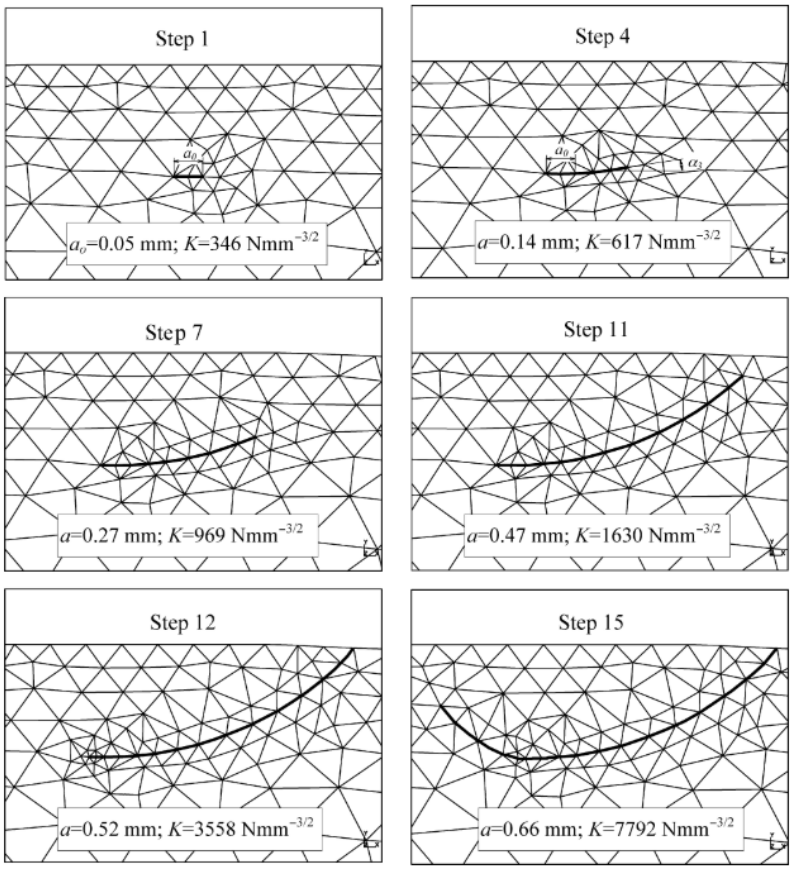


Figure 2.18: Simulation results of fatigue crack growth using FEM based on fracture mechanics [38].

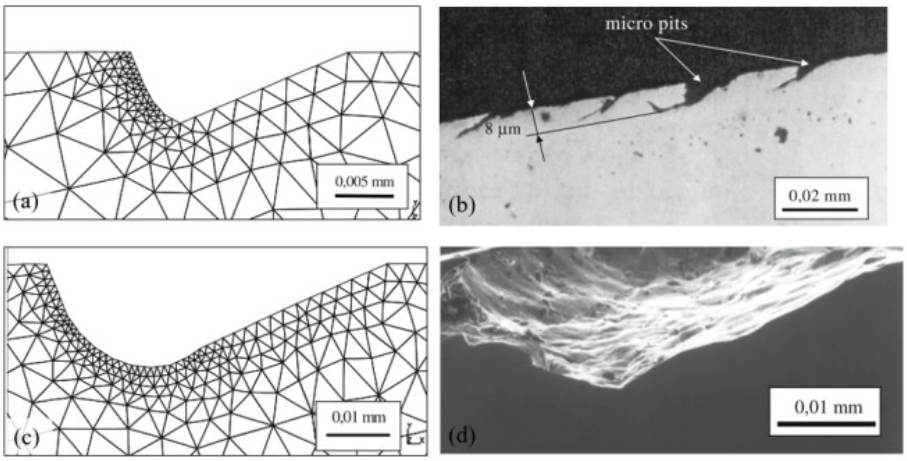


Figure 2.19: (a) A predicted micropit in comparison to (b) micropits generated experimentally; (c) A predicted macropit compared with (d) a macropit generated experimentally [158].

In summary, physics-based methods for gear wear prediction are available for common gear wear mechanisms, especially for fatigue wear, of which detailed modelling of lubricated contacts and crack initiation and propagation are achieved. The accuracy of physics-based models appear reasonable as suggested by their experimental tests with well-controlled operating conditions. Physics-based models are suitable for wear prediction when the models' required inputs (e.g., material constants, wear coefficients) can be obtained and the operating conditions remain consistent. In addition, physics-based models are based on physical laws or relationships that have been extensively studied, and their outputs usually have clear physical meaning, which is often desired for engineering applications. However, physics-based models have the following main limitations that restrict their wide applications.

- Current modelling capabilities of physics-based methods are insufficient to account for the complexity of gear wear that at least includes: (i) numerous impacting factors (e.g., material, heat treatment, geometric errors, surface roughness, lubrication, additives, and temperature), (ii) the variation of those factors through the wear process, and (iii) the co-existence of multiple wear mechanisms. Therefore, the accuracy and reliability of physics-based methods are often questioned in practical usage.
- Using physics-based methods requires the necessary understanding of the gear wear mechanism (expertise required). Moreover, collecting the input parameters for physical models is often costly.
- Using complex physical models for improved accuracy usually requires prohibitive computational resources, which makes them less cost-effective than other methods.

2.4.2 Statistical data-driven methods

Data-driven methods for gear wear prediction rely only on observed data related to gear wear, which can be event data or CM data [18]. Because event data that registers past gear faults related to wear are relatively scarce, CM data are primarily used in data-driven

methods. The fast development of sensor technology has made it possible to measure multiple types of information concerning gear wear in real time. The obtained CM data are very versatile, such as vibration, oil temperature, wear debris quantity, load, and speed. When CM data are collected, a statistical model is commonly employed to be fitted to the data in a probabilistic manner, and it is often called a statistical data-driven approach [48].

Si et al. [48] presented a comprehensive review of statistical data-driven methods for RUL estimation, and the statistical models were classified into two groups: models that use direct CM data, and models that use indirect CM data, as shown in Figure 2.20 [48]. For direct CM data, commonly used statistical models were further classified into regression-based, Wiener process, Gamma process, and Markovian-based models, according to how the machine degradation process is simplified. For indirect CM data, stochastic filtering-based, covariate-based hazard, and hidden Markov model and hidden semi-Markov model based methods were identified as the main methods. The bridging role of feature extraction between indirect and direct CM data was also elaborated. That is, when extracted features from indirect CM data can be used to characterise the machine condition, those statistical models for direct CM data can also be applied to the extracted features.

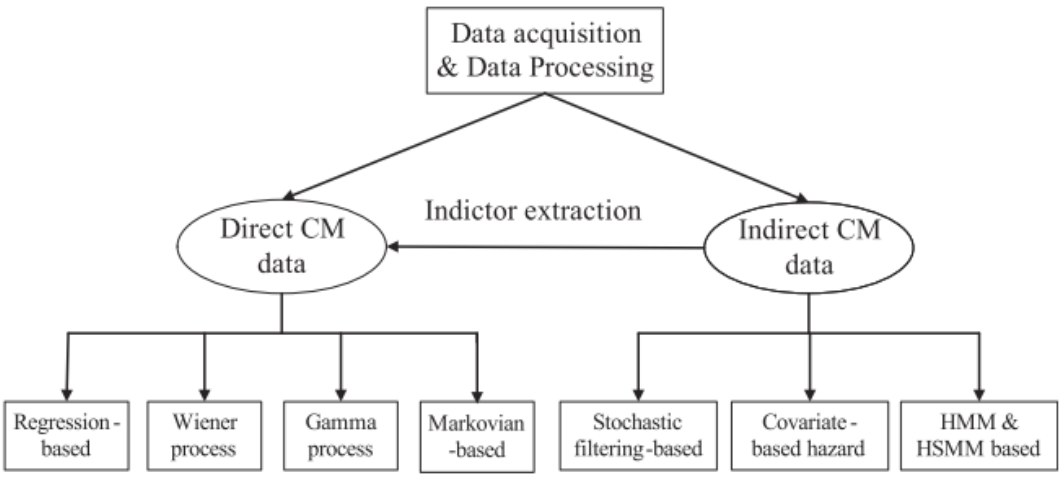


Figure 2.20: Classification of commonly used statistical data-driven methods for RUL estimation [48].

CM data such as vibration signals and wear debris features are commonly used for gear wear prediction. Feature extraction is usually applied to the raw data before statistical modelling, with rare exceptions where the raw data is used directly. Wang et al. [159] reviewed health indicators for prognosis and health management of bearings and gears in the literature, which are extracted from vibration signals mainly through signal processing and conventional machine learning techniques. However, many health indicators were engineered for fault detection and diagnosis and may not be effective for wear prediction [15, 137]. He et al. [160] applied particle filters to a vibration-based health indicator that were computed using six condition indicators (e.g., residual RMS, FM0) and oil debris mass for RUL prediction in a gear pitting test. Zaidi et al. [161] extracted time-frequency features from motor current using discrete wavelet transform (DWT) and employed hidden Markov models (HMMs) to predict gear failures of automotive starter motors. Kocijan and Tank [162] applied a Gaussian process model to the time series of a spectral feature of vibration signals to predict gear wear, and test results showed that the model provided warnings of critical wear 15 to 20 hours in advance. However, the relationship between the selected feature and actual gear wear was not investigated.

Statistical data-driven methods using wear debris features are limited in the literature, and most existing studies usually attempt to establish physical models describing the variation of wear debris concentration in a gear wear process. Feng et al. [8] modelled the wear debris concentration based on Kragelsky's model and indicated that the evolution of an index of particle coverage area obtained using an OLVF (see section 2.2.3) was aligned with the simulation results. However, a quantitative criteria to determine if gears enter the severe wear stage was absent, and the generalisation capability of the proposed model was not demonstrated through different tests. Zhu et al. [163] used particle filters on the viscosity and dielectric constant of the lubricating oil to predict the RUL of wind turbine gearboxes, and the authors derived physical models to relate the two oil properties to wear debris concentration. However, the relationship between oil properties and actual gear wear has not been well understood, and the accuracy of this method could be impacted by many factors that were not considered in the laboratory test. For example, the oil viscosity can be notably affected by oil oxidation, external contamination, additive depletion, etc. [47],

which was considered only to be affected by temperature and wear particles. Hong et al. [164] proposed using the peak rate of wear debris generation as the indicator of the end of RUL based on a physical model that related wear particle generation to surface roughness evolution, and empirical data from wind turbine gearboxes were used to validate the method. However, it can be observed from the empirical data that the rate of wear debris generation varied significantly throughout the wear process, and there could be multiple peaks of wear particle generation, bringing challenges to use the indicator in practice.

Current statistical data-driven methods for gear wear prediction based on feature extraction, or feature engineering, face the following main challenges.

- Handcrafted features for gear wear prediction have limited generalisation capability. There is still no health indicator or a set of health indicators that have been shown effectiveness for different gear wear mechanisms under different operating conditions [138].
- The difficulty of feature engineering is expected to increase significantly as multi-dimensional CM data becomes increasingly available [48].
- To extract useful features and fit appropriate statistical models to the features require considerable expertise in signal process and statistics, which could limit their popularity in industrial applications.

2.4.3 Machine learning methods

Machine learning methods are being increasingly used for gear wear prediction and other tasks in gear prognosis, and they can be distinguished from the two methods mentioned above due to their unique features. Similar to statistical data-driven methods, machine learning methods for gear wear prediction are also driven by CM data of gears, meaning that little knowledge of the physics of the gear wear process is required. However, compared to statistical modelling in which the relationship between inputs (history CM

data in this case) and outputs (future CM data) is explicitly approximated by a statistical model such as a linear regression model, machine learning implicitly models the process by learning from observed data without presuming the relationship between inputs and outputs [165]. Thus, the advantage that little expertise in complex statistical modelling is required, and the increasing availability of CM data have given rise to machine learning methods in gear wear prediction. Among various machine learning techniques, deep learning has demonstrated its superiority over conventional approaches through numerous applications. Therefore, this subsection reviews conventional machine learning methods for gear wear prediction (section 2.4.3.1) and new potentials brought by deep learning approaches (section 2.4.3.2).

2.4.3.1 Conventional machine learning approaches

Gear wear prediction via conventional machine learning approaches is usually realised through the following steps: gear wear data acquisition and preprocessing, feature extraction, model selection, training, and testing (predicting gear wear). There are many conventional machine learning models available, and examples of those for time-series prediction include support vector machine (SVM), random forest, and neural networks [166]. Wang et al. [167] compared the performance of recurrent neural networks (RNNs) and neuro-fuzzy systems in multiple gear fault prediction tests (including severe abrasive wear, scuffing, and pitting). Vibration-based features were extracted and used for model training and testing. Results showed that the neuro-fuzzy systems generally outperformed RNNs in one-step-ahead prediction of the evolution of the features. However, actual gear wear conditions were not examined to validate the correlation between the extracted features and gear wear. Kundu et al. [168] employed a random forest regression methodology to predict gear pitting propagation using correlation coefficient-based features extracted from residual vibration signals and achieved an average RUL prediction accuracy of 6.7% in five gear pitting tests when fusing vibration signals from sensors installed at different positions. Nevertheless, the prediction performance of the proposed feature was not compared with other commonly used vibration-based features such as RMS and FM4.

2.4.3.2 Deep learning approaches

Compared to conventional machine learning approaches, deep learning approaches do not require the feature extraction step and make the implementation of machine learning easier for engineers without expertise in signal processing and feature selection. Deep learning techniques have been successfully used in prediction tasks such as stock price prediction [169], traffic prediction [170], precipitation forecast [171], and video prediction [172, 173]. The development of transfer learning and other deep learning techniques also provide promising solutions to improve the generalisation capability of gear wear prediction models for different machines under different operating conditions [138].

Currently, most deep learning applications in gear condition monitoring focus on gear fault detection and diagnosis, and research on gear wear prediction are limited. Xiang et al. [174] and Yan et al. [175] employed modified long short-term memory (LSTM) networks to process extracted features from vibration signals to predict RUL of gears. LSTM networks are famous for learning long-term dependencies in sequential inputs and have shown better performance than traditional RNNs [176]. Experimental results showed that the proposed models could predict the RUL of gears with reasonable accuracy when most data points from the late stage of the tests were given. However, complicated feature extraction procedures from raw vibration signals were needed to relieve the impact of background noise. The lack of separate gear tests (e.g., under different operating conditions) to validate the generalisation capability of the models could also be an important limitation. Lo et al. [177] trained a 1-D CNN (another commonly used neural network to process sequential data) on raw vibration data to estimate gear wear by comparing the test vibration signal and normal vibration signals (from the normal wear stage) and abnormal signals (from the wear-out stage) in the feature space. The proposed model was validated using data from a gear wear test and a benchmark bearing dataset, and results showed a low estimation error (10%) for wear depth. Nevertheless, the target data, i.e., vibration signals and the wear depth, at the end of a gear's service life are necessary for the model to work, which are impossible to obtain for in-service gears. The authors also indicated that there was significant variability across different gears in terms of their operating conditions and

service lives, making it difficult to transfer the target data from test gears to others.

Although little feature engineering is needed, gear wear prediction using deep learning approaches still faces challenges for its further development. An important issue is the data quality for training deep learning models, which represents the degree of correlations between CM data and actual gear wear [17], since the models only rely on CM data without any help from physical relationships. The learning difficulty of deep neural networks could be lowered significantly by using direct CM data instead of indirect CM data. For example, gear vibration signals can be easily impacted by background noise, inherent manufacturing errors, varying operating conditions, etc. [5, 15], and wear debris data that have much closer relationships with gear wear can still suffer from modifications by gear contacts and measurement errors [10, 49].

2.5 Summary

This chapter has reviewed common gear wear mechanisms, severity levels, and their surface features, and main methods for gear wear monitoring and prediction. Visual inspection and surface imaging, and OA and WDA have shown strong capabilities in gear wear monitoring and their strengths and weaknesses were examined. Both methods utilise images as key inputs, and thus, common techniques for image acquisition and conventional and deep learning based image analysis techniques for gear wear monitoring have been reviewed. For gear wear prediction, three main methods, namely, physics-based methods, statistical data-driven methods, and machine learning methods, were reviewed, and the potentials and challenges of using deep learning approaches for gear wear prediction were presented. In summary, the following research gaps in existing research on gear wear monitoring and prediction can be identified through this literature review.

1. A method to obtain direct and comprehensive gear wear information that does not require gearbox disassembly is not available. The advantage of visual inspection and surface imaging in accessing direct wear information is largely forsaken for gear

applications due to this reason.

2. Limited research has been done on the morphological relationship between worn gear surfaces and wear debris in the same wear process. Because wear debris may undergo deformation or fracture due to their entrapment between gear contacts, this relationship needs to be experimentally examined in a gearbox setup for a better understanding of the gear wear process and informed use of WDA.

3. A method to automate the assessment of gear wear mechanism and severity based on gear surface appearance has not been seen in the literature. To date, gear wear assessment still heavily relies on human experts and/or processed information through qualitative and quantitative image analysis, which have various drawbacks such as only extracting limited surface information and time-consuming. Development of an automated method could help address issues of low turnaround, high cost and subjectivity associated with manual assessment.

4. Little research has been done on the potential of gear wear prediction using gear tooth images, although deep learning has been successfully applied in similar prediction tasks (e.g., video prediction). Gear tooth images are direct CM data that are able to capture actual gear wear with high accuracy and reliability, from which advanced deep learning model could learn gear wear dynamics.

Chapter 3

Methodology

This research aims to develop advanced techniques for gear wear monitoring and prediction using direct gear wear data. In the previous chapter, existing methods and techniques for gear wear monitoring and prediction have been reviewed, and four gaps in the research area have been identified. Based on the gaps, the four objectives of this research are restated here.

1. Currently, direct and detailed imaging of gear teeth often requires gearbox disassembly, making the collection and utilisation of direct gear wear information notably difficult. Therefore, the first objective of this research is **to develop a method to lower the difficulty of obtaining direct and comprehensive gear wear information to the extent that gearbox disassembly is not required.**

2. Due to the difficulty in collecting wear evolution information in a wear process as mentioned above, little research has been done on the morphological relationship between worn gear surfaces and wear debris in the same wear process, which is vital for proper use of wear debris analysis for gear wear monitoring. Therefore, the second objective of this research is **to experimentally investigate the qualitative and quantitative relationship between worn gear surfaces and wear debris in the evolution of a gear wear process.**

3. The assessment of gear wear mechanism and severity still heavily relies on trained analysts, which is often associated with a high cost and low efficiency. Therefore, the third objective of this research is **to develop a system that automatically assesses gear wear mechanism and severity.**

4. Gear tooth images containing direct and comprehensive wear information have not been used for gear wear prediction, despite images being increasingly used for similar prediction tasks. Therefore, this research's fourth and last objective of this research is **to develop an image-based gear wear prediction model using a deep learning approach.**

This chapter outlines the methodology adopted in this research to achieve the above objectives. Figure 3.1 illustrates the overall structure of the methodology and the connections between the four technical chapters that correspond to the four objectives (Chapters 4 to 7). Firstly, gear wear tests were conducted to generate the wear data used in this research. Then, to achieve Objective 1, replication compounds were selected to replicate worn gear teeth to obtain gear teeth moulds without gearbox disassembly, and a wide variety of direct wear information was extracted from gear mould images via image acquisition and image analysis. To achieve Objective 2, gear mould images and wear debris images, which were obtained by imaging the wear debris slides made from collected oil samples, were analysed and compared to investigate the relationship between worn gear surfaces and wear debris in the same gear test. Then, to achieve Objective 3, deep learning models (CNNs) were trained on gear mould images and wear debris images to automate the process of gear wear mechanism and severity assessment. Lastly, to achieve Objective 4, a deep learning model (GAN) was trained on time series of mould images to learn the gear wear dynamics and predict the evolution of gear wear.

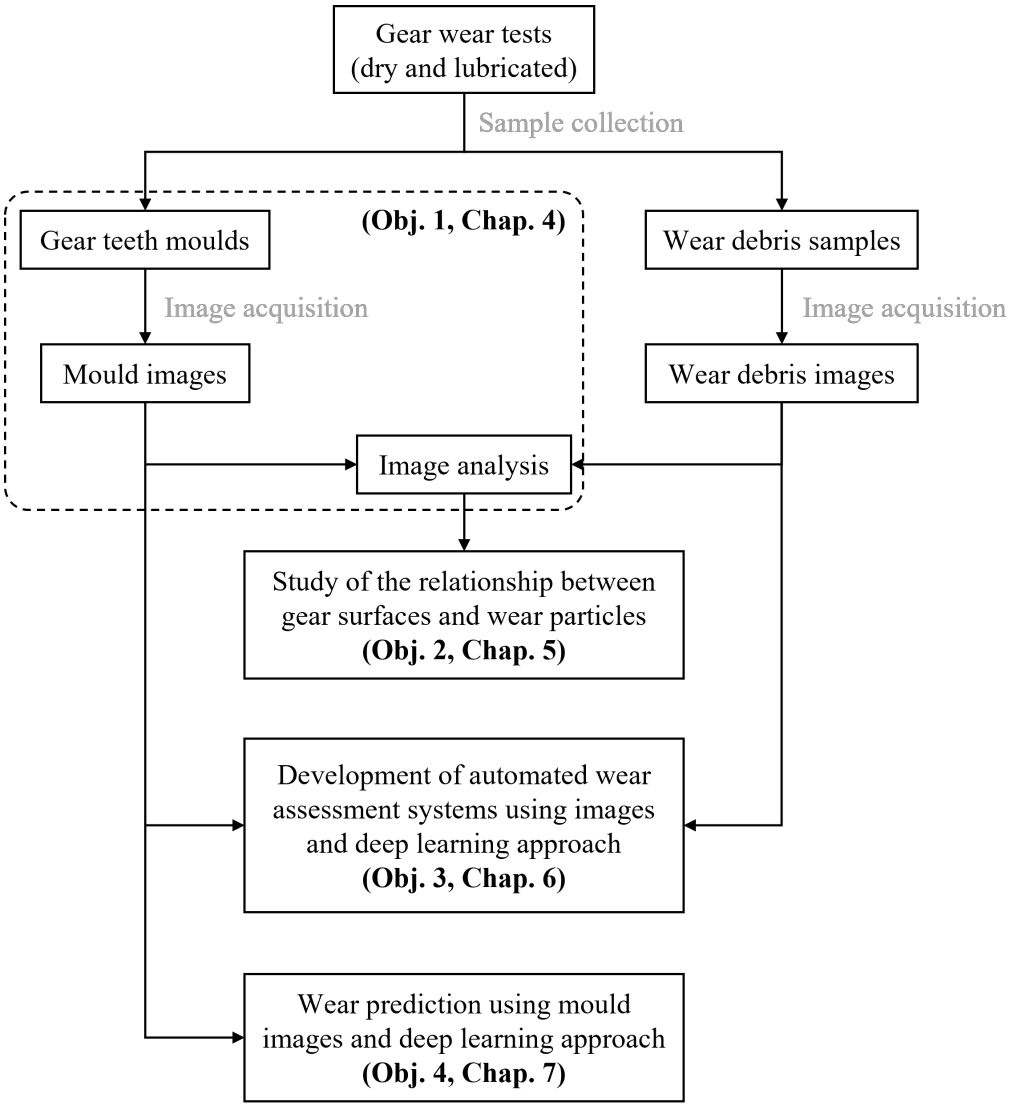


Figure 3.1: The overall structure of the methodology of this research.

In this chapter, section 3.1 details the experimental facility and gear wear tests, and section 3.2 introduces the methods to collect wear samples in the tests, followed by section 3.3 that overviews the image acquisition process and image analysis techniques. Image-based gear wear assessment and wear prediction using deep learning approaches are briefly explained in sections 3.4 and 3.5, respectively. The summary of this chapter is presented in section 3.6. It is worth noting that Chapters 4, 5, 6, and 7 have a dedicated section that provides details of the methodology to achieve their corresponding objective.

3.1 Experimental facility and gear wear tests

Gear wear tests, i.e., a long-running lubricated test and a relatively short dry test, were conducted on a single-stage spur gearbox test rig, as shown in Figure 3.2. The test rig consists of an ABB three-phase induction motor that provides input power (nominal power: 4 kW) with a variable speed, a single-stage spur gearbox, and a magnetic particle brake that provides a variable load.

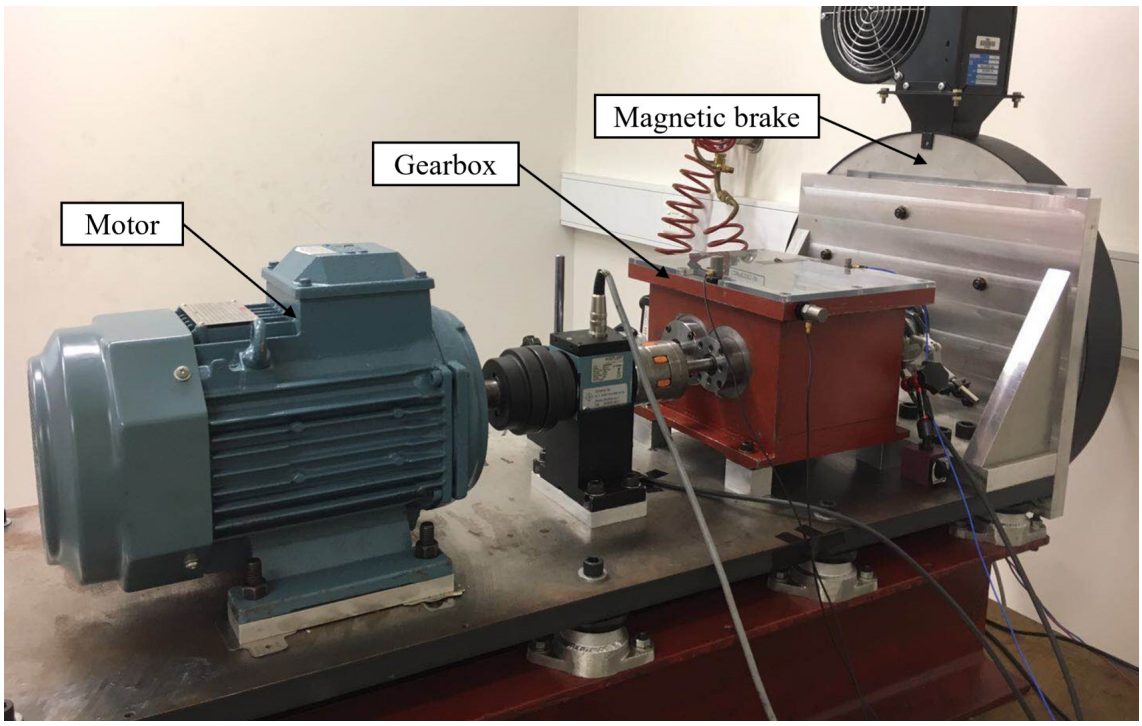


Figure 3.2: The single-stage spur gearbox test rig.

The pair of spur gears used in the lubricated test, as shown in Figure 3.3, consists of a driving gear with 19 teeth and a driven gear with 52 teeth. Another pair of gears with the same configuration and material was used in the dry test. The gears used in the two tests were manufactured by KHK, and their main technical specifications were summarised in Table 3.1. The elemental composition of the gear material is shown in Table 3.2.

Figure 3.3: The pair of spur gears used in the lubricated test. The left is the driven gear and the right is the driving gear. The yellow marked teeth are the ones for moulds making during the test. The dry test used another gear pair with the same configuration and material.

Table 3.1: Main technical specifications of the spur gears used in the two gear wear tests.

Parameters	Value
Number of teeth	19 (driving gear) / 52 (driven gear)
Module	2 mm
Face width	20 mm
Pressure angle	20°
Material	Normalised JIS S45C (equivalent to AISI 1045)
Tooth hardness	$\sim$ 190 HB

Table 3.2: Nominal elemental composition of JIS SC45 steel.

Element	**Weight %**
Carbon	0.42 - 0.48
Manganese	0.6-0.9
Phosphorus	0.03
Sulphur	0.035
Silicon	0.15 - 0.35
Iron	Balance

The lubricated test was designed to study a gear fatigue process by collecting and analysing gear teeth moulds and wear debris samples. Before the test, the two gears were manually roughened ($S_a = 3.2$ µm, measured using a laser scanning microscope with a 20x objective lens) to accelerate the initiation of fatigue pitting. This procedure was based on the relationship between surface roughness and surface fatigue life in rolling-sliding contacts [151, 178–180]. The gears were lubricated using splash lubrication in the test, using Shell mineral oil with a kinematic viscosity of 146 m^2s^{-1} at 40°C and 14.7 m^2s^{-1} at 100°C. The torque of the input shaft of the gearbox was kept constant (20 Nm) through the test by the magnetic brake, while its rotational speed was adjusted to avoid the overheating of the magnetic brake. The specific input speeds were: 20 Hz from 0 to 0.57 million cycles, 10 Hz from 0.57 to 0.86 and 2.5 to 3.25 million cycles, and 15 Hz from 0.86 to 2.5 million cycles. In total, the lubricated test ran for around 3.25 million cycles (68 hours), and the driving gear was eventually severely worn by macropitting and abrasive wear. During the test, the test rig was stopped 40 times to apply the surface replication method to obtain gear teeth moulds without disassembling the gearbox. The time interval between each stoppage of the gearbox was roughly 80 thousand cycles, and shorter intervals were used during the beginning and closing stage of the test to capture the rapid wear evolution on gear teeth.

The dry test was designed to study a gear scuffing process by collecting and analysing gear teeth moulds. A high wear rate was expected [27] as no lubrication would be used. Due to this reason, the two gears were not manually roughened, and a lower input torque (5 Nm) was selected so that the duration of the test is reasonably long enough for collecting multiple batches of gear teeth moulds to study the wear evolution. A constant input speed (10 Hz) was used throughout the dry test. In total, the dry test ran for 54.6 thousand cycles (around 4 hours), and both of the two gears were severely worn by scuffing at the end of the test. During the test, the test rig was stopped 11 times for moulds making. The time interval between each stoppage of the gearbox was roughly 6 thousand cycles, and shorter intervals were used during the beginning stage of the test to capture the rapid wear evolution on gear teeth.

3.2 The sampling of gear wear evolution

in order to directly study the evolution of gear wear, gear teeth moulds, which are direct records of gear wear, and oil samples with wear debris, which are by-products of gear wear, were collected when the gear test rig was stopped during the tests. Techniques and procedures used to collect these samples are described below.

3.2.1 Development of a moulding method to collect gear teeth moulds

A moulding method, or surface replication, was developed to acquire direct records of gear wear, i.e., gear teeth moulds, during gear wear tests without gearbox disassembly. There are many types of surface replication compounds available, and several key considerations such as replication resolution and deformation were incorporated in the selection process. Microset 101 [181] (thixotropic silicone polymers) was finally chosen as the replication compounds in this research due to its high-resolution, fast curing, low shrinkage, low thermal expansion, and ability to be used on vertical surfaces. Its replication quality was tested using a resolution target and a surface comparator in terms of the replication resolution, accuracy, and profile reproduction. More details about the development of the moulding method can be found in Chapter 4. Table 3.3 shows key properties of the replication material.

Table 3.3: Key properties of the surface replication compounds (Microset 101) [181].

Property	Value
Resolution	0.1 µm
Working life	3 min at 25°C
Curing time	20 min at 25°C
Usable temperature	-10°C to 180°C
Cured hardness	28 - 30 Shore A
Shrinkage	$< 1\%$
Linear expansion coefficient	2.8×10^{-6} °C^{-1}

Before each test, six gear teeth (three on the driving gear and three on the driven gear) were selected and marked for moulds making, as shown in Figure 3.3, where the selected

teeth were marked in yellow. As mentioned above, the test rig was stopped multiple times during the two tests, and at those times, the lid of the gearbox was taken off for moulds making. Before applying the replication compounds, the marked gear teeth needed to be carefully cleaned and any oil on the teeth must be absorbed for the proper curing of the replication compounds. Then, the replication compounds were injected using a dispenser with an extended mixing nozzle (see Figure 3.4) to cover the marked worn tooth flanks and fill in the valley between gear teeth, as shown in Figure 3.5. After the curing time of 20 minutes, a batch of six gear teeth moulds was taken off and stored separately. In total, 41 batches of moulds (including a batch made before the test) were obtained in the lubricated test, and 12 batches (also including a batch before the test) of moulds were obtained in the dry test.

Figure 3.4: The replication compounds dispenser used in this research [181].

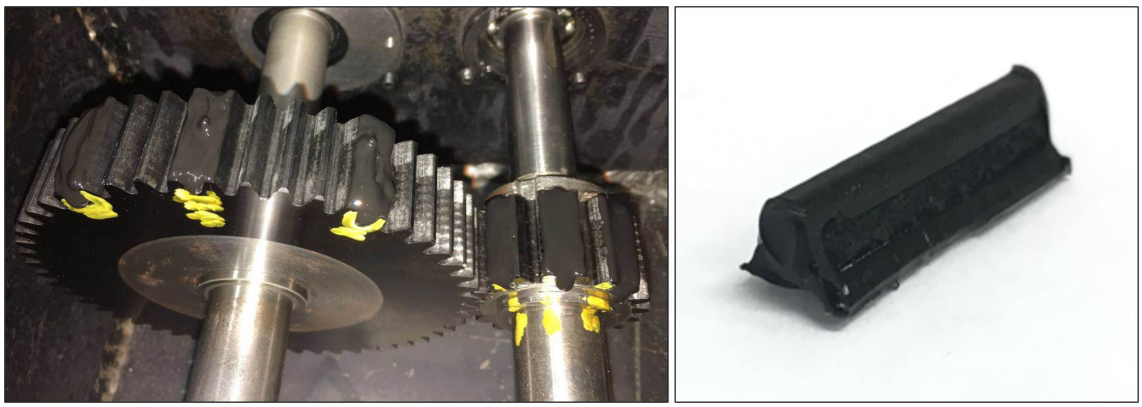


Figure 3.5: Curing moulds on gear teeth in the dry test (left) and a cured gear tooth mould (right).

3.2.2 Collection of wear debris slides using the filtergram method

Wear debris is by-product of gear wear, and it also carries various important gear wear information. Therefore, wear debris slides were acquired in the lubricated test via oil sampling and the filtergram method. Oil sampling was carried out just before the test rig was stopped for moulds making, and the sampling location was 1 cm below the oil surface and underneath the gear meshing region. A guiding channel was specially designed and 3D printed to ensure the accuracy and consistency of the oil sampling position, as illustrated in Figure 3.6. In total, 40 oil samples were obtained in the lubricated test, and the volume of each oil sample was around 20 ml. It is worth noting that an equal amount of clean oil was added to the gearbox after each sampling.

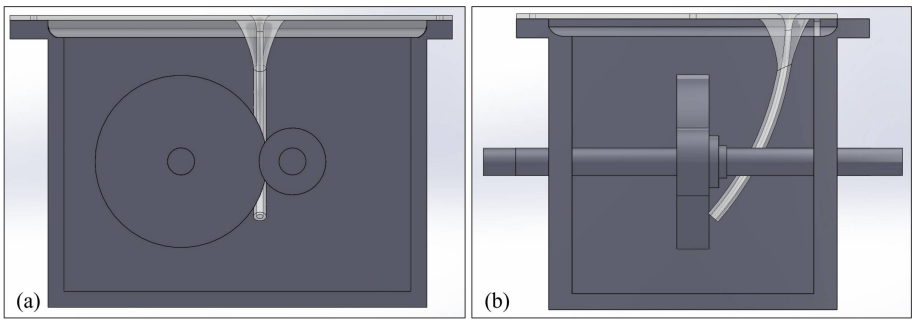


Figure 3.6: The design of the guiding channel (in light grey) installed on the gearbox (in dark grey). (a) Front sectional view. (b) Right sectional view.

Then, the oil samples were processed using the filtergram method [182] to make wear debris slides. The method involved separating wear debris from oil using membrane filters and fixing the filters on clean glass slides. The membrane filter used in this research had a pore size of 3 µm, meaning that only wear debris larger than 3 µm could be captured. For each wear debris slide, 1 ml of the oil sample was extracted from the sampling bottle and used in slide making. Before oil extraction, the sampling bottle was sonicated to ensure an even distribution of wear debris in the oil, and a low-power and short (around 10 mins) sonication was chosen to avoid violent collisions of wear debris [122, 182]. The membrane filter on the glass slide was cleared to enhance the contrast between particles and their background during imaging.

3.3 Image acquisition and analysis

In order to obtain a wide variety of wear information, two types of microscopy, i.e., optical microscopy and laser scanning confocal microscopy (LSCM), were selected to capture 2D and 3D images of collected gear wear samples (i.e., gear teeth moulds and wear debris slides), respectively. To capture wear features at the micro scale, an Olympus BX53 optical microscope [183] with a DP72 digital camera [184] and a Keyence VK-X200 3D laser microscope [185] were selected to be the suitable imaging facilities. Then, a series of image processing and analysis techniques were employed to extract various gear wear information for gear wear monitoring. Figure 3.7 shows the different image acquisition and analysis techniques chosen in this research to obtain different gear wear information.

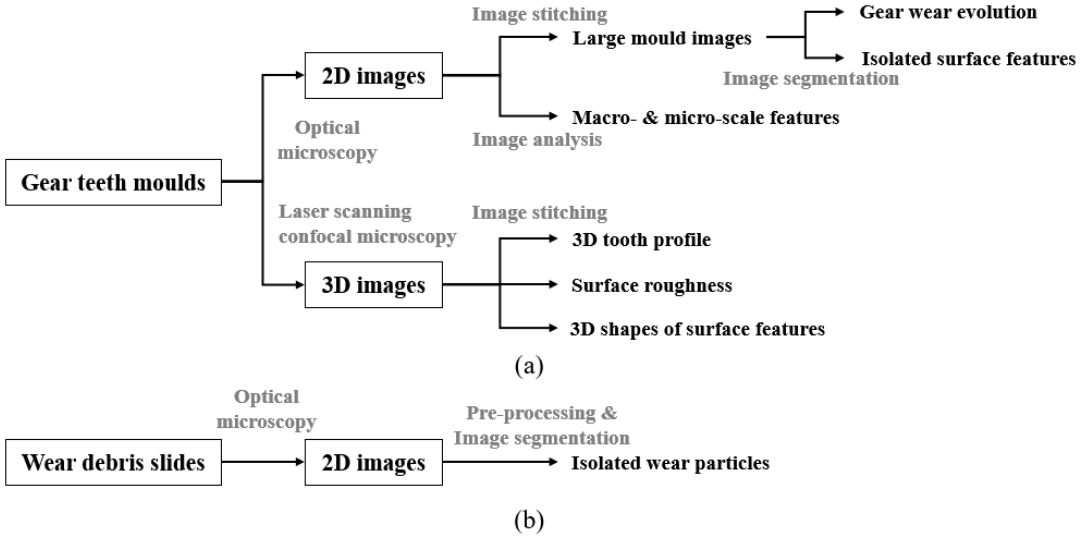


Figure 3.7: Image acquisition and analysis for (a) gear teeth moulds and (b) wear debris slides.

Gear teeth moulds contain microscopic wear features and macroscopic appearances on worn gear teeth, and thus different objective lenses were employed to cover the broad size range. For example, a 5x objective lens of the optical microscope was used to capture the macroscopic appearance of gear teeth, while a 20x lens was used to capture microscopic wear features such as micropits and surface cracks. A mould holder was specially designed and 3D printed to properly fix the moulds. For the 2D images captured using the 5x

objective lens, image stitching was necessary to generate large mould images covering the entire addendum or dedendum of a tooth flank, and the gear wear evolution could be visualised in great detail through time series of the large mould images in a wear process. To quantitatively describe important surface features such as macropits, the features need to be separated from their background (i.e., the tooth surface), and advanced image segmentation techniques were employed for this purpose. Using the 3D images, the height information was extracted to calculate various 3D surface roughness parameters such as S_a and S_q and carry out 3D characterisation of surface features such as measuring pitted depth. Moreover, 3D tooth profiles could also be obtained by stitching multiple 3D images along the profile, and they allowed the calculation of wear depths by subtracting a worn tooth profile from the original one.

For wear debris slides where wear debris from a few microns to about a hundred microns were dispersed, a 20x objective lens was selected for the optical microscope to capture 2D wear debris images. Key numerical features of wear debris, such as concentration, size, aspect ratio, and roundness, can be extracted using the 2D images to be used in comparison with those of worn gear surfaces. It is worth noting that the concentration of wear debris in this research is defined as the percentage of area covered by wear particles within the total area of captured images [95]. 3D images of wear debris were not incorporated due to their limited sample size. The extraction process started from preprocessing the raw 2D images to improve their overall quality via noise reduction and contrast enhancement. Then, image segmentation was used to separate wear debris from their background, and various numerical features were selected and calculated for proper particle characterisation. More details about the feature extraction step can be found in Chapter 5.

The extraction of key features of worn gear surfaces and wear debris from images allowed a detailed investigation of their relationship in the same wear process. Both qualitative and quantitative comparisons were made between worn gear teeth and wear debris. Qualitatively, the evolution of the wear mechanisms and their severity on gear teeth and that of the types and sizes of wear debris were compared using time series of mould images and wear debris images in the lubricated test. Quantitatively, the evolution of crucial

indicators of gear wear severity, including pitted area and wear debris concentration, were compared. For other important numerical features of macropits and wear debris, detailed comparisons regarding their data distributions, evolution, and principal components were designed and conducted. Further information can be found in section 5.3.

3.4 Image-based wear assessment using deep learning

Deep learning models (CNNs) were developed to automate the identification of gear wear mechanisms and severities using mould images and wear debris images. A deep learning approach was adopted due to its strong capabilities in learning intricate patterns in large datasets over other approaches [63], and CNNs were chosen due to their numerous successful applications in image analysis [132]. As illustrated in Figure 3.8, automated gear wear assessment was achieved via the following steps.

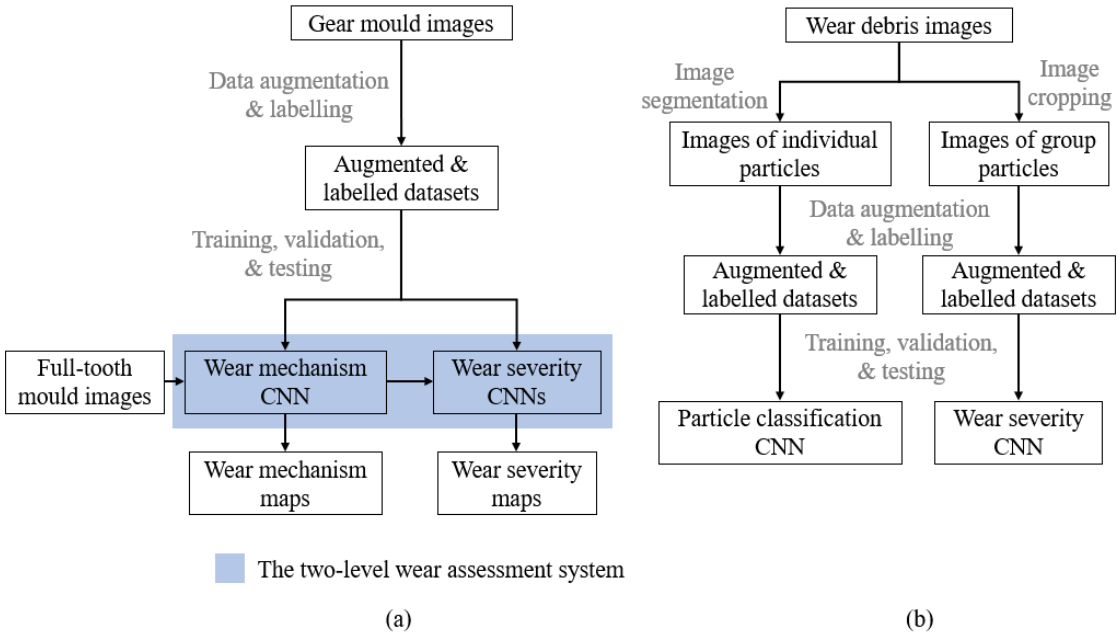


Figure 3.8: Image-based gear wear assessment using deep learning. (a) Wear assessment CNNs based on gear mould images. (b) Wear assessment CNNs based on wear debris images.

1. Preparation of gear wear images. Currently, open-source image datasets of gear wear that can be used to train deep learning models for gear wear assessment are not available. Therefore such datasets were purpose-built in this research. Large mould images that directly recorded gear wear evolution in the tests and wear debris images that directly recorded the by-products of gear wear in the lubricated test were suitable data to train CNNs for automated gear wear assessment. Wear debris images were converted to two types of images: images of individual particles for particle classification by image segmentation, and images of group particles for wear severity assessment by image cropping. This procedure was necessary because the CNN built for particle classification required that each input image contained only one wear particle while the other CNN built for severity assessment did not have this requirement. More details about data preparation can be found in Chapter 6.

2. Data augmentation and labelling. The number of acquired raw images (both mould images and wear debris images) were insufficient for training deep CNNs, and thus different strategies of data augmentation were designed to increase their numbers (see Chapter 6). Then, the labelling process in which the images were manually classified into different wear mechanisms or particle types and different severity levels, was conducted following established guidelines (e.g., gear standards) to ensure its correctness and consistency. In the end, more than one hundred thousand labelled mould images were generated covering four wear mechanism categories: normal, abrasive wear, scuffing, and pitting, and three severity levels for each of the latter three wear mechanisms: mild, moderate, and severe. For wear debris images, more than ten thousand labelled images of individual particles were generated covering four particle types: rubbing, laminar, fatigue, and cutting particles, and nearly ten thousand labelled images of group particles were obtained covering three severity levels: mild, moderate, and severe.

3. Construction, training, validation and testing of the CNNs. The CNNs used to automate gear wear assessment were constructed based on a widely recognised architecture, and its six variants were developed for different classification tasks (one

wear mechanism CNN and three wear severity CNNs for mould images, and one particle classification CNN and one wear severity CNN for wear debris images). Then, the augmented and labelled datasets were split into training, validation, and testing datasets. The training and validation datasets were used in the training process, where the validation datasets were used for monitoring the existence of overfitting, and the testing datasets were used to test the models' generalisation to new data.

4. Applying trained CNNs in a two-level wear assessment system. A two-level inference system was developed to integrate wear mechanism assessment and wear severity assessment. In the system, the wear mechanism CNN classifies mould images into different wear mechanism category and abnormal images were sent to the corresponding wear severity CNN for severity assessment. The developed system was tested using full-tooth mould images to generate, for the first time, colour-coded wear mechanism maps that clearly showed the distribution of identified wear mechanisms on the entire tooth flank, and wear severity maps that visualise the severity distribution for each identified mechanism.

3.5 Image-based wear prediction using deep learning

A deep learning model, i.e., generative adversarial network (GAN), was developed for the first time to learn gear wear dynamics from direct records of gear wear evolution, i.e., mould image sequences (time series of mould images). The approach was inspired by studies on video prediction using deep learning techniques since a sequence of mould images for the same tooth is conceptually similar to a video. As illustrated in Figure 3.9, image-based gear wear prediction was realised via the following steps.

1. Construction of full-tooth gear mould images. Gear wear occurred on the entire active tooth flank in the gear tests, and it had different manifestations (e.g., mechanisms, severity) in different regions, especially along the tooth profile direction. If the training samples only captured a limited portion of the tooth surface, it could

be difficult for the wear prediction model to learn the overall gear wear dynamics, lowering the reliability of wear prediction. Therefore, full-tooth mould images were constructed by stitching multiple smaller mould images, and they showed overall how different wear mechanisms developed on gear teeth. More details about the construction process can be found in Chapter 7.

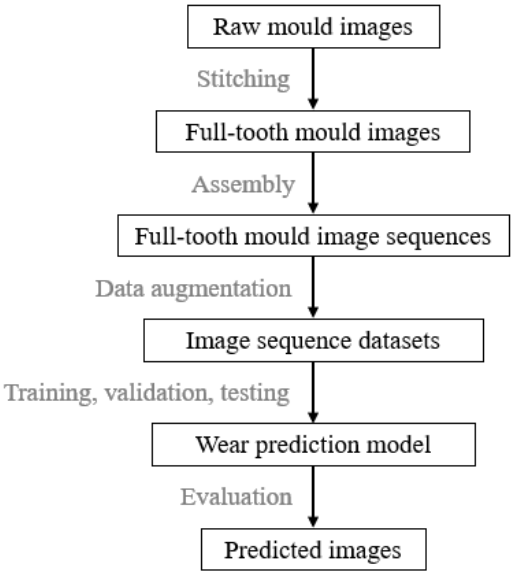


Figure 3.9: Image-based gear wear prediction using deep learning.

2. Assembly of individual full-tooth mould images into sequences. A certain number of the full-tooth mould images of the same gear tooth were selected and assembled in chronological order, forming a sequence of mould images. Considering that visual inspection and surface imaging could be carried out at different time intervals for different machines, mould image sequences with different time intervals were assembled. The sequences were similar to short videos that visualised the gear wear dynamics.

3. Data augmentation to build the image sequence datasets. The number of sequences of full-tooth mould images was insufficient for deep learning training, and the size of each image was too large, requiring prohibitive computing resources. In order to address the issues, image cropping was introduced to reduce the image size and

increase the number of sequences, and the cropping was performed only along the tooth width direction to ensure that each sequence covered the wear evolution along the tooth profile direction.

4. Construction, training, validation, and testing of the wear prediction model. An image sequence is a time series of images, and convolutional long short-term memory (ConvLSTM) networks are commonly used to process such data [171]. They can be regarded as a combination of convolutional operations that are famous for handling images, and LSTM networks that are famous for handling time series. Therefore, a ConvLSTM network was constructed to generate future mould images using past ones. A combination of three loss functions was designed to increase the accuracy and sharpness of predicted images. A CNN was employed to distinguish generated mould images from real mould images to increase the realism of predicted images. In this way, a GAN with a generator (ConvLSTM) and a discriminator (CNN) was built for gear wear prediction. Then, the prepared mould image sequences were split and used to train, validate, and test the wear prediction model.

5. Evaluation of predicted images. The predicted images were evaluated both qualitatively and quantitatively during testing regarding the prediction accuracy and image realism. The prediction accuracy measures the closeness between predicted images and their ground truth, while realism measures how realistic and sharp the predicted images are. Two commonly used image metrics were adopted to evaluate the overall accuracy and realism of predicted images, and two wear specific metrics were designed to evaluate the prediction accuracy of gear wear propagation.

3.6 Summary

In summary, this chapter provided an overview of the methodology designed to achieve the four research objectives. Key stages were introduced, including experimental tests, image acquisition and analysis, and using images and deep learning for automated gear

wear assessment and prediction. The methodology to achieve the four objectives can be summarised as follow.

- To achieve Objective 1 (to develop a method to obtain direct and comprehensive gear wear information without gearbox disassembly), a moulding method was developed to obtain gear teeth moulds that replicate worn gear surfaces. Optical microscopy and LSCM were selected to capture 2D and 3D images of the moulds, which contained both macro- and micro-scale wear features. A variety of direct gear wear information was extracted via image analysis techniques and utilised to monitor gear wear.

- To achieve Objective 2 (to investigate the relationship between the features of worn gear surfaces and those of wear debris), important numerical features of macropits and wear debris that were generated in the same wear process were selected and extracted from gear mould images and wear debris images via image segmentation and analysis. Both qualitative comparisons based on the images and quantitative comparisons based on extracted features were conducted to gain novel insights into the gear wear process.

- To achieve Objective 3 (to develop an automated system for gear wear mechanism and severity assessment), deep learning models (CNNs) were developed to identify gear mould images and wear debris images of different wear mechanisms and severity levels. Large datasets of the two types of images were purpose-built to ensure the accuracy and reliability of the automated assessment. A two-level inference system that integrated the assessment of wear mechanism and severity was also developed to convert full-tooth mould images into colour-coded maps clearly showing wear conditions on the tooth surface.

- To achieve Objective 4 (to develop a gear wear prediction model using direct wear information), a deep learning model (generative adversarial network (GAN)) was developed to learn gear wear dynamics from time-series of gear mould images. Large datasets of mould image sequences that directly recorded gear wear evolution were

built for the first time. A combination of three loss functions was designed to increase the prediction accuracy and sharpness, and an adversarial training strategy was developed to enhance the prediction realism.

Chapter 4

Application of Surface Replication Combined With Image Analysis to Investigate Wear Evolution on Gear Teeth: A Case Study

This chapter addresses the first objective of this research, i.e., **to develop a method to obtain direct and comprehensive gear wear information without gearbox disassembly**, via the combination of surface replication and image analysis. As illustrated in Figure 3.1, this chapter is the foundation of this thesis because its main outputs, i.e., gear mould images, are the core inputs of the other three chapters. **This chapter is a modified version of the published journal article below,** and the main modifications include additional experimental results and a condensed methodology section, of which the details about gear test rig and procedures were moved to Chapter 3.

- **H. Chang**, P. Borghesani, W. A. Smith, and Z. Peng, "Application of surface replication combined with image analysis to investigate wear evolution on gear teeth – A case study," *Wear*, vol. 430–431, pp. 355–368, 7 2019.

Abstract

Gear wear is a critical degradation process in many applications, and its monitoring is important to ensure the safe operation of machines. The case study presents a novel approach to obtain comprehensive and high-resolution wear information for monitoring the evolution of gear wear. This is achieved by combining a moulding method and non-contact imaging-based measurements in two- and three-dimensions and without gearbox disassembly and tooth removal. Optical and laser scanning confocal microscopy and image processing tools were employed to provide quantification of surface roughness, wear severity and wear depth on gear teeth during a lubricated endurance test. Also, a detailed qualitative analysis on the progression of the two main wear mechanisms (abrasive wear and fatigue pitting) was provided, alongside an evaluation of the diagnostic capabilities offered by the proposed methodology. The insights on gear wear offered in this case study also represent a valuable guideline for researchers and engineers in gear condition monitoring. Moreover, the proposed wear analysis technique can be used to characterise and understand wear phenomena in other conditions and geared transmissions.

4.1 Introduction

As reviewed in section 2.1, wear is a material degradation and removal process that can eventually lead to the failure of mechanical components and systems. Gear systems, widely used to transmit power, often suffer from abrasive wear and/or contact fatigue. Serious gear wear often induces excessive vibration and noise and significantly impacts the remaining useful life (RUL) of gears [5, 186]. To mitigate the impact of gear wear on the safety and reliability of machines, operators often put in place a schedule of regular inspections [5] and/or install online condition monitoring sensors. The most popular technologies include oil and wear debris analysis (OA/WDA) [26, 187], vibration [5, 55] and acoustic emission [188].

However, the effectiveness of gear wear monitoring systems and inspection schedules is

strongly dependent on the knowledge of gear wear dynamics. Thus, laboratory testing and analysis of wear evolution on gear teeth (both surface topography and tooth profile) are conducted to understand the progression of different wear mechanisms and their interactions. Observing and detailing different wear phenomena at the micro-scale has been an important and challenging task throughout the history of modern tribology [23, 152, 189, 190]. Surface topography at a micro level is known to affect the wear rate [23,27] and surface fatigue life [179,180] in lubricated contacts, while tooth profile deviation affects transmission error [5], vibration [191], and wear rate [146]. Also, modelling different gear wear modes often needs quantitative morphological information of corresponding surface features as input and/or validation [37, 68, 154]. Moreover, numerous condition monitoring attempts [26, 55, 59, 192] based on vibration or acoustic emissions (AE) require the ability to directly assess the surface condition to establish correlations and validate diagnostic techniques.

In the past, many surface inspection techniques have been applied to examine wear of gear teeth at the micro-scale [193, 194]. These techniques worked directly on gear teeth, thus requiring their transport and fitting within high-resolution instruments. This step increases the cost and complexity of the procedures since the dismantling of gearboxes is usually needed, and destructive tooth removal is required to perform a detailed analysis on and/or under the tooth surface, making the techniques suitable only for analysing the final stage of a gear wear process [156]. Researchers aiming at regular inspections of gear teeth then usually have to abandon microscopy and accept the significantly lower-quality output of a simple visual inspection [57, 195].

As mentioned in section 2.2, there is little research on direct gear wear monitoring without gearbox disassembly. The most promising method is based on using a moulding material to replicate the tooth surface, thus allowing microscopic analysis without gear disassembly and tooth removal. The few studies adopting this approach differ in the moulding material and analysis technique. Andersson [66,67] used acetyl cellulose films and acetone to obtain replicas of gear teeth, but the flexible films were intrinsically unable to include profile information (only micro-scale patterns were impressed in the replicas). Flodin [64, 68]

made replicas of gear teeth using a methyl methacrylate-based resin as the moulding material, obtaining hard moulds that retained tooth profile information. However, the author himself suggested the moulds had serious deformation issues, which led to the loss of information regarding the true gear teeth profiles, and there were thought to be problems with the replication of surface roughness. In addition, the surface information was ultimately obtained from the moulds using only a stylus, without the benefits of more sophisticated examination methods (e.g. optical imaging), leading to the loss of important information such as micro surface features. More recently, Moorthy and Shaw [69, 70] obtained replicas of gear teeth using silicone polymers and used them to observe surface damage by imaging and measuring tooth profile changes using a coordinate measuring machine of which the sensing probe will come into contact with samples in measurement. But silicone polymers may deform when being in contact with the probe, which might lead to measurement error. Also, the approach cannot obtain surface roughness and micro surface features such as small pits and microcracks that are important in monitoring wear evolution on gear teeth.

Therefore, this chapter aims to investigate the capabilities of combining surface replication with image analysis in wear monitoring and wear analysis on gear teeth, allowing the extension of the method to other gears and conditions. This chapter is organised as follows. Section 4.2 introduces the methodology adopted in the case study, and section 4.3 presents the results. Discussion is detailed in section 4.4, followed by the summary of this study in section 4.5.

4.2 Methodology

In order to obtain direct and comprehensive gear wear information without gearbox disassembly, a novel method that combines surface replication and image analysis was developed and applied in the lubricated gear wear test to investigate gear wear evolution. The gear test rig and procedures are briefly introduced in section 4.2.1, and section 4.2.2 details the development of surface replication of gear teeth, i.e., the moulding method, followed by

section 4.2.3 introducing the process of image acquisition and analysis to investigate gear wear evolution.

4.2.1 Gear test rig and procedures

The lubricated spur gear wear test was conducted on a single-stage spur gearbox rig, as shown in Figure 3.2. Full details about the test rig and the test procedures can be found in section 3.1, and only key information is presented in this section.

The driving gear has 19 teeth while the driven gear has 52 teeth, as shown in Figure 3.3. The two gears had an initial surface roughness of 3.2 µm, which was manually roughened to accelerate the initiation of fatigue pitting. The gear material is normalised JIS S45C (equivalent to AISI 1045) medium carbon steel with a hardness of around 190 HB. Six teeth (three of the driving gear and three of the driven gear) were selected and marked to make moulds, as shown in Figure 4.1. Considering that through visual inspection, all teeth experienced surface fatigue uniformly, three moulds (2 for analysis and comparison plus 1 spare one) on each gear are sufficient. In principle, a mould could be made on any tooth after the lubricant on the surface was removed. Practically, it involves removing the lubricant, cleaning the surface, applying the mould material and a curing period before removing the mould from the gear. The total duration for making one mould is approximately 30 minutes. If the above procedures were used to make 6 moulds one by one, it would take a total of 180 minutes at each stop, resulting in a significant increase in the total testing hours. To minimise the duration of the mould making process so that the test could be completed in a reasonable time frame, a total of 6 moulds (three moulds on each gear) were made simultaneously. To do so, 3 teeth on each gear were randomly selected within a 90-degree circular sector indicated in Figure 4.1, which were above the oil surface and had space for the moulding material to be applied to properly cover both tooth flanks. The selected tooth surfaces were cleaned altogether, and the mould materials were applied to them at one go. After curing (20 minutes), 6 moulds were ready to be collected and the test was continued. The total duration for collecting the 6 moulds at

one stop was about 40 minutes.

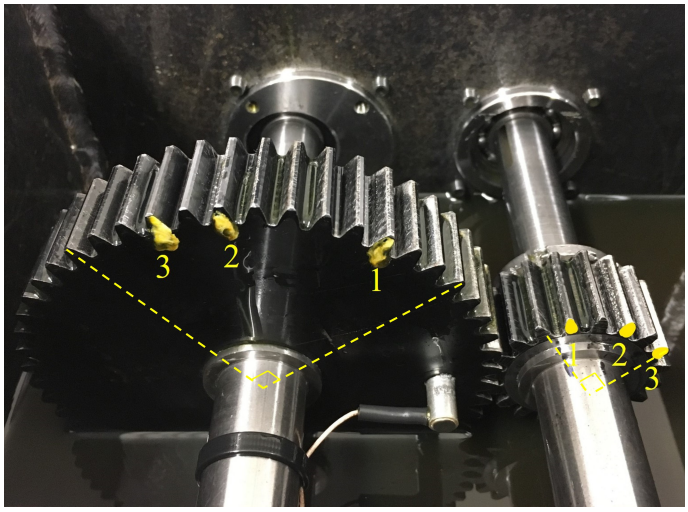


Figure 4.1: The driven (left) and driving (right) gears. The marked teeth were monitored continually using the method developed in section 4.2.2.

During the test, the test rig was stopped at pre-defined intervals to make moulds without disassembling the gearbox. In total, there were 40 intervals during 3.25 million running cycles (68 hrs), and shorter intervals were used at the beginning and the later stage of the test where rapid changes to the teeth occurred.

4.2.2 Surface replication of gear teeth

Three aspects of the performance of the surface replication material, or the moulding material, are critical to gear wear monitoring. They are: (i) the micro-scale resolution, i.e. the capability to fill micro-valleys and conform to micro-scale features on tooth surfaces; (ii) the accuracy at which moulds reproduce surface topography, i.e. surface roughness, without bias introduced, for instance, by deformation due to residual stress in the curing process; and (iii) the capability to reproduce tooth profiles (i.e. reproducing not only local surface roughness, but also macro-level profile shape and any changes therein, expected to be up to a few hundred micrometres for the module 2 gears tested in this study). Microset 101 (thixotropic silicone polymers) was chosen as the moulding material in this study,

which features high resolution, fast curing, low shrinkage, low thermal expansion, and ability to be used on vertical surfaces, and it overcomes the issues of the acetyl cellulose films (loss of the surface profile information) used in [67] and the methyl-methacrylate resin (large deformation) used in [68]. The high resolution and low shrinkage of the silicone polymers are due to the excellent wettability and low surface tension of the silicone polymer, which are related to the low intermolecular forces between methyl groups and the flexibility of the backbone of silicone [196]. A detailed evaluation of the replication quality of the moulding material was conducted through comparing the resolution, accuracy, and profile retention to those of original engineering surfaces, and the procedures and results are presented in Appendix A. Nominal technical information of the moulding material can be found in Table 3.3.

During each interval in the test, the marked teeth were cleaned using compressed air and the remaining oil was removed using absorbent materials. Then, the moulding material was applied to the teeth using a dispenser with an extended mixing nozzle (see Figure 3.4) until the valleys between gear teeth were filled (see Figure 3.5). After the curing time (20 minutes at 25°C), the moulds were taken off (see Figure 3.5) and properly stored for future analysis.

4.2.3 Image acquisition and processing for wear analysis

The moulds of gear teeth were examined using optical microscopy and laser scanning confocal microscopy (LSCM), mainly for qualitative and quantitative analysis, respectively. The choice of an imaging-based surface measurement was determined by the need to preserve detailed profile information, which would be compromised using contact-based techniques (e.g. profilometer) on a shape-preserving yet relatively soft moulding material.

An optical microscope was used to capture 2D images of the moulds, with the twofold objectives of qualitatively observing the wear evolution and quantitatively estimating the growth of pitted area on gear teeth. In order to obtain a qualitative and comprehensive picture of the tooth wear evolution, a 5x magnification lens was used to image the entire

width of a gear tooth (axial direction). Seven images, each covering an area of 2.8 (length) x 2.1 (height) mm, were captured and stitched to obtain a high-resolution (a few microns) representation of the entire addendum or dedendum of a gear tooth (Figure 4.2 (a) is an example for the dedendum).

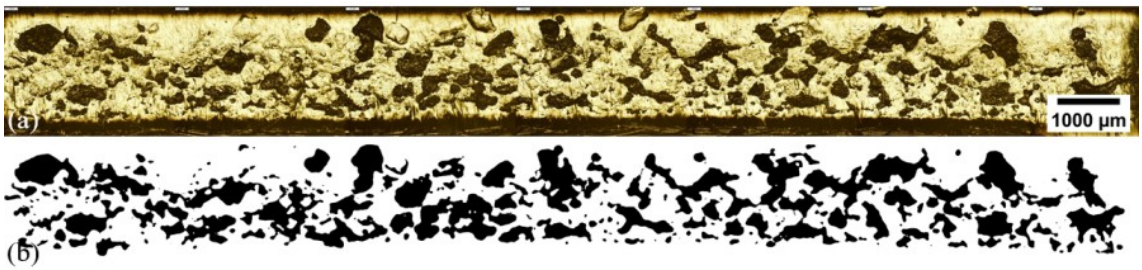


Figure 4.2: (a) A stitched 2D image of the dedendum on driving gear teeth and (b) its binary image. Individual images were acquired using optical microscopy (5x objective lens).

The stitching procedure was done on both the addendum and dedendum of the moulds, resulting in a series of 14 images per mould. The height of each image (2.1 mm) was sufficient to cover more than half of the active profile in the radial direction. For the quantification of pitted area, a 30x30 pixels median filter was applied to the stitched 2D images (typically 9400x1030 pixels) to remove noise, followed by a binarisation process using Otsu's method [33]. The latter transforms the images into logical matrices classifying each pixel as either pitted (black/0) or not pitted (white/1). An example of this transformation is given in Figure 4.2 (b). Finally, the pitted area was calculated by summing and scaling inverted logical matrices (where 1 stands for pitted). In addition to these images, a higher magnification of 20x was used for examining micro features (e.g. cracks) in specific areas of the tooth.

LSCM was employed to capture 3D images of radial strips along the radial direction of each mould and of additional areas where features of interest (e.g., cracks, pits and wear marks) were located. A single radial strip was obtained by stitching seven 1.4x1 mm images captured with a 10x objective lens. The resulting strip (e.g. Figure 4.3) contained the height and spatial information of a complete tooth profile plus two unworn portions

at the root and tip (owing to tip relief) of the tooth. The latter were used as references in the evaluation of the wear depth and the comparison of profiles at different stages of the test. In particular, the estimated tooth profiles were obtained by averaging the strip in the axial direction to reduce the effects of localised height variations.

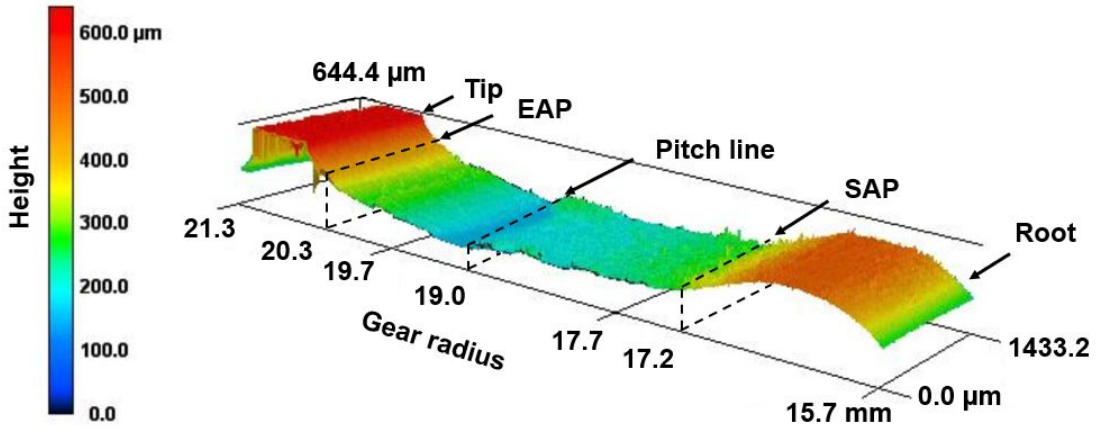


Figure 4.3: A stitched 3D image of the profile of driving gear tooth 2 (SAP: start of active profile, EAP: end of active profile). Individual images were acquired using LSCM with a 10x objective lens.

Similar strips were obtained to evaluate surface roughness yet based on 8 images of the active profile only and with an 20x objective lens. The surface roughness index S_a was calculated based on ISO 4288:1996 on three individual strips at different axial locations across the tooth width (left, centre, and right), and the average of the three S_a values was used. The 20x lens was also used for 3D images on specific areas of the mould in correspondence to identified features of interest. This magnification factor results in a lateral resolution of 0.7 µm and a depth resolution of 0.5 µm.

4.3 Results

The surface replication method was applied to gear teeth in a lubricated spur gear test where the test rig was stopped at pre-defined intervals for mould making. Then, the moulds of gear teeth were imaged using optical microscopy and LSCM, and image analysis

techniques were employed to quantify the gear wear evolution. The observations on the pitting process, including its initiation and propagation and its interactions with abrasive wear and plastic deformation, are reported. Terminologies for gear wear used in this paper are consistent with ANSI/AGMA 1010-F14 [21]. More specifically, macropitting is a type of fatigue wear on soft tooth surfaced caused by repeated Hertzian stresses at the macroscale (this case), while micropitting is another type of fatigue wear caused by repeated Hertzian stresses and plastic flow at the asperity scale and often occurs on hard gear teeth.

4.3.1 Wear evolution on gear teeth at the macroscale

As mentioned in section 4.2.1, three driving gear teeth and three driven gear teeth were monitored throughout the test. It was found that those driving gear teeth and driven gear teeth had the same or very similar wear patterns. Therefore, only one of the driving gear teeth and one of the driven gear teeth were used here as an example to show the wear process.

Figure 4.4 shows the dedendum of tooth 2 of the driving gear at seven different times during the test. Note that the darker areas in the figure represent macropits (confirmed to be protrusions on the moulds). Overall, there were three identified stages in the evolution of the driving gear tooth surface. The first stage involved polishing or running-in from the beginning (Figure 4.4 (a)) to 0.13 million cycles (Figure 4.4 (c)). The driving gear experienced a quick polishing or running-in stage when its original rough surface was effectively smoothened in less than 0.013 million cycles (11 mins). The second stage showed the fast growth of small and large macropits from 0.13 to 2.17 million cycles (Figure 4.4 (f)). In the second stage, The macropitting process initiated around the start of active profile (SAP) and then propagated towards and beyond the pitch line (Note that the driving gear meshes dedendum first). Furthermore, small macropits appeared before 0.13 million cycles, and then quickly spread to the entire dedendum by about 0.86 million cycles. Large macropits (length $>$ 500 µm) were mainly distributed near the SAP early in

the test and then progressed to the pitch line at the end. The final stage experienced the stabilisation (zero growth, albeit with some fluctuations) of the macropitting process in the period of 2.17 to 3.25 million cycles (Figure 4.4 (g)), in which macropits grew and new macropits formed, while many existing macropits shrank in size and even disappeared, probably as a consequence of abrasive wear and plastic deformation.

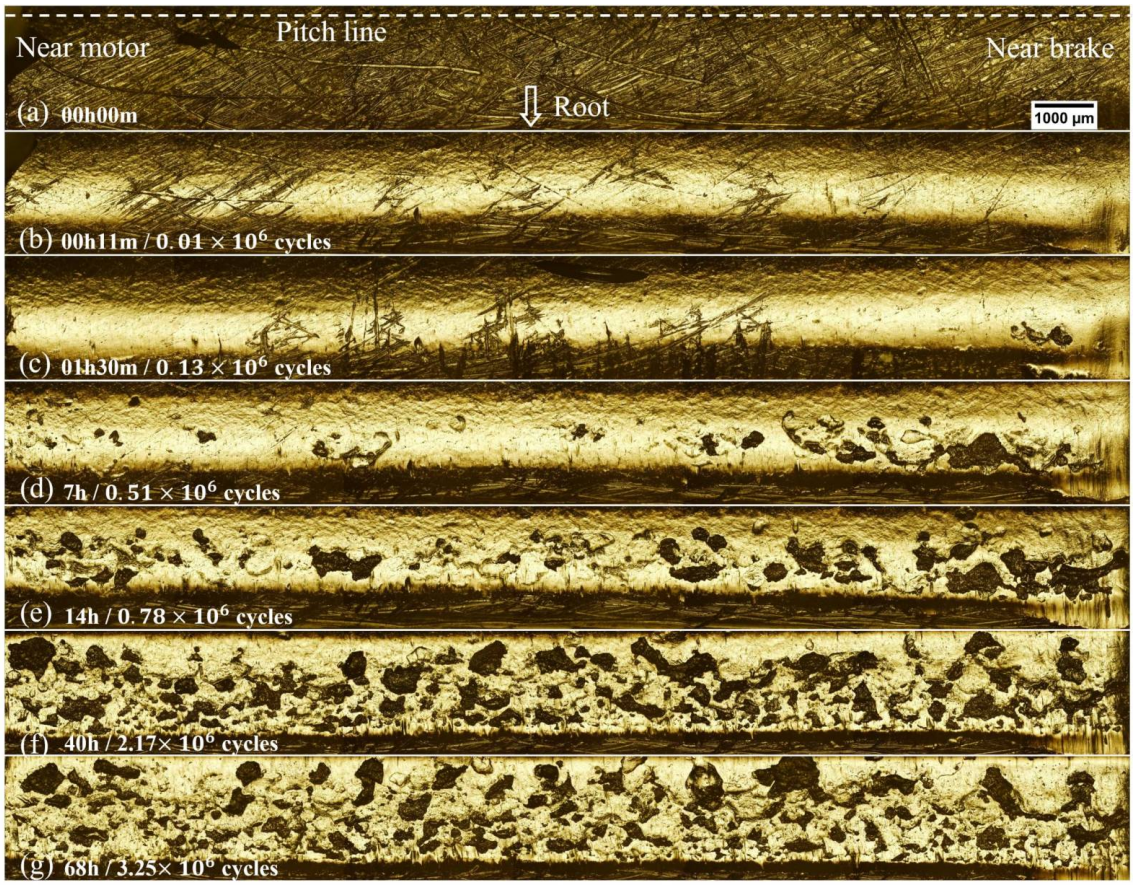


Figure 4.4: (a-g) Wear evolution on the dedendum of driving gear tooth 2 (0-68 hours; 0-3.25 million cycles).

Figure 4.5 shows the addendum of tooth 2 of the driving gear at six different times. Unlike the dedendum, the addendum of the driving gear teeth experienced predominantly abrasive wear, with little macropitting present. The growth of abrasive wear can be identified by the increase of parallel scratches oriented in the direction of sliding and the removal of initial random scratches. There were no obvious macropits on the addendum, but small

pits of around 10 to 20 µm were observed from the initial stage (0.01 million cycles). In addition, indentations, a type of plastic deformation [21], (protrusions in the moulds) can be seen on the addendum, which had similar sizes to macropits on the dedendum, and they were potentially caused by wear debris trapped between the two contact surfaces.

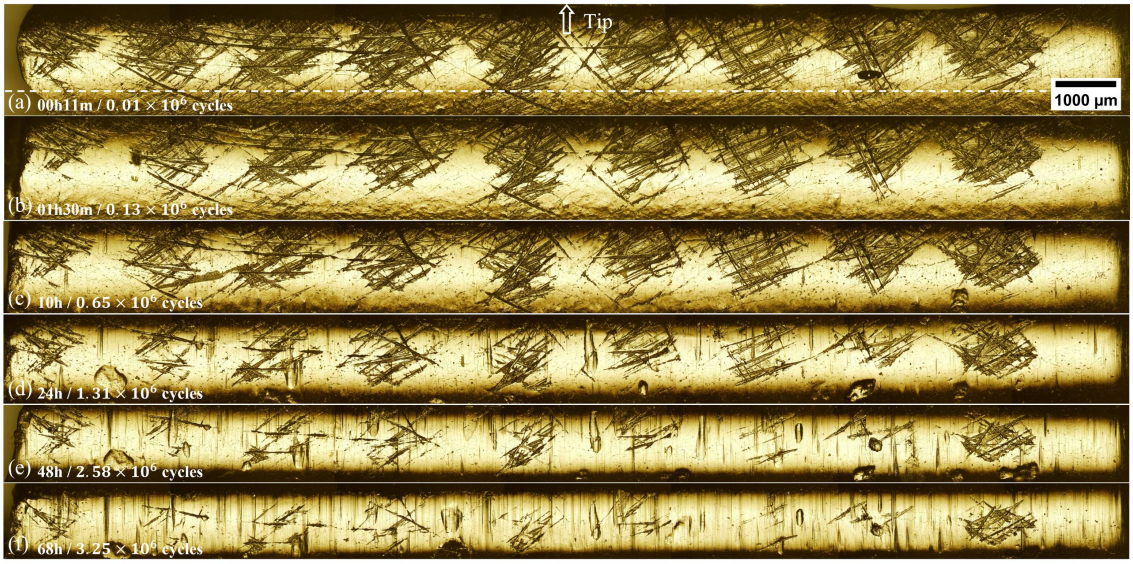


Figure 4.5: (a-f) Wear evolution on the addendum of driving gear tooth 2 (0-68 hours; 0.01-3.25 million cycles).

Figure 4.6 shows the dedendum of tooth 1 of the driven gear at seven different times. In general, the dedendum of the driven gear experienced much slower and less severe wear than the driving gear did. This is exemplified by the slow polishing process, which was not completed at the end of the test. At about 21h (i.e., 1.1 million cycles), in addition to abrasive wear, macropitting commenced, evidenced by the small number of small macropits shown in Figure 4.6 (d). In comparison to the macropitting process of the driving gear, the macropitting growth rate of the driven gear was much slower (see the changes in the macropit numbers from Figure 4.6 (d) to (f)). It can be seen in Figure 4.6 (g) that at the end of the test, most of the macropits were distributed along an axial line. Similar to the driving gear, obvious abrasive wear can also be noticed around the SAP, but it is not as severe as the driving gear case. The driven gear dedendum also showed many indentations with similar sizes and shapes to those on the driving gear addendum.

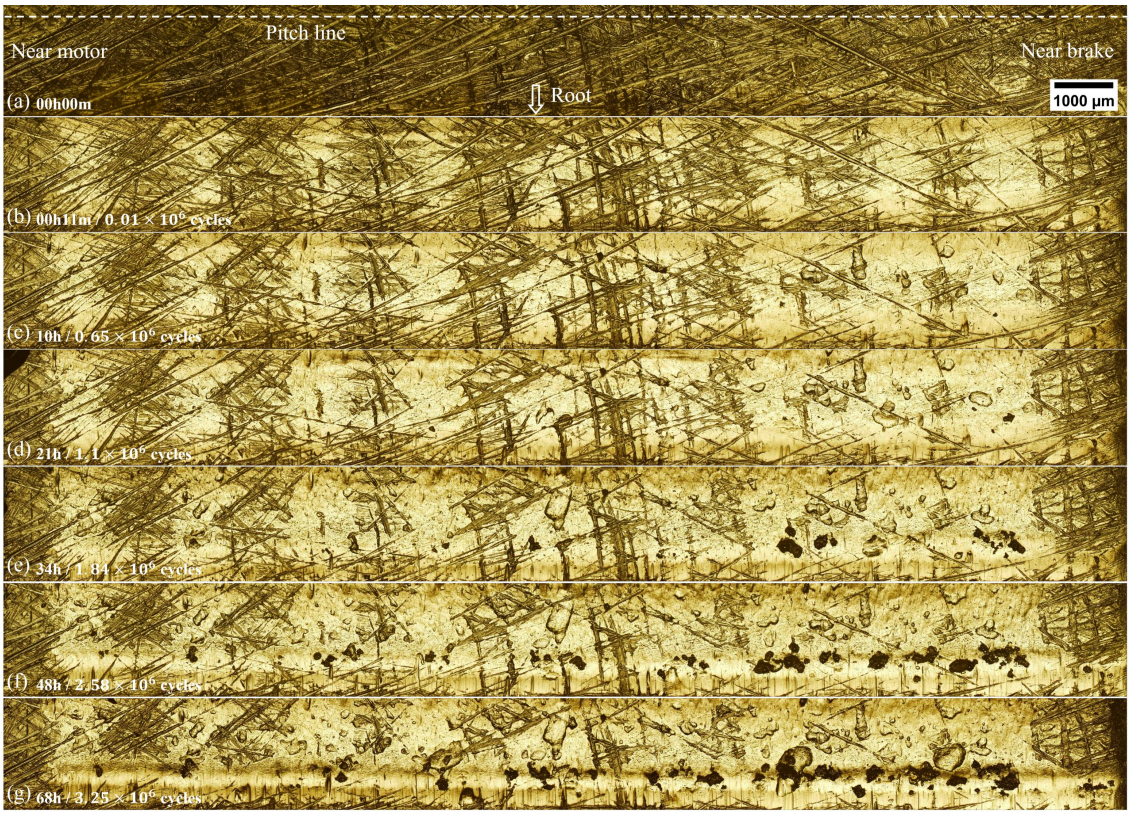


Figure 4.6: (a-g) Wear evolution on the dedendum of driven gear tooth 1 (0-68 hours; 0-3.25 million cycles).

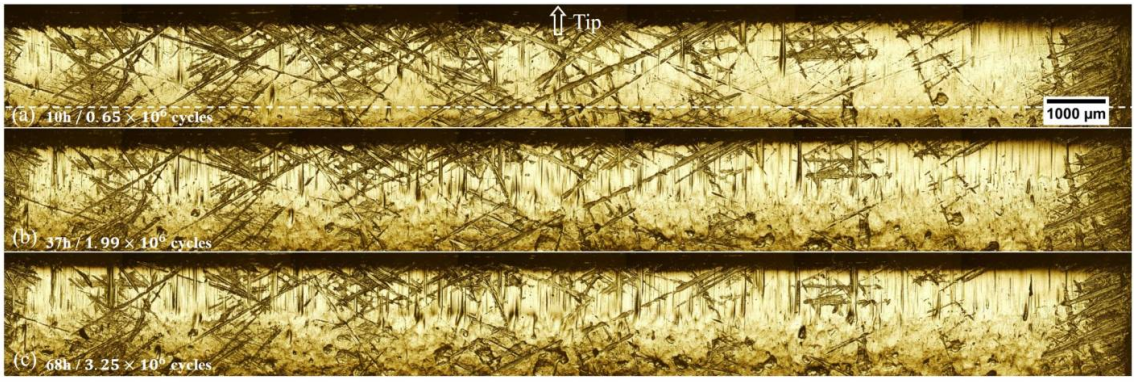


Figure 4.7: (a-c) Wear evolution on the addendum of driven gear tooth 1 (10-68 hours; 0.65-3.25 million cycles).

Figure 4.7 shows the addendum of the driven gear at three different times. The addendum of the driven gear had the same dominant wear mechanism as on the driving gear adden-

dum, which was abrasive wear. Compared to the features on the addendum of the driving gear, larger and denser parallel scratches were found on the driven gear addendum. It can also be noticed that there were more initial scratches removed at the left edge than the right edge, suggesting the existence of slight misalignment.

4.3.2 Surface roughness evolution on gear teeth

Figure 4.8 shows the surface roughness (S_a) evolution of the driving gear in the test. As mentioned in section 4.3.1, the driven gear teeth had a much less obvious change than the driving gear, so only the surface roughness of the driving gear teeth was tracked.

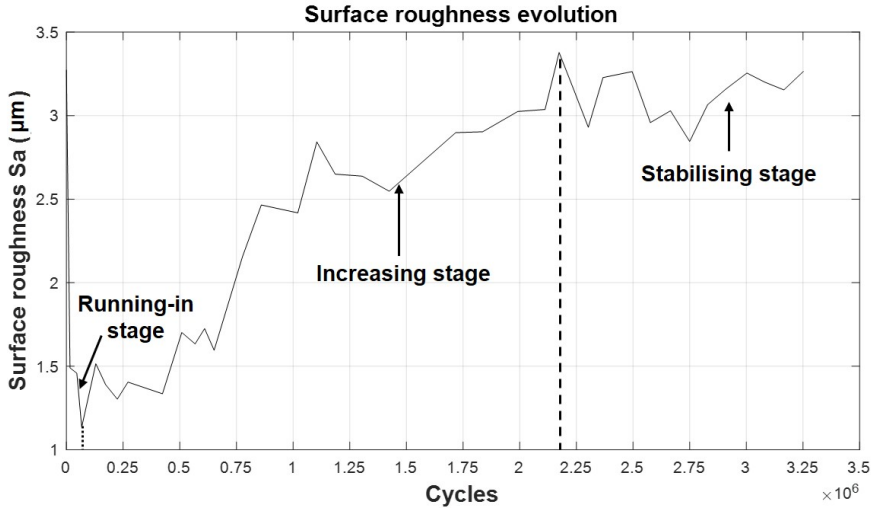


Figure 4.8: Surface roughness evolution of driving gear tooth 2.

Similar to the macro-scale wear evolution presented in section 4.3.1, three stages can be identified on the S_a data in Figure 4.8. The three stages were the running-in stage (0 to 0.1 million cycles) with the surface roughness decreasing, the increasing stage (0.1 to 2.2 million cycles), and the stabilising stage (2.2 to 3.3 million cycles) with no significant change in the surface roughness. The initial surface roughness of the driving gear teeth was $S_a = 3.2$ µm, and it was reduced rapidly to 1.1 µm in the first stage. The maximum surface roughness, occurring at the end of the second stage, was $S_a = 3.4$ µm. Within

the second stage, two sub-stages can be identified: the fast growth stage (0.13 to around 1 million cycles) and the slow growth stage (1 to 2.2 million cycles).

4.3.3 Macropitting initiation on gear teeth

Fatigue pitting was found to be the main wear mechanism that characterised the driving gear surface. Thus, a closer examination of the initiation and the propagation of macropitting was implemented qualitatively and quantitatively.

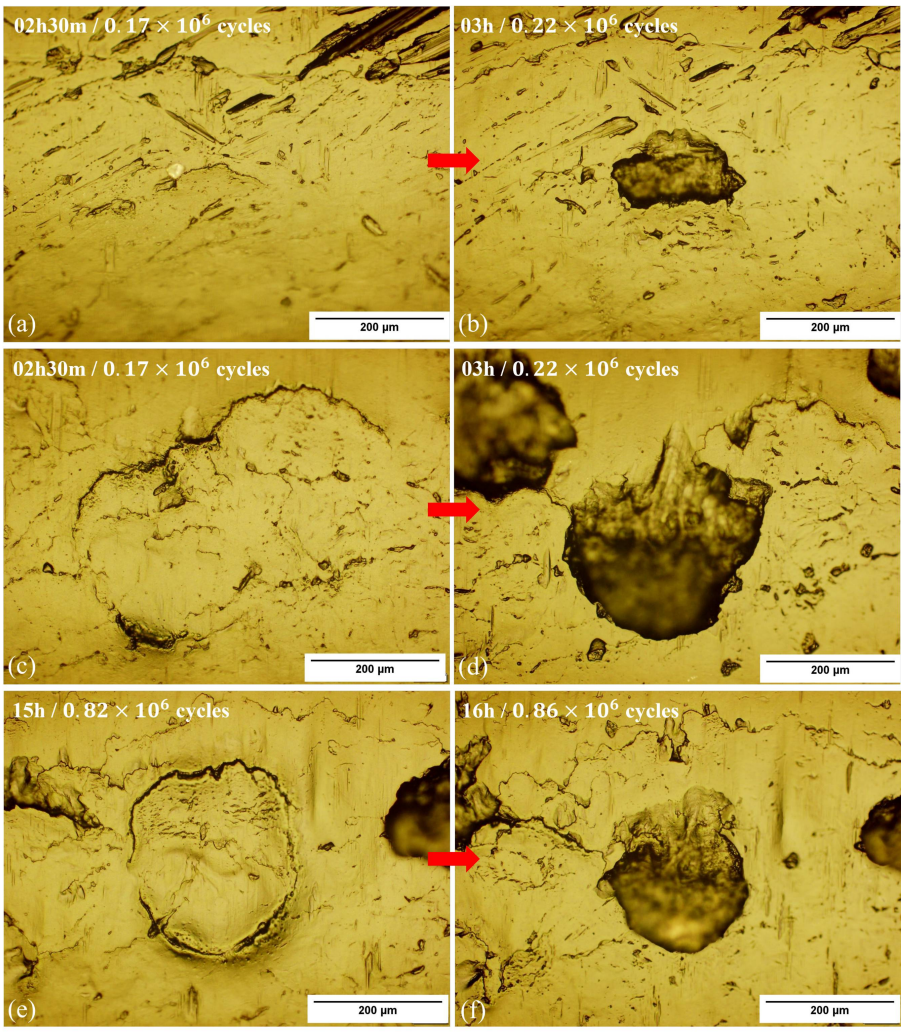


Figure 4.9: The driving gear tooth surfaces (a, c, e) before and (b, d, f) after macropits formed.

Figure 4.9 presents three examples of tooth surface before and after a macropit formed. Before materials came off the tooth surface (Figure 4.9 (a), (c), and (d)), cracks that almost enclosed the soon-to-be macropit can be clearly observed. For many macropits observed throughout the test, cracks could be seen before they formed, and it was found that on formation, the macropits were usually smaller than the area the cracks had enclosed.

Figure 4.10 shows 3D shapes of three macropits at different stages (see section 4.3.1). Pit 1 was the earliest initiated macropit, while the other two formed during the fast growth and stabilisation stages, respectively. It was found that most macropits had a wedge shape and the sharp end always pointed in the same direction as the relative sliding velocity, as shown in Figure 4.10 (a-c). The depth of newly-formed macropits was in the range of $30 \sim 90$ µm, and large macropits tended to be deeper than small ones. After macropits formed, their depth was observed to decrease in general, as shown in Figure 4.10 (d).

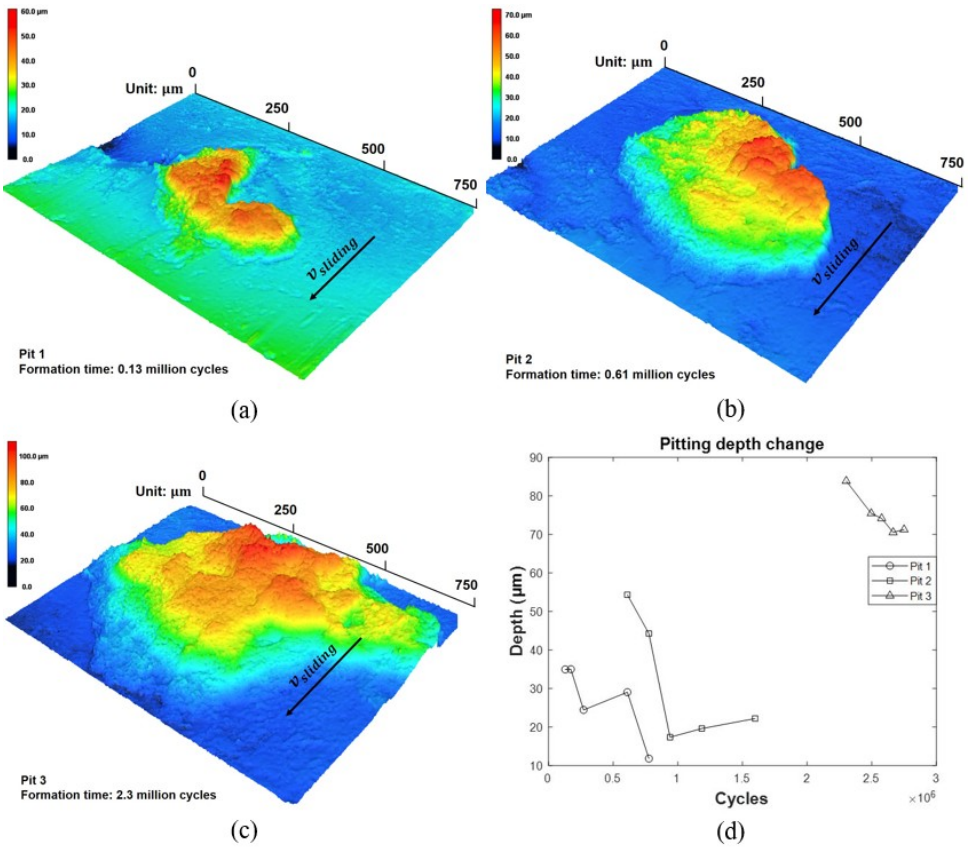
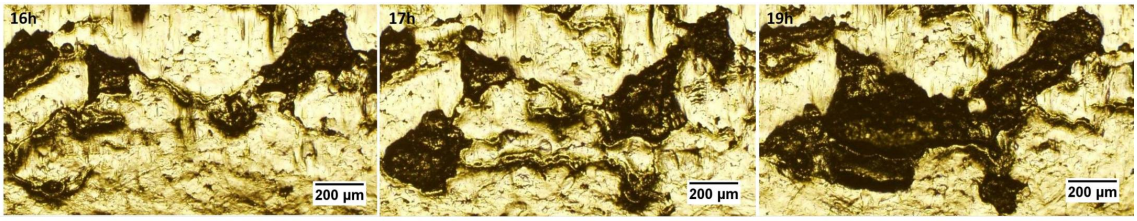


Figure 4.10: (a-c) 3D shapes of three macropits and (d) their depth changes at five times.

4.3.4 Macropitting propagation on driving gear teeth

It was found macropitting propagated in three modes. They are: (i) the ongoing initiation of new macropits that are relatively far from other pits (Figure 4.11 (a)), (ii) self-expansion of macropits caused by new cracks growing from original ones (Figure 4.11 (b)), and (iii) merging of neighbouring macropits into a larger macropit (Figure 4.11 (c)).



(a) (b) (c)

Figure 4.11: (a-c) Merging of small macropits into a large macropit on driving gear teeth (16-19 hours; 0.86-1.02 million cycles).

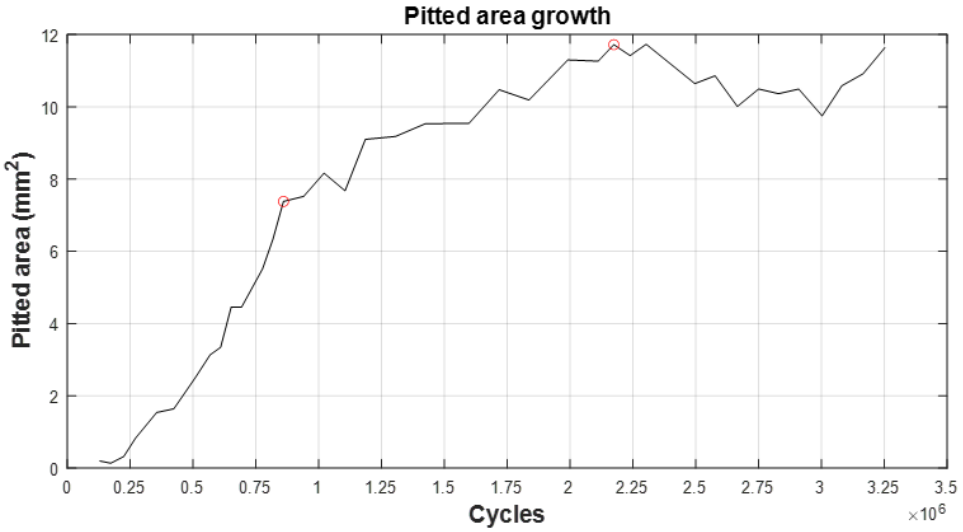


Figure 4.12: Driving gear pitted area evolution throughout the test.

Quantitative analysis of macropitting propagation on driving gear teeth was realised using microscopic images and image analysis as described in section 4.2.3. The growth of total pitted area on tooth 2 on the driving gear is shown in Figure 4.12, obtained by summing the logical matrix of Figure 4.2 (b) and calibrating. Only one tooth was selected because

the macropitting severity was uniformly repeated across different teeth. There were also three identified stages of pitted area growth, which were the fast growth stage (0.13 to 0.86 million cycles), slow growth stage (0.86 to 2.17 million cycles), and the stabilisation stage (2.17 to 3.25 million cycles). Macropitting started before 1h30m (0.13 million cycles) and quickly expanded to around 7.4 mm^2 out of the area of the dedendum of 40 mm^2 at 0.86 million cycles. The maximum pitted area was 11.7 mm^2 at 2.17 million cycles.

In addition to the pitted area, it is also important to find the pitting distribution on gear teeth. Figure 4.13 and Figure 4.14 show the evolution of macropitting distribution in the axial and radial directions, respectively. The grayscale value of each small rectangle in the two figures corresponds to the pitted area percentage at a certain axial or radial position on the gear tooth surface at a certain time. In the axial direction, before 1 million cycles, the more heavily loaded (brake) side has notably denser macropitting than other regions, but by the end, macropitting is evenly distributed along this direction. In the radial direction, macropitting started around the SAP and spread towards the pitch line at a steady rate. In the end, severe macropitting (pitted area percentage larger than 30% in Figure 4.14) was mainly located around the SAP and the pitch line.

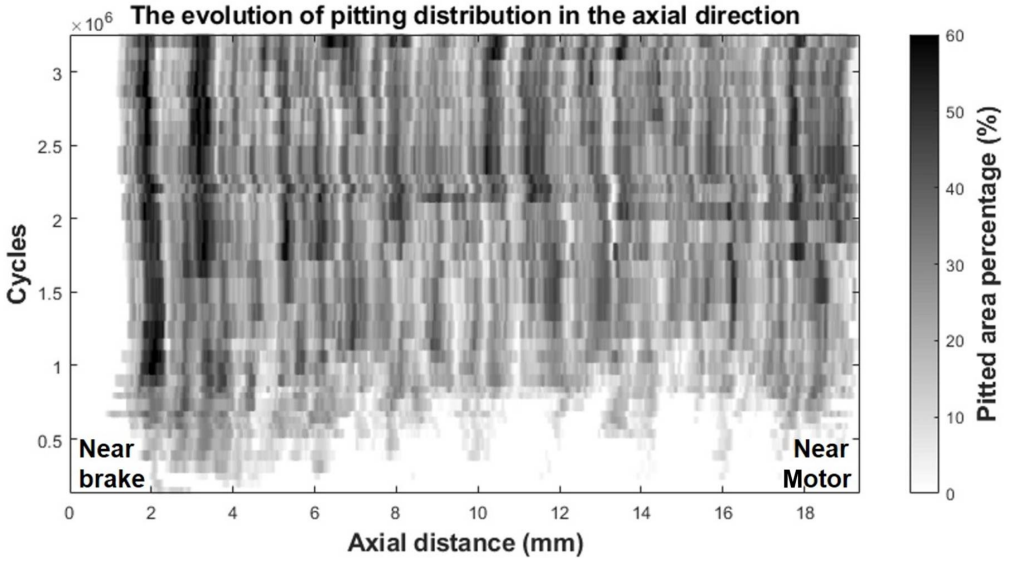


Figure 4.13: The evolution of macropitting distribution in the axial direction on the driving gear tooth 2.

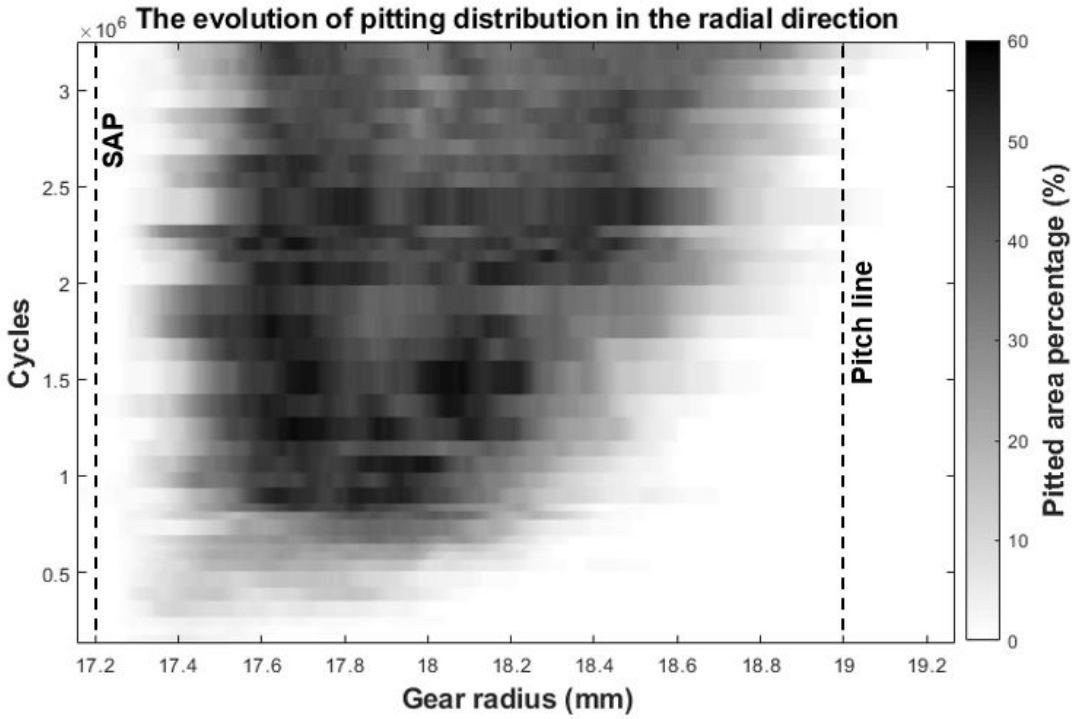


Figure 4.14: The evolution of macropitting distribution in the radial direction on the driving gear tooth 2.

4.3.5 Tooth profile change of the driving gear

As wear progresses from the micro to macro level, geometric change occurs to the tooth profile. The modified tooth profile can be obtained using LSCM by scanning the moulds from the tip to the root (see section 4.2.3). Figure 4.15 shows the average profiles of tooth 2 on the driving gear at three different times – 0h11m, 48h and 68h, or 0.013, 2.58, and 3.25 million cycles – corresponding to the three stages described in section 4.3.1. Figure 4.16 shows the wear depth after the fast growth stage (orange) and at the end of the test (blue), which was obtained by subtracting profiles at 48h and 68h in Figure 4.15 by the profile at 11m, and it also shows the moving of the contact line of minimum wear towards the tooth tip on driving gear teeth, except the tip and the root. A few observations of tooth profile changes are as follows. (a) After 3.25 million cycles, the maximum wear depth was 80 µm which occurred at the dedendum, while the minimum wear depth, except for

two ends of the active profiles, was 14 µm which occurred 0.2 mm outside the pitch line at the addendum. (b) It can be seen clearly from both figures that most material removal happened during the fast growth stage. (c) The dedendum wore substantially more than the addendum. (d) During the fast growth stage, the pitch line had nearly zero wear ($\sim$ 1 µm), but during the stabilization stage, the area around the pitch line experienced relatively severe wear.

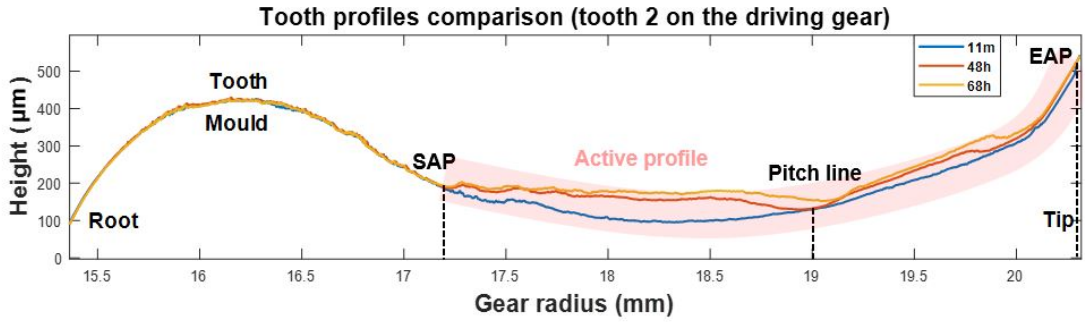


Figure 4.15: Tooth profile changes of the driving gear (tooth 2) with the active profile marked in light red starting from the SAP and ending at the EAP.

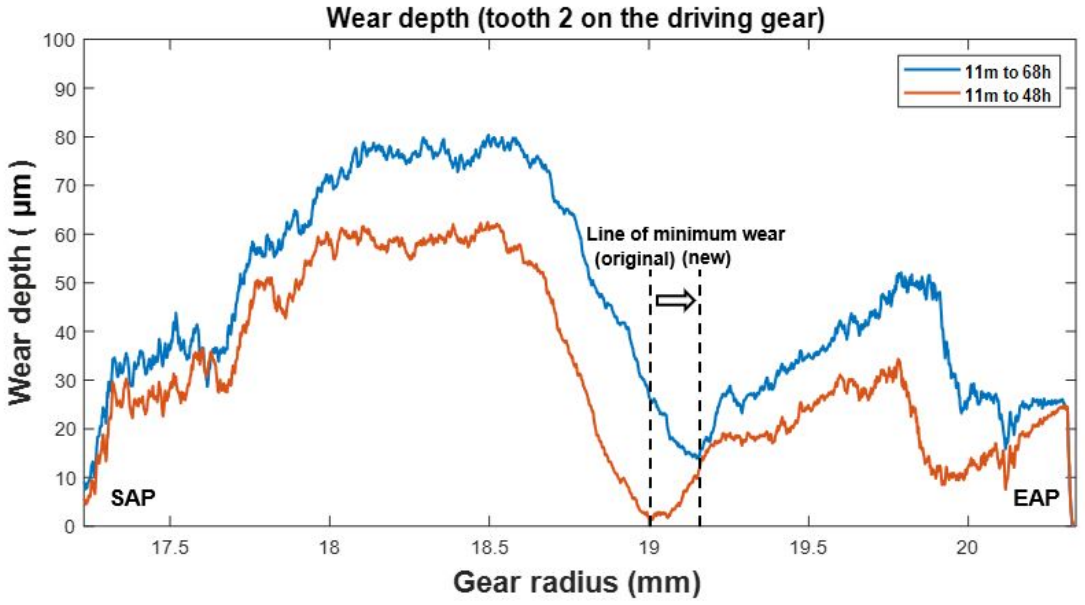


Figure 4.16: Wear depth on the driving gear tooth 2.

4.4 Discussion

4.4.1 Evaluation of the quality of the proposed method on gear wear monitoring

The combination of the surface replication and image analysis methods presented in this case study has allowed us to monitor and quantitatively measure the evolution of the tooth surface and profile in the wear process. The micro-scale examination of the surface damage on gear teeth was possible thanks to the high resolution (see section 4.2.2 for definition) of the moulding material (0.1 µm) and the imaging facilities (ranging from a few microns to 0.7 µm). Both macropits up to $\sim$ 500 µm (see Figure 4.4) and small pits and cracks (see Figure 4.17) on gear teeth were clearly visible and detailed. The rigidity of the moulding material, coupled with non-contact measurements (imaging), enabled the tracking of tooth profile evolution.

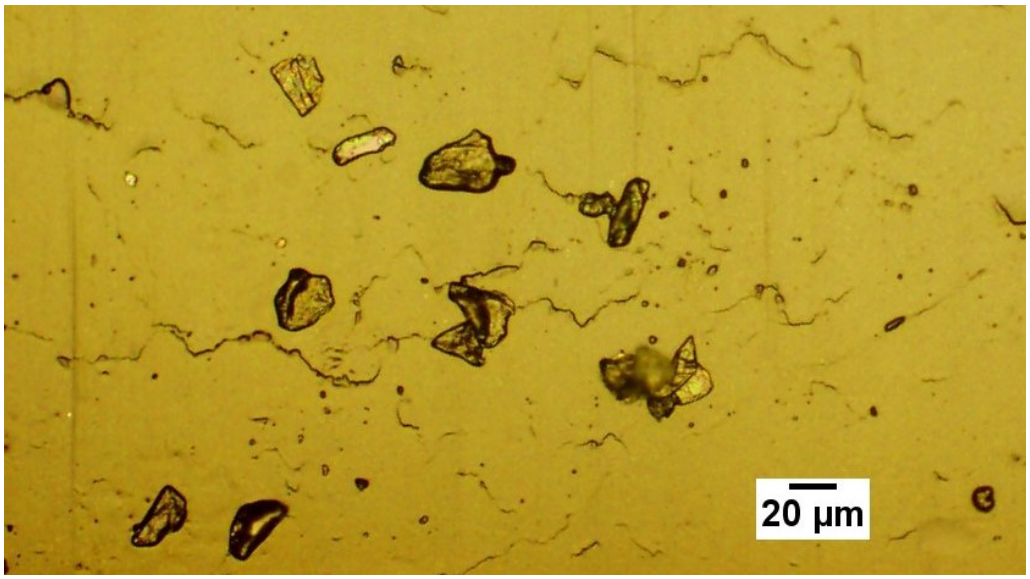


Figure 4.17: Small pits and cracks on the addendum of the driven gear tooth 2.

The quality of the procedure (and thus the validity of the case study results) was also validated against reference samples of known geometry and surface roughness. This additional validation is discussed in Appendix A.

4.4.2 The identification of macropitting, abrasive wear and plastic deformation and their interactions

By analysing the surface morphology of the gear teeth moulds, three types of surface damage, namely, macropitting, abrasive wear and plastic deformation were detected. Abrasive wear and plastic deformation were the main surface damage modes in the running-in period in which the initial rough surfaces were smoothened (see Figure 4.8). Macropitting (to be specific, progressive macropitting [21]) caused by repeated Hertzian stresses [20] existed for the majority of the test. Most macropits were larger than 100 µm, and many macropits were larger than 500 µm in length. Macropitting only occurred on the dedendum of both gears (see Figure 4.4 to Figure 4.7), which is aligned with results of many theoretical and experimental studies on macropitting [116, 179, 197–199].

Along with the macropitting process, abrasive wear co-existed as the secondary wear mechanism and can be identified by parallel scratches or furrows left on one or both contact surfaces [21, 200]. Large scratches longer than 100 µm were observed around the SAP and on the addendum of gear teeth of both gears (see Figure 4.5 and Figure 4.7), and short scratches were found all over the active profiles of gear teeth (see Figure 4.18). The location of the operating pitch line can also be identified by the absence of scratches or furrows, as shown in Figure 4.18.

Plastic deformation of a material is due to Hertzian stresses exceeding the yield strength of the material [21]. Plastic deformation is a well-known surface damage in gear contacts [21, 44, 201], but its records at the microscale on gear teeth are limited. The evidence of plastic deformation in this test is as follows. At the macroscale, rough surfaces of both gears were smoothened effectively in just 11 minutes of running (0.013 million cycles), and almost no large scratches were left on driving gear teeth (see Figure 4.4 (a)-(b)). At the microscale, abundant signs of plastic flow of materials filling or burying the large furrows left by the roughening process employed before the test were observed on gear teeth, as shown in Figure 4.19. It can be seen that aside from those large furrows, the area is relatively smooth, and there are only a few scratches.

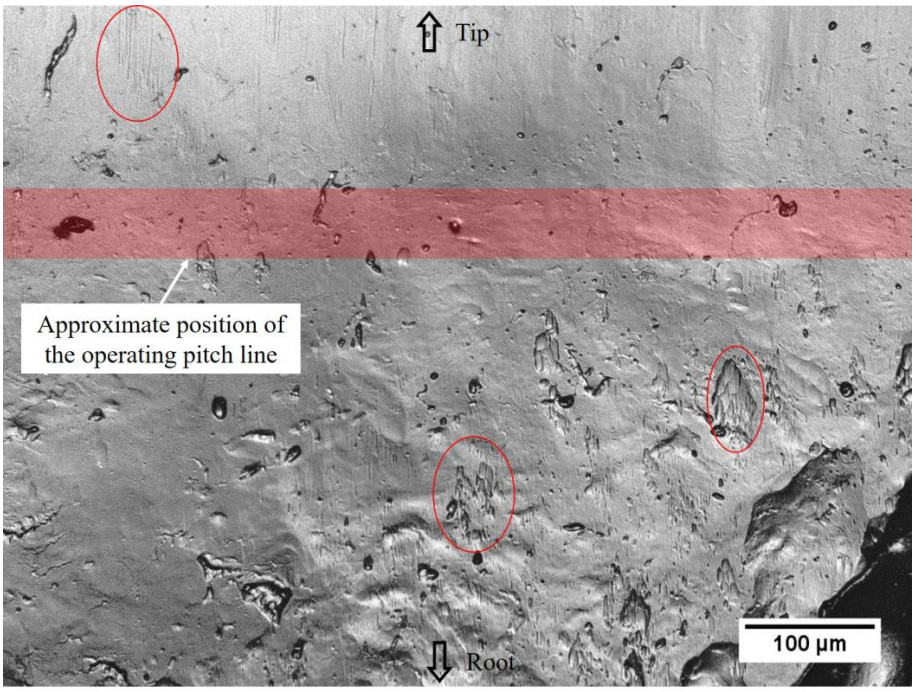


Figure 4.18: Parallel scratches around the operating pitch line on the driving gear (2.5 million cycles, circled are scratches or furrows caused by abrasive wear).

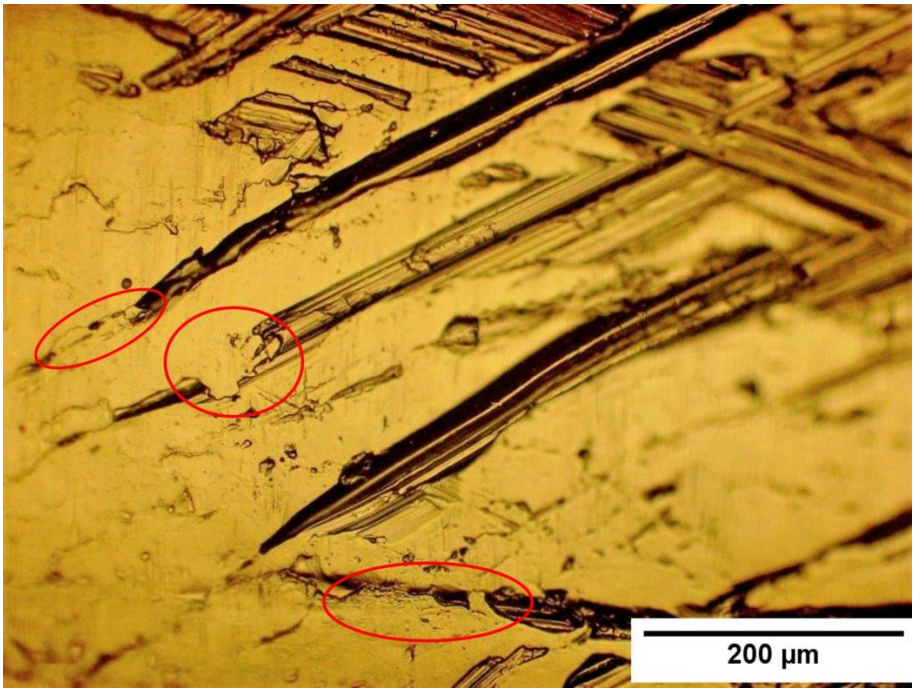


Figure 4.19: Plastic deformation of materials on driving gear teeth (circled are signs of plastic flow).

The three surface damage modes, macropitting, abrasive wear, and plastic deformation, did not act independently on the tooth surfaces. During the first 0.013 million cycles of the driving gear, plastic deformation and abrasive wear were the two main wear mechanisms observed. From 0.013 to 0.13 million cycles, abrasive wear steadily developed. Relatively uniform contact pressure due to the flattened surface and the work hardening effect is suggested as the main cause of the lower incidence of plastic deformation. Then, from around 0.13 million cycles, macropitting initiated on the driving gear and quickly propagated, becoming the dominant wear mechanism, which was reflected in the large pitted area and large wear depth on the dedendum of driving gear teeth. As shown in Figure 4.16, compared to the addendum of driving gear teeth where severe abrasive wear occurred but no macropitting appeared, the wear depth on the dedendum was nearly twice as much as the addendum after 2.58 million cycles.

The interactions between the three wear mechanisms were mainly reflected in the shrinking of the pitted area, which can be noticed easily from 2D images, as shown in Figure 4.4, and the pitted area growth curve in Figure 4.12. The combined effect of abrasive wear and plastic deformation could be the cause for reducing the pitted area. Firstly, abrasive wear constantly removed materials from the dedendum, reducing the pitted area and the depth of macropits. Secondly, edges of newly formed macropits had a relatively high stress concentration compared to the non-pitted areas, and thus, plastic flow was likely to occur in these regions, resulting in the partial filling of the macropits.

4.4.3 Insights into macropitting initiation and propagation

Most signs on the moulds of gear teeth before macropitting was initiated were found to be cracks. Some cracks were closed and later resulted in macropits of the same shape. Based on 3D images of newly appeared macropits, inclination angles of crack faces with respect to the tooth surface were estimated in the range of 10-20 degrees, smaller than the 20-30 degrees adopted in many crack propagation models [37, 157] (this can be material- and operating condition-specific). Furthermore, the direction of crack propagation was found

to be opposite to the direction of the relative sliding velocity (see Figure 4.10 (a-c)). This statement is supported by the fact that macropitting only occurred on the dedendum of gear teeth of both gears in the test, and those regions experienced 'negative sliding' - the direction of the relative sliding velocity was opposite to the direction of the rolling velocity.

The location of macropitting initiation was initially around the SAP where rippling, which is periodic, wave-like plastic deformation [21], was found before those regions became pitted, as shown in Figure 4.20. The emergence of rippling indicated that this region experienced large tangential forces. After the initial stage, the location of macropitting initiation spread towards the pitch line where contact stresses were higher, as shown in Figure 4.4 and Figure 4.14. Apart from the locations of macropitting initiation, the average size of macropits also changed with time. Early-formed macropits (Figure 4.6 (c) and Figure 4.10 (a)) were smaller and shallower than late-formed ones (Figure 4.10 (c)) on average. The same trend might also apply to the generation of wear particles, fatigue particles in particular, which were found in literature usually to increase in size in the late stage of a gear test [202].

As described in section 4.3.4, three modes of macropitting propagation were observed: on-going initiation of new macropits, self-expansion of and merging of neighbouring macropits.

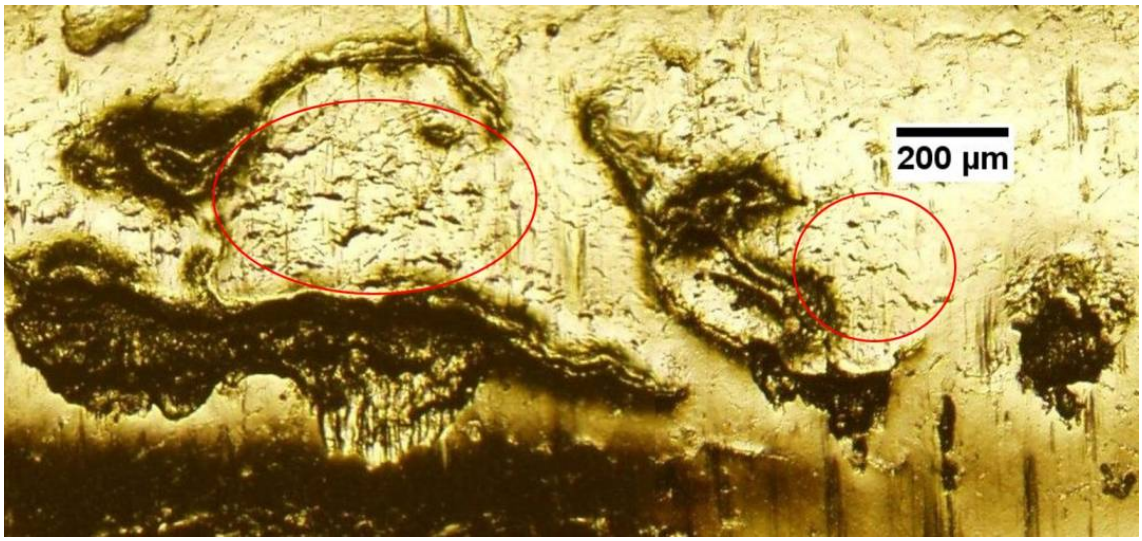


Figure 4.20: Rippling before macropitting propagated on the driving gear (0.27 million cycles, circled is the sign of rippling, a type of plastic deformation).

4.5 Summary

In this chapter, a combination of surface replication and image analysis was established and applied to investigate the wear evolution on gear teeth in the lubricated gear test. The replication quality of the moulding material (thixotropic silicone polymers) – including resolution, precision and profile retaining ability – enabled a series of detailed analyses of tooth surface evolution due to different wear mechanisms. Optical microscopy and LSCM were used to inspect gear tooth moulds, and image analysis allowed tracking the qualitative evolution of abrasive wear, macropitting and plastic deformation, and quantifying gear wear in terms of pitted area, surface roughness, and wear depth.

Insights drawn from this case study were summarised as follows:

- Three material degradation modes were detected: abrasive wear and plastic deformation that smoothened gear teeth in the short running-in period, followed by macropitting - the primary wear mechanism for most of the test, while abrasive wear was the secondary mechanism.
- Macropitting initiated around the SAP on driving gear teeth and propagated to the pitch line. The average macropit size increased with the propagation.
- Three growth stages of macropitting were observed: the fast growth stage (0.13 to 0.86 million cycles), the slow growth stage (0.86 to 2.17 million cycles), and the stabilisation stage (2.17 to 3.25 million cycles).
- Three stages of surface roughness (S_a) evolution were observed: the running-in stage (0 to 0.1 million cycles), the increasing stage (0.1 to 2.2 million cycles), and the stabilising stage (2.2 to 3.3 million cycles).
- The maximum wear depth (80 µm) occurred on the dedendum of the driving gear teeth, whereas the line of minimum wear-depth moved towards the tip.

This study has shown the potential of the proposed methodology for laboratory gear

wear investigation, with possible extension to field applications when the gearbox can be stopped and opened at regular intervals. This is not uncommon in industries characterised by seasonal production (e.g. food processing) and (with lower frequency) when major overhauls occur in usually continuous production systems. Moreover, despite being specific to this case study, the insights gathered on wear evolution of gear teeth, can be used as experimental evidence to guide the development and/or validate the results of gear-wear models.

Chapter 5

Investigation on the Relationship Between Gear Macropits and Wear Particles in a Gear Fatigue Process

This chapter addresses the second objective of this research, i.e., **to investigate the relationship between the features of worn gear surfaces and those of wear debris**, via comparing and analysing gear mould images and wear debris images obtained in the same fatigue wear process. This study is based on the development in Chapter 4, which effectively lowered the difficulty in collecting direct gear wear information, and it examines the links between worn gear surfaces and wear debris both qualitatively and quantitatively. The novelties of the method are as follow. (i) The scarcity of tribological data of gear teeth during a wear process is solved, and for the first time, gear macropits and wear particles during a wear process are collectively characterised and studied. (ii) Quantitative comparisons between pitting particles and gear macropits in terms of their tribological features are investigated and better understood.

This chapter is a modified version of the published journal article below, and the main modifications include a condensed methodology section of which the details about the experimental test and wear sample collection were moved to Chapter 3.

- **H. Chang**, P. Borghesani, and Z. Peng, "Investigation on the relationship between macropits and wear particles in a gear fatigue process," *Wear*, vol. 484-485, p.203724, 11 2021.

Abstract

Macropitting is one of the most common gear wear mechanisms, and monitoring its initiation and propagation is critical to ensure the safe operation of gearboxes. Wear debris analysis (WDA) is widely used to monitor and diagnose gear macropitting and its diagnostic capability relies on a good understanding of the relationship between wear particle features and those of actual gear macropits. However, the study on this relationship has been impeded by ongoing challenges in characterising gear macropits during the wear process. As a result, detailed experimental data of macropits and the quantitative relationships between wear particles and macropits evolving in the fatigue process are scarce.

In order to address the above issues, this study examines macropits and wear particles generated on gear surfaces and their evolution in a gear fatigue process through employing a moulding method to obtain 38 batches of gear teeth moulds in the wear process. 2D characterisations of macropits on the moulds and sampled wear particles were conducted using optical microscopy. Qualitative and quantitative relationships between different features of macropits and wear particles were investigated. Results showed the concentration of wear particles had a similar trend to pitted area growth, yet with a delay. This behaviour was also found in 4 non-dimensional numerical features characterising the geometry of particles and pits. The results of this study provide further insights into the evolution of gear contact fatigue and can assist in its monitoring.

5.1 Introduction

As reviewed in section 2.1.3, macropitting is one of the most common wear mechanisms of gears [43,203], and it is the result of gear surface fatigue accumulated under cyclic rolling-sliding contacts [33,204]. The initiation of macropitting occurs when fatigue micro-cracks form due to stress concentrations at the contact surface (related to surface asperities, wear debris, etc.) or subsurface (related to inclusions, voids, etc.), and the cracks propagate under cyclic loads until the detachment of gear material [151, 178, 199]. The development of macropitting can be impacted by many factors, including load, surface roughness, lubricants, slide-roll ratio, etc [46,179,180].

Severe macropitting often induces excessive vibration and noise and may eventually lead to tooth breakage [20,199,205]. Thus, machine condition monitoring (MCM) techniques are often employed to detect the presence of macropitting to better maintain the gearbox. Wear debris analysis (WDA) is one of the most widely used techniques for MCM. WDA has been used offline for decades and recently implemented online with the development of online wear debris sensors [101]. Both offline and online WDA examine wear debris characteristics, including concentration/quantity, morphology, colour, and composition, to estimate the wear rate and wear mechanism of a machine [127]. The capability of WDA to estimate the wear rate and wear mechanism is based on the fact that wear debris is the by-product of a wear process [206]. This means that the total amount of wear debris in lubricants equals the material loss of contact components if without wear debris removal. However, studies have revealed that wear debris may be entrained into contacts between components after its generation, resulting in its deformation or fracture [10–12]. Thus, the size and morphology of wear particles may be different from their original states. Therefore, the strength and characteristics of the correlation between wear debris features and worn surface condition must be assessed to establish the reliability of wear-debris-based gear monitoring.

There are numerous studies on wear particle characterisation [109–113] and worn surface examination [64, 110, 114–116]. However, studies that correlate characteristics of wear

particles and worn surfaces are few [105–107, 207]. In [105], the authors examined worn surfaces and wear particles generated under lubricated sliding contacts between steels and classified them into six wear regimes, based on a qualitative assessment of their morphology (e.g. smoothness for worn surfaces, size and colour for wear particles). In [106], the authors qualitatively compared two textural parameters obtained from co-occurrence matrices of images of wear particles and worn surfaces in a set of pin-on-disc tests and suggested differences in the two parameters between wear particles and worn surfaces. The surface roughness of wear particles and worn surfaces were measured in different wear stages under rolling and sliding contacts [107, 207]. It was found that wear particles and worn surfaces share a similar trend of surface roughness changes.

Despite these interesting studies, the characterisation of the evolution of both surfaces and particles throughout the wear process is limited, especially for gears, due to the difficulty of frequent image acquisition of worn gear teeth. Due to this reason, microscopic images of worn gear teeth in the wear process are scarce. Recently, a method that combined surface replication, or moulding, and image analysis to monitor gear wear was proposed in [121], and it allows the morphology of worn gear teeth to be recorded in great detail in the wear process.

This chapter is organised as follow. Section 5.2 elaborate the methodology of the study, and section 5.3 presents results and discussion, followed by summary in section 5.4.

5.2 Methodology

The methodology to investigate the relationship between gear macropits and wear particles is illustrated in Figure 5.1. This section firstly briefly introduces the experimental test and data acquisition in section 5.2.1 and then details image processing and feature extraction to obtain numerical features to characterise macropits and wear debris in section 5.2.2, followed by data analysis to determine the relationship between the features of gear macropits and wear particles in section 5.2.3

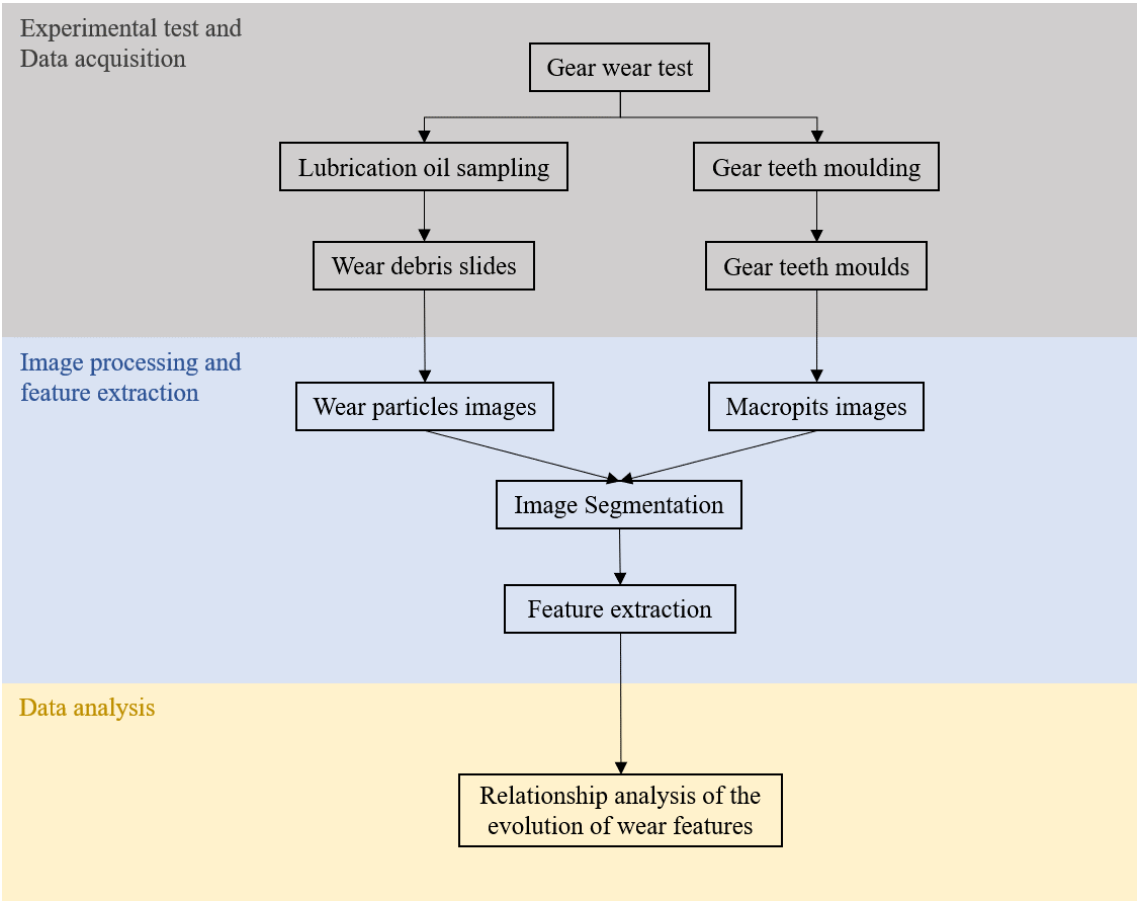


Figure 5.1: Diagram of the methodology to investigate the relationship between gear macropits and wear particles.

5.2.1 Experimental test and data acquisition

The lubricated gear wear test was designed to promote macropitting and conducted on the test rig, as shown in Figure 3.2. The gear material, load, and lubrication were carefully selected to minimise the chance of occurrence of micropitting, scuffing and abrasive wear [20]. Furthermore, in order to accelerate the initiation of macropitting, the two gears were manually roughened before the test. This procedure is based on the established effect of surface roughness on surface fatigue life in rolling-sliding contact, as reported by many experimental and theoretical works (see [151, 178–180]). The test ran for a total of 3.25 million cycles (of the input shaft), and at the end, the driving gear (pinion) was severely

worn by macropitting. Full details about the test can be found in section 3.1.

In order to monitor the gear wear process, lubrication oil samples and gear teeth moulds were collected using the methods detailed in section 3.2. The test was paused 40 times for oil sampling and moulds making. The intervals between two pauses were designed to be denser in the initial stage of the test, and the average interval was 0.08 million cycles. Six gear teeth (3 for the driving gear, 3 for the driven gear) were marked, as shown in Figure 3.3, and six moulds were obtained during each pause.

An oil sample was obtained before the test rig was stopped each time for making gear moulds. The oil sampling point was 1 cm below the oil surface and underneath the gear meshing region, and a 3D-printed guiding channel was used to ensure the accuracy and consistency of the position of the sampling point, as shown in Figure 3.6. After each sampling, an equal amount of clean oil was added to the gearbox to ensure that the oil quantity in the gearbox remained unchanged. A total of 40 oil samples were collected during the test.

Wear debris slides were made from oil samples using the filtergram method [208], in which the membrane filter on the glass slide was cleared for a better contrast between particles and their background so that high-quality images of wear particles could be acquired for quantitative analysis. The membrane filter has a pore size of 3 µm to collect wear particles whose sizes are larger than the pore size. For each slide, the same amount (1 ml) of oil sample was extracted from the oil sample bottle after sonication to ensure the consistency of comparisons. In total, 38 good-quality gear teeth moulds of a driving gear tooth and 10 wear debris slides were used in image acquisition described in the next section. Choosing fewer wear debris slides was because the small size and dispersed state of wear debris on the slide required much more time to image than the moulds and the 10 slides were selected to cover the entire duration of the test with similar intervals.

Gear teeth moulds and oil samples obtained from the gear tests were examined using optical microscopy, as mentioned in section 3.3. The imaging procedure aimed to capture the overall appearance of the pitted region and wear particles so that the concentration and

other numerical features of macropits and wear particles can be extracted. A specially designed holder with an involute profile was designed for the moulds to ensure the consistency of the imaging position. Due to the different scales of macropits on the moulds and wear particles on the slides, different imaging strategies were adopted. For the moulds, a 5x objective lens was used, and 7 images were captured along the face width, which were then stitched into a complete scan (see Figure 5.2) of the pitted region on the tooth flank. For the wear particle slides, a 20x objective lens was used, and 40 images were captured at different locations in a circular ring of the membrane filter. The circular ring was where most particles were evenly distributed, and it had an outer diameter of 11 mm and an inner diameter of 5 mm. In total, the 40 images make up around 20% of the circular ring area. Figure 5.3 shows a sample wear particle image captured by optical microscopy.

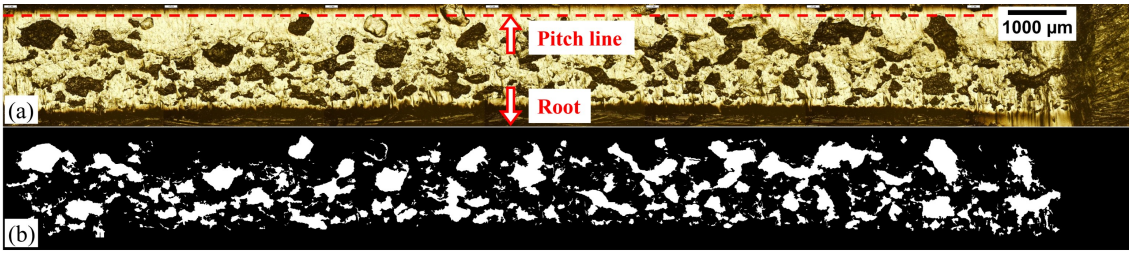


Figure 5.2: (a) A stitched image of the pitted region on a gear tooth mould (macropits are in black) and (b) its segmented binary image.

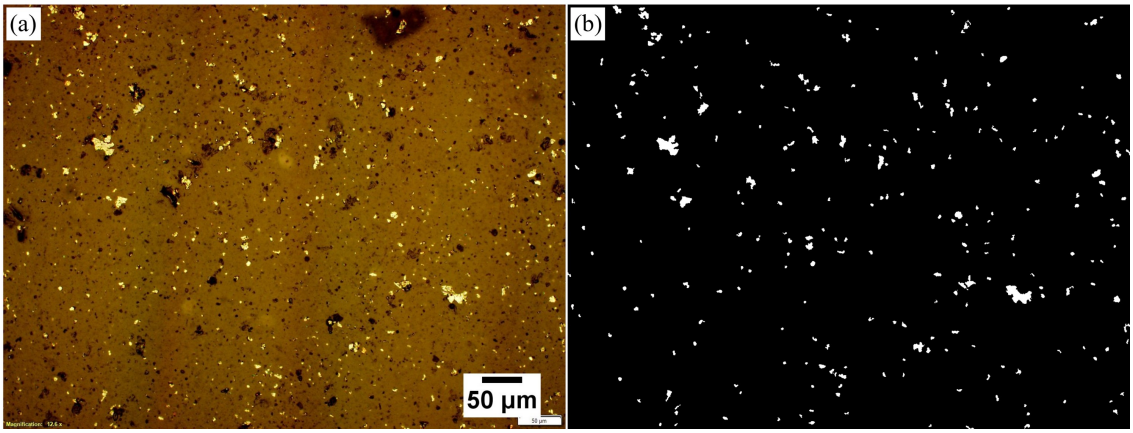


Figure 5.3: (a) A wear particle image captured by optical microscopy (particles are the bright ones) and (b) its segmented binary image.

5.2.2 Image processing and feature extraction

5.2.2.1 Image processing

A series of image processing techniques were applied to transform the raw 2D images into usable forms from which different numerical features can be extracted. The main task was image segmentation that separated either macropits or wear particles from their background and separated connected macropits.

The high contrast between objects of interest (either macropits or wear particles) and the background (see Figure 5.2 (a) and Figure 5.3 (a)) allows the images to be segmented using simple histogram-based thresholding methods. Each channel of the RGB image was assigned a separate threshold using the method, and the three channels were combined into a new binary image. Figures 5.2 (b) and 5.3 (b) show the segmented binary images of the 2D images of macropits and wear particles, respectively.

After segmentation by thresholding, the binary images of wear particles were ready for feature extraction, but those of macropits needed further processing. The reason is that some macropits merged in the wear process, as shown in Figure 5.4 (a) to (b), while discrete wear particles would not be affected by this phenomenon.

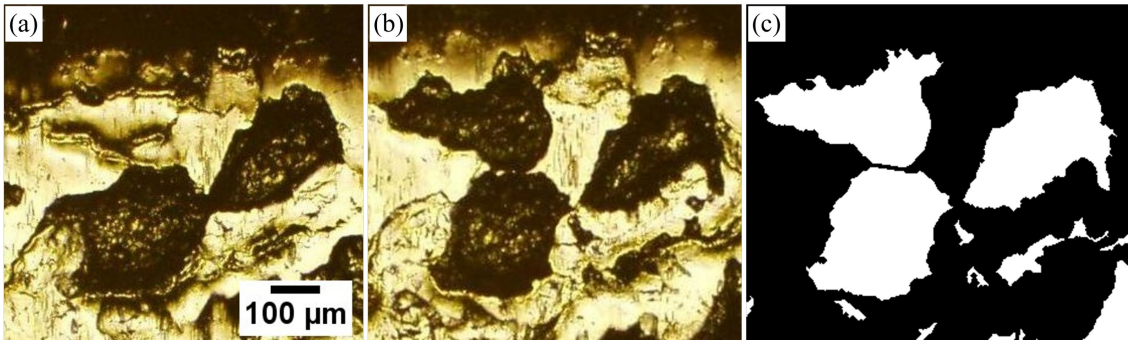


Figure 5.4: An example of a new macropit formed and merged with old macropit; (a) at 0.86 million cycles, (b) at 0.94 million cycles, a new pit formed in the top left region and was connected to pit below, (c) segmented binary image of (b) after marker-controlled watershed segmentation.

The merged macropits could not be separated simply by thresholding, and they will result in wrong numerical features such as area and aspect ratio. To solve this problem, marker-controlled watershed segmentation [123] was applied to the binary images after thresholding. Watershed-based segmentation is often used to break connected or overlapped regions in an image into distinct objects, and the marker-controlled method minimises over-segmentation, which is a common problem associated with other methods of watershed segmentation. Figure 5.4 (c) shows the segmentation result of Figure 5.4 (b) using marker-controlled watershed transformation.

5.2.2.2 Pitted area and concentration of wear particles

The images after image processing, as described in section 5.2.2.1, were then used to extract two measures of macropits' propagation and one measure of the particle concentration in oil. Regarding macropits, two measures of area coverage were defined: (i) the area coverage of macropits $c_{pits}(t)$, calculated as the total area of all pits over the total area of the active tooth flank (see Equation 5.1); and (ii) the area coverage of the pitted region $c_{pittedregion}(t)$, obtained as the overall area of the flank showing pits over the total area of the active tooth flank (see Equation 5.2).

$$c_{pits}(t) = \frac{A_{pits}(t)}{A_{active\,tooth\,flank}} = \frac{A_{pits}(t)}{51.31\text{mm}^2} \tag{5.1}$$

$$c_{pitted\,region}(t) = \frac{A_{pitted\,region}(t)}{A_{active\,tooth\,flank}} = \frac{A_{pitted\,region}(t)}{51.31\text{mm}^2} \tag{5.2}$$

In Equation 5.1 and Equation 5.2, $c_{pits}(t)$ is the area coverage of macropits at time t, and $A_{pits}(t)$ is the total area of macropits in the large mould image of the time t. $A_{active\,tooth\,flank}$ is the area of the active tooth flank which is a constant, and $A_{pitted\,region}$ is the area of the pitted region which is defined as a rectangular area that contains most of the pitted area $A_{pits}(t)$ (99% in this study).

The main difference between the two quantities is that the former only considers the sum of the areas of the discrete macropits, whereas the second is based on the "pitted region", i.e., the continuous region impacted by macropitting. $c_{pitted\,region}(t)$ is always larger than $c_{pits}(t)$, since it includes additional portions of the surface between pits. An example of the difference between the two areas is shown in Figure 5.5.

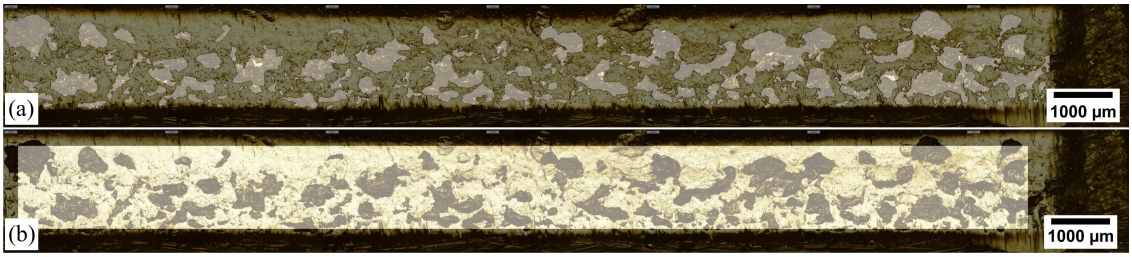


Figure 5.5: (a) Pitted area and (b) pitted region highlighted in a gear tooth mould image.

The second measure is closer in concept to the 'pitted area' used in gear standards [21,209], which describes the spread of macropitting instead of accurately calculating the area of every macropit. Moreover, this measure is less affected by the shrinkage of macropits compared to $c_{pits}(t)$. This phenomenon, as shown in Figure 5.6, can be caused by abrasive wear that gradually make the pits shallower and smaller and plastic deformations of the edges of macropits that made material flow in the pits [210].

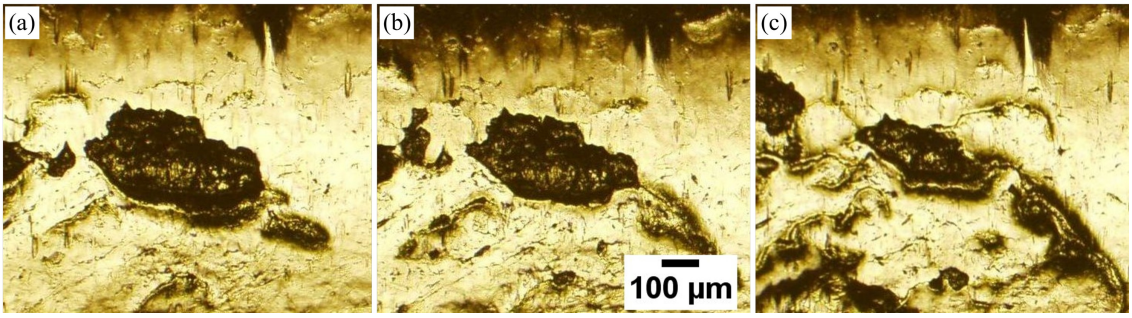


Figure 5.6: An example of shrinkage of macropits in the wear process; (a) at 0.36 million cycles, (b) at 0.42 million cycles, (c) at 0.51 million cycles.

The "derivatives" of $c_{pits}(t)$ and $c_{pitted\,region}(t)$ with respect to time were also defined (see Equation 5.3 and Equation 5.4) using a finite-difference approach in order to investigate

their relationships with particles.

$$\dot{c}_{pits}(t_i) = \frac{c_{pits}(t_{i+1}) - c_{pits}(t_{i-1})}{t_{i+1} - t_{i-1}} \qquad 1 \leq i \leq 37 \tag{5.3}$$

$$\dot{c}_{pitted\,region}(t_i) = \frac{c_{pitted\,region}(t_{i+1}) - c_{pitted\,region}(t_{i-1})}{t_{i+1} - t_{i-1}} \qquad 1 \leq i \leq 37 \tag{5.4}$$

In Equation 5.3 and Equation 5.4, i is the number of gear inspection, t_i is in million cycles, $t_0 = 0$ and both parameters are assumed $c_{pits}(0) = c_{pitted\,region}(0) = 0$.

The concentration of wear particles for the oil sample collected at time t was instead defined as the percentage of area covered by wear particles within the total area of captured images at time t (see Equation 5.5), rather than the number of wear particles per unit volume of oil. Both the pit- and particle-related quantities used in this paper are commonly used in gear teeth inspection [21] and wear debris analysis [95].

$$c_{particle}(t) = \frac{\sum_{i=1}^{40} A_{i,particles}(t)}{A_{sample} \times 40} = \frac{\sum_{i=1}^{40} A_{i,particles}(t)}{14.5\text{mm}^2} \tag{5.5}$$

In Equation 5.5, $c_{particle}(t)$ is the concentration of wear particles for the oil sample collected at time t, $A_{i,particles}(t)$ is the total area of wear particles in the ith sample image (out of 40), and A_{sample} is the area of each sample image which is a constant. A_{pits} and $A_{particles}$ were calculated using the binary images as shown in Figure 5.2 (b) and Figure 5.3 (b).

5.2.2.3 Numerical features of individual macropits and wear particles

The 2D images after processing as described in section 5.2.2.1 were used to extract various numerical features of individual macropits and wear particles, relating to their size, shape, and edge details, as summarised in Table 5.1. Note that the major or minor axis length was extracted from the ellipse whose second moment of area is the same as the object's. A_{box}

and A_{convex} are the area of the object's bounding box, the smallest rectangle containing the object, and the area of the object's convex hull, respectively.

Table 5.1: 2D numerical features [111, 117, 211] extracted from individual macropits and wear particles images.

No.	Numerical features	Formula
1	Area (µm^2)	A
2	Perimeter (µm)	L
3	Equivalent diameter (µm)	$d_{eq} = \sqrt{\frac{4A}{\pi}}$
4	Major axis length (µm)	l_a
5	Minor axis length (µm)	l_b
6	Aspect ratio	$AR = \frac{l_a}{l_b}$
7	Eccentricity	$e = \sqrt{1 - \frac{l_b^2}{l_a^2}}$
8	Extent	$ex = \frac{A}{A_{box}}$
9	Solidity	$s = \frac{A}{A_{convex}}$
10	Circularity	$c = \frac{4\pi A}{L^2}$

5.2.3 Analysis of extracted quantities

The evolution of the area coverage of macropits (c_{pits}, $c_{pitted\,region}$ and their derivatives) and the concentration of wear particles ($c_{particles}$ were compared to identify any relationship. The relationships between the 10 numerical features of individual macropits and wear particles were investigated in three aspects: (i) data distributions, (ii) evolution and (iii) principal components. The analysis used 10,952 observations of wear particles' features (obtained from 10 oil samples) and 12,374 observations of macropits' features (obtained from 38 gear inspections). The purposes and details of the analyses are provided below.

Firstly, the data distributions of the two feature sets (one for macropits and the other for wear particles) were analysed using exploratory analysis, which quantifies the central tendency and variability of the features of macropits and wear particles and provided summary statistics including minimum, maximum, range, mean, median, 25 percentile,

75 percentile, skewness, and kurtosis. Box plots were then used to visualise and compare the results.

Secondly, the evolution of the 10 numerical features of macropits and wear particles in the test were compared to reveal their relationships in the time domain. For observations corresponding to the same time and the same objects (either macropits or wear particles), the average value was obtained and used in time plots. The observations of macropits came from 38 different times, and those of wear particles came from 10 different times.

Finally, the two feature sets were analysed using principal component analysis (PCA). PCA is commonly used in dimensionality reduction for multivariate statistics, and it is an effective method to reveal the driving principles of the variables, i.e. which variables change together [212]. It has been used for wear particle features [213] and other condition monitoring signals [59] but not for macropits. Before PCA, the dataset was standardised, i.e., the raw scores subtracting the sample mean and divided by the sample standard deviation, due to different units and scales of the variables. For example, macropits had the size of usually hundreds of microns while wear particles were of only tens of microns.

5.3 Results and Discussion

In this section, an overview of the evolution of gear macropits and wear particles is presented firstly. Then, the evolution of pitted area and wear particles' concentration are compared, followed by the relationships of their numerical features.

5.3.1 Overview of the evolution of gear macropits and wear particles

For the ease of understanding of the gear wear process, a qualitative comparison between mould images and wear particle images at different stages in the test was carried out. Figure 5.7 shows an overview of the evolution of gear macropits and wear particles and tabulates wear mechanisms of gear teeth (in the order of their severity) and different types

of wear particles present in oil samples (in the order of their proportions). Changes in the severity of different wear mechanisms and proportions of different wear particles with respect to previous ones were also observed qualitatively and annotated near the images. It is worth noting that for each time, 40 wear particle images were used for evaluation and only one of them is presented in the figure.

Figure 5.7: The evolution of gear macropits and wear particles. (a1) – (j1) mould images, (a2) – (j2) wear particle images. Texts summarise wear information, including running cycles, wear mechanisms on gear surface in the order of severity, their changes with respect to previous ones, wear particle types in the order of proportion, their changes with respect to previous ones, and approximate sizes. Note '↑', '↓' and '-' means increase, decrease and no apparent change, respectively.

As shown in the figure, gear macropits started near the root and quickly propagated towards the pitch line before 0.86 million cycles and then the propagation slowed down until the pitted area peaked around 2.17 million cycles when macropits occupied the majority of the dedendum. Afterwards, the propagation stabilised, but the shrinkage of

pits was noticeable. Meanwhile, the concentration of wear particles in the oil samples quickly increased before 0.86 million cycles and peaked around 1.99 million cycles, after which the concentration decreased significantly. Moreover, it can be observed that the darkness of oil samples followed a similar trend to wear particle concentration.

5.3.2 The evolution of pitted area and wear particles' concentration

Figure 5.8 (a-b) show the evolution of the area coverage of macropits and pitted region (c_{pits} and $c_{pitted\,region}$) and of the wear particles' concentration. It can be observed from the two figures that the growth trend of either of the two parameters of pitted area was evidently different from that of wear particles' concentration, especially after 2 million cycles. However, both plots show a three-stage evolution of pitted area, in which both parameters quickly rose before 1 million cycles, then peaked around 2 million cycles, and fluctuated or stabilised afterwards. After around 2.2 million cycles, there was a substantial decrease in c_{pits} while $c_{pitted\,region}$ was more stable, which shows the use of pitted region was effective to minimize the impact of macropits shrinkage.

The increase of macropitting to its peak could be driven by two factors. (i) The cyclic condition of lubricated rolling-sliding contacts possibly caused the ongoing initiation and propagation of macropitting. (ii) The roughened gear surfaces likely caused localised stress concentrations and thus could accelerate the initiation and propagation of macropitting in the early stage. Then, the stabilisation and decline of macropitting could be caused by three mechanisms: (i) the flattening of the involute profile after around 2 million cycles (see previous work in [121]), possibly increasing the nominal contact area; (ii) abrasive wear continuously removing material and making pits shallower and smaller; and/or (iii) increased localized contact stress around pits causing plastic deformations of their edges, which can flow into the macropits.

Figure 5.8 (c-d) compares the wear particle concentration with the derivatives of the two area-coverage measures of macropitting. The plots of the derivatives have been smoothened with a 3-time-step moving average. It can be observed that while parti-

cle concentration did not align well with the derivatives of pit measures, the alignment became much better with a shift of 0.85 million cycles along the horizontal axis.

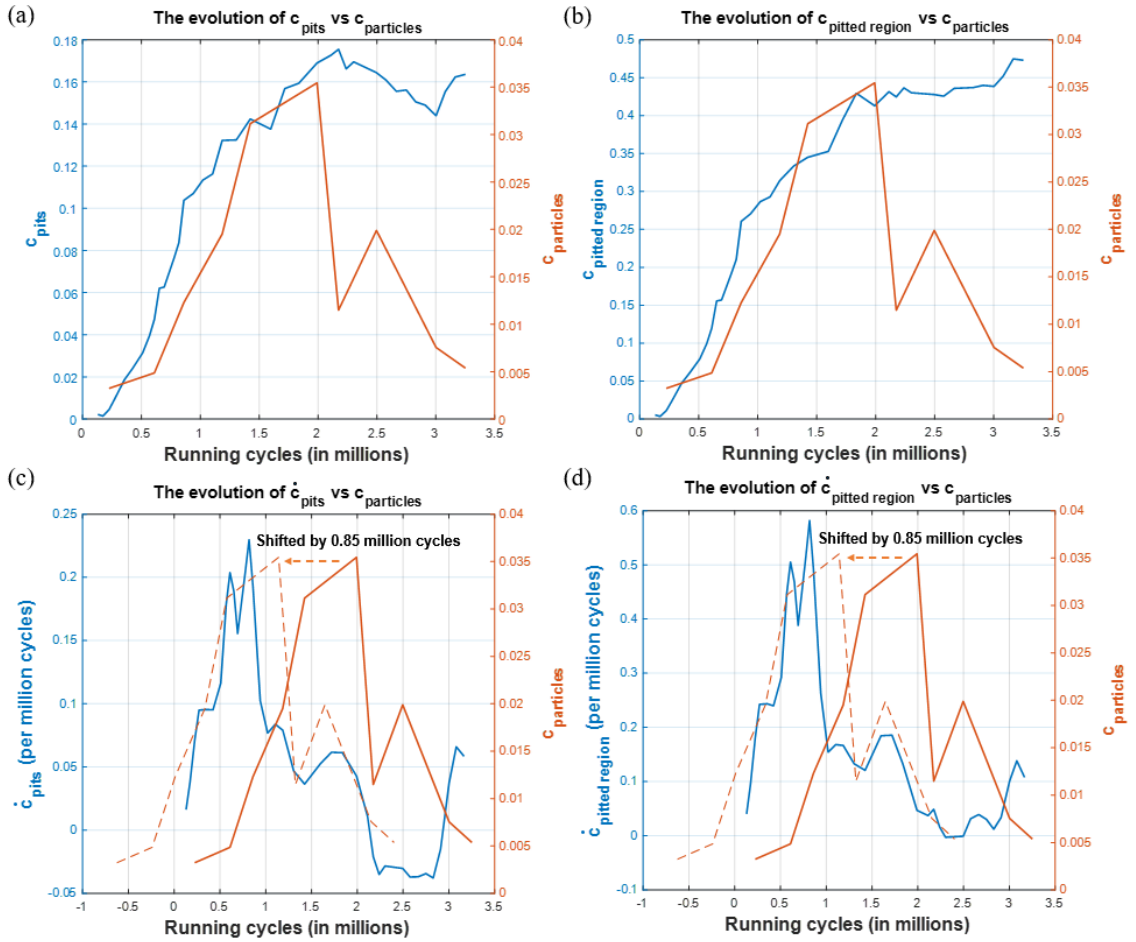


Figure 5.8: The evolution of the pitted area and wear particles' concentration in the test; (a) the evolution of c_{pits} vs $c_{particles}$; (b) the evolution of $c_{pitted\,region}$ vs $c_{particles}$; (c) the evolution of $\dot{c}_{pits}$ vs $c_{particles}$; (d) the evolution of $\dot{c}_{pitted\,region}$ vs $c_{particles}$. The plots for $\dot{c}_{pits}$ and $\dot{c}_{pitted\,region}$ were smoothed using a moving average window of 3 points.

A possible explanation for the superior alignment of the derivatives is that wear particles in each oil sample are expected to be mostly new particles that were just released into the oil. This is justified under the assumption that most history particles (especially in a large size) have settled and therefore not captured at the oil sampling location (see section 5.2.1 for the oil sampling procedure). Moreover, particles smaller than 3 microns are not likely to be included in the analysis due to the filter's pore size and imaging resolution.

This could also contribute to the "removal" of historical particles from samples, assuming that recirculated particles are likely to be ground into a smaller size.

However, the delay between the propagation of macropitting and wear particles' concentration was a surprising finding, especially considering the above hypothesis on the removal of historical particles. A possible explanation is that the low hardness of the gears used in the test allowed primary wear particles to be pressed and entrained into the contact surfaces [10, 103, 214], as apparent in Figure 5.9, and released in the oil later.

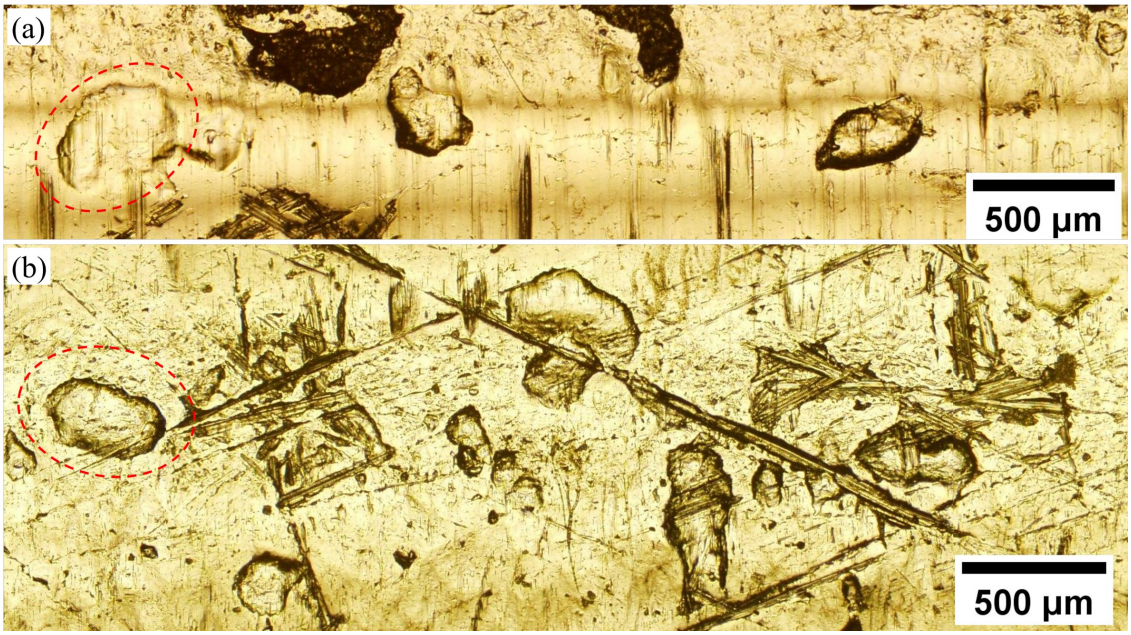


Figure 5.9: Dents on the tooth surfaces that have similar sizes and shapes to macropits; examples are highlighted with red dashed circles; (a) the tooth surface of the driving gear at 2.3 million cycles; (b) the tooth surface of the driven gear at 0.94 million cycles.

5.3.3 Relationships between numerical features of individual macropits and wear particles

The investigation results of the relationships between numerical features of individual macropits and wear particles are detailed in three aspects in this section, which are the data distributions, evolution, and principal components of the numerical features.

5.3.3.1 Data distributions of the numerical features

Figure 5.10 visualises the descriptive statistics of the numerical features of macropits and wear particles, and the complete comparison of the statistics is shown in Table 5.2.

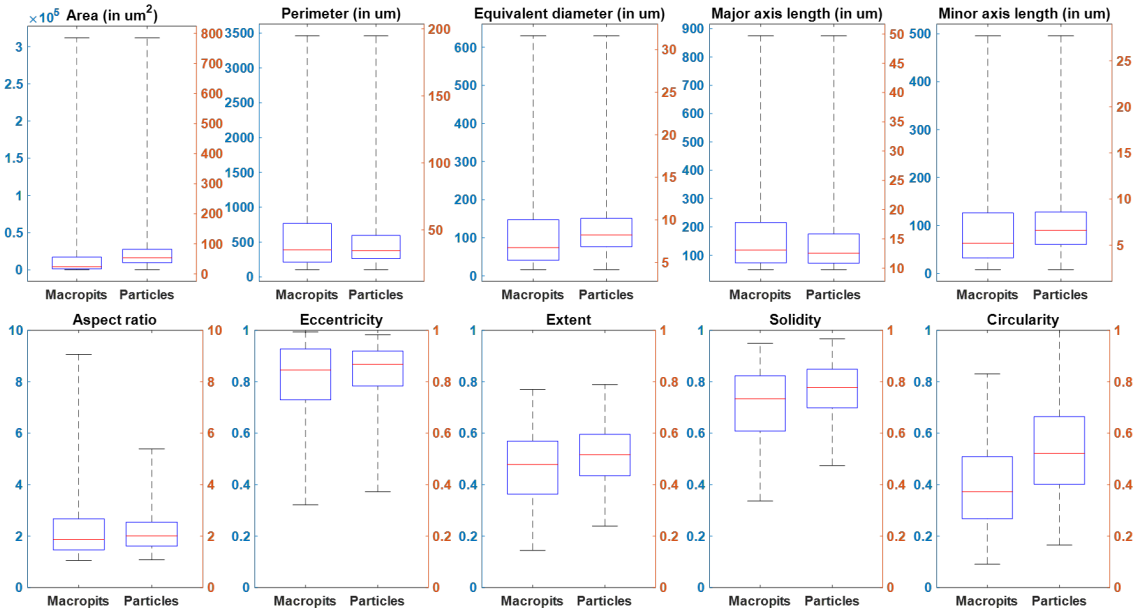


Figure 5.10: Box plots of the numerical features of macropits and wear particles.

Table 5.2: Descriptive statistics of the numerical features of individual macropits and wear particles.

	Minimum		Maximum		Range		Mean		Median		25 percentile		75 percentile		Skewness		Kurtosis	
	M	P	M	P	M	P	M	P	M	P	M	P	M	P	M	P	M	P
Area (μm^2)	216.6	13.6	3.1E5	785.3	3.1E5	771.7	1.7E4	74.9	4.3E3	53.4	36.9	1.3E3	1.7E4	81.4	4.1	4.2	23.3	26.8
Perimeter (μm)	104.2	20.4	3.5E3	194.8	3.4E3	174.4	580.5	41.3	387.9	34.6	28.5	211.7	767.9	45.9	2.0	2.8	7.9	13.6
Equivalent diameter (μm)	16.6	4.2	630.1	31.6	613.5	27.5	111.0	9.1	74.3	8.2	6.9	41.1	147.5	10.2	1.9	2.2	7.3	10.3
Major axis length (μm)	49.3	9.7	874.6	49.7	825.3	40.0	166.3	14.4	118.0	12.5	10.8	73.1	215.9	15.8	1.9	2.5	7.4	11.3
Minor axis length (μm)	7.9	2.3	495.7	27.7	487.8	25.4	92.6	7.3	62.9	6.6	5.1	32.5	126.2	8.6	1.8	1.9	6.7	8.6
Aspect ratio	1.1	1.1	9.1	5.4	8.0	4.3	2.3	2.2	1.9	2.0	1.6	1.5	2.7	2.5	2.0	1.2	7.7	4.6
Eccentricity	0.3	0.4	1.0	1.0	0.7	0.6	0.8	0.8	0.8	0.9	0.8	0.7	0.9	0.9	-0.9	-1.4	3.3	4.8
Extent	0.1	0.2	0.8	0.8	0.6	0.6	0.5	0.5	0.5	0.5	0.4	0.4	0.6	0.6	-0.2	0.0	2.3	2.4
Solidity	0.3	0.5	0.9	1.0	0.6	0.5	0.7	0.8	0.7	0.8	0.7	0.6	0.8	0.8	-0.5	-0.4	2.4	2.5
Circularity	0.1	0.2	0.8	1.0	0.7	0.8	0.4	0.5	0.4	0.5	0.4	0.3	0.5	0.7	0.4	0.3	2.4	2.4

In the box plots, the upper and lower whisker represents the maximum and minimum value, respectively, and the red line is the median value. The upper and lower edge of the box represents 25 and 75 percentile value, respectively. Figure 5.10 shows that macropits are much larger than wear particles regarding their 2D sizes, as reflected by box plots of all the dimensional features. In terms of the non-dimensional features, the data distributions of macropits are comparable to those of wear particles. For all features, similar central tendency and variability were observed between macropits and wear particles, as reflected in the relative size and position of boxes in relation to the whiskers.

5.3.3.2 The evolution of the numerical features

Figure 5.11 shows the evolution of 6 numerical features (out of 10) of macropits and wear particles in the test. Firstly, it can be observed from Figure 5.11 (a) and (b) that the area and major axis length of macropits and wear particles evolved with notably different trends, and applying a time shift of -0.85 million cycles to the plots of wear particles (indicated by dashed lines, as done for the concentration results) did not improve the result. The poor correlations were also observed for the remaining dimensional features, including perimeter, equivalent diameter and minor axis length (not reported for the sake of brevity). On the contrary, Figure 5.11 (c)-(f) show similar variations between macropits and wear particles in terms of 4 non-dimensional features, including aspect ratio, extent, solidity, and circularity, if a time shift of -0.85 million cycles is applied to the plots of wear particles. The fact that the same delay between pits and particles is observed in the concentration study (section 5.3.2) and these 4 non-dimensional features provides additional support to this finding. It also suggests that non-dimensional features of particles might have more diagnostic value than dimensional geometrical quantities.

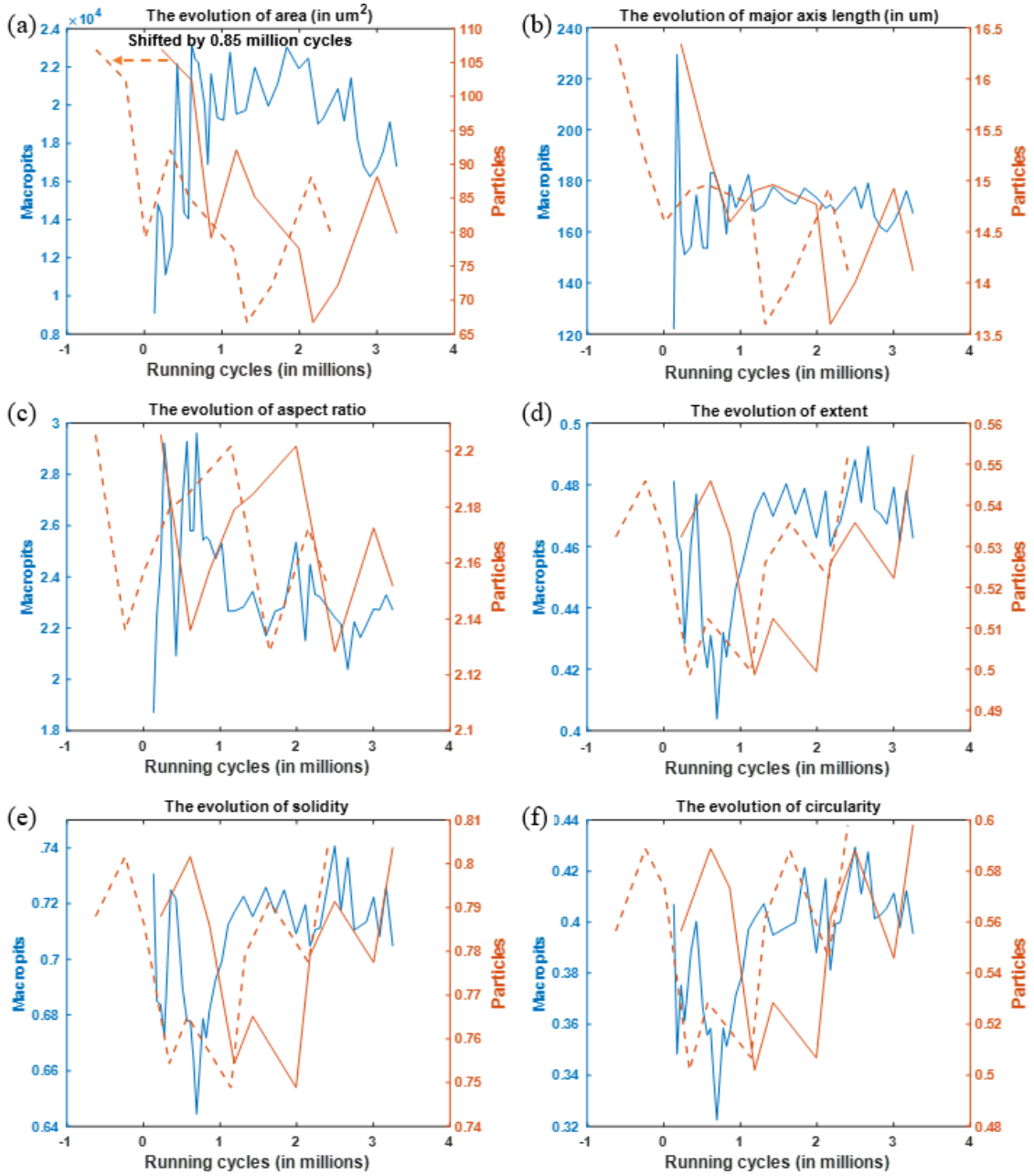


Figure 5.11: The evolution of 6 numerical features of macropits and wear particles: (a) area, (b) major axis length, (c) aspect ratio, (d) extent, (e) solidity, and (f) circularity. The solid lines represent actual values, and the dashed orange lines represent time-shifted values (all by -0.85 million cycles) with respect to the solid orange lines.

5.3.3.3 Principal components of the numerical features

The principal components of the two feature sets were analysed and visualised using PCA as described in section 5.2.3. Figure 5.12 show the PCA results of the numerical features of macropits and wear particles, including the number of principal components obtained and the visualisation of their relationships.

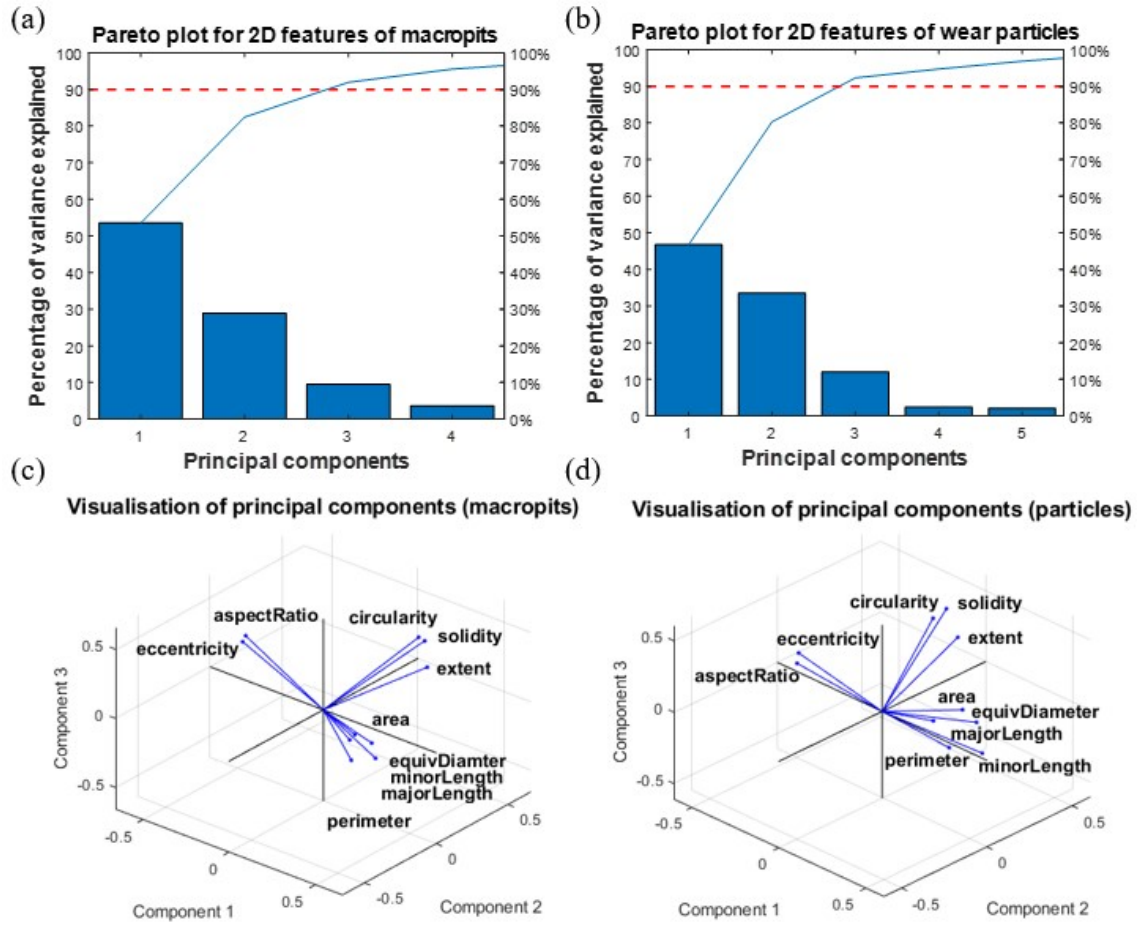


Figure 5.12: PCA results of 2D features; (a-b) Pareto plots of macropits and wear particles, respectively; (c-d) 3D biplots of principal components of macropits and wear particles, respectively.

It can be seen from the pareto plots in Figure 5.12 (a-b) that both macropits and wear particles needed 3 principal components to account for more than 90% of the variance. Figure 5.12 (c-d) show the principal component coefficients of the 3 components in 3D

space, and it is evident the 2D features had the same grouping with a clear physical meaning for macropits and wear particles. The 3 groups are (i) size features, including area, perimeter, equivalent diameter, major axis length, and minor axis length, (ii) overall shape features, including aspect ratio and eccentricity, and (iii) edge-detail features, including circularity, solidity and extent. The results suggested that the 10 numerical features of macropits and wear particles had very similar principal components.

5.4 Summary

A novel and integrated method was applied to investigate both qualitative and quantitative relationships between gear macropits and wear particles in a lubricated gear test. The method adopted the technique of moulding and imaging to acquire accurate and detailed wear information on gear teeth and effectively solved the scarcity of tribological data of gear teeth during the wear process. Sampled macropits and wear particles were imaged using optical microscopy, and the 2D images were used to extract concentration and 10 numerical features of macropits and particles related to their size, shape, and edge details. The evolution of the pitted area and macropits' concentration were compared. For numerical features, their data distribution, evolution, and principal components were compared. In general, this study suggests that using only the concentration of wear particle is insufficient for timely gear wear monitoring. The most relevant findings of the investigation are summarised as follow.

- The evolution of the concentration of wear particles showed a similar trend to that of the pitted area derivative but with a strong delay ($\sim$ 0.85 million cycles).
- The evolution of 4 non-dimensional features, including aspect ratio, extent, solidity and circularity, of wear particles and macropits showed similar trends if we consider the same delay ($\sim$ 0.85 million cycles).
- PCA shows strong covariance of most non-dimensional features, suggesting that only a few are needed for surface/particle characterisation.

This study thus encourages an interesting research direction for future work. An investigation of the reasons for the observed delay would be aided by conducting gear tests with different configurations (gear materials, speeds, loads, etc.). This would also contribute to a more in-depth analysis of the detailed information conveyed by the non-dimensional features discussed in this paper.

Chapter 6

Automated Gear Wear Mechanism and Severity Assessment Using Images and a Deep Learning Approach

This chapter addresses the third objective of this thesis, i.e., **to develop an automated system for gear wear mechanism and severity assessment**, by developing convolutional neural networks (CNNs) to identify gear mould images and wear particle images of different wear mechanisms and severity levels. This work is built upon the moulding-imaging method developed in Chapter 4, which allowed easy acquisition of gear mould images that contain rich and detailed surface features for wear assessment. The potential of mould images being utilised for gear diagnostics is investigated and compared with wear debris images using the same deep learning approach.

This chapter expands on the following published journal article in which only mould images were used for gear wear assessment,

- **H. Chang**, P. Borghesani, and Z. Peng, "Automated assessment of gear wear mechanism and severity using mould images and convolutional neural networks," *Tribology International*, vol. 147, p.106280, 2 2020.

The expansion aimed to investigate the feasibility of CNNs in classifying wear particle images into different wear mechanisms and severity levels, based on the validated performance of CNNs for mould images in similar classification tasks in the article above, and a similar methodology was adopted for the classification of wear particle images.

Abstract

A novel methodology for automated wear mechanism and severity assessment combining surface replication, imaging and deep learning is proposed. A large dataset of images of gear teeth moulds was built and covers abrasive wear, macropitting and scuffing, and three severity levels for each mechanism, i.e., mild, moderate and severe. Wear debris image datasets were also built and cover four types of wear particles, including rubbing, laminar, fatigue, and cutting, and three severity levels. A two-level inference methodology for mould images was designed, based on a first convolutional neural network (CNN) for wear mechanism identification, followed by three CNNs for wear severity estimation. The first level obtained a test classification accuracy of 98.22% and the second of 95.16% on average. The two-level system was also applied to full flank mould images to generate wear mechanism and severity maps showing the geographical distribution of wear. The particle classification CNN achieved a test classification accuracy of 88.97%, and the severity CNN using wear debris images scored 90.11%.

6.1 Introduction

The assessment of wear mechanisms and severity plays an important role in gear wear monitoring because the information is critical for engineers to analyse the root cause of

wear-related fault and to determine when the maintenance should be carried out based on the severity of the fault [78, 114, 215]. As reviewed in Chapter 2, current MCM techniques are mainly for fault detection and diagnostics, with some having capabilities in assessing wear mechanism and severity. Specifically, vibration and acoustic emission (AE) techniques are primarily for localised fault detection and diagnosis rather than distributed wear monitoring [5, 56, 78]. Oil and wear debris analysis (OA/WDA) is a more direct and effective technology for wear analysis compared to vibration and AE and has shown capabilities in wear mechanism and severity assessment based on oil features such as viscosity and acidity [47] and wear debris features such as size, quantity, morphology, and colour [47, 112, 216]. However, discrete wear debris features are inadequate to reconstruct a gear tooth surface, and its assessment accuracy can be affected by deformation of wear debris, accumulation of historical particles, exogenous contaminants, etc [27, 47, 134].

A conventional method to get a direct and reliable assessment of gear wear mechanism and severity is to visually inspect a tooth surface, possibly with the aid of cameras for enhanced magnification [7, 78]. In this case, the appearance of the tooth surface is compared against gear standards by trained analysts to identify the presence of specific wear mechanisms and their severity. However, there are three problems involved in this vision-based method, and they are (i) high cost of employing trained analysts, (ii) subjectivity associated with human judgement, and (iii) limited capability of human experts performing such tasks [7, 112, 216, 217]. The third problem refers to the limited resolution of the human eye or portable optical instruments (e.g. cameras) [121, 218]. Micro-scale observations require gears to be disassembled and sent to high-resolution imaging facilities such as scanning electron microscopes (SEM) and often require the teeth to be cut to get cross-sectional views [7, 121]. The cost involved in this process and the destructive nature of thorough optical surface analysis (tooth removal) often result in researchers and engineers accepting low-resolution inspections. Based on previous gear wear research that employed surface replication of gear teeth [66–68, 70], a feasible solution of combining surface replication and image analysis has recently been proposed for gear wear investigation [121]. In that methodology, surface replication, or moulding, is applied to gear teeth with a short pause in the gear operation but without disassembling the gearbox. The moulds are then

analysed with high-resolution optical and laser-scanning microscopes allowing the reconstruction of both micro-level features and macro-level profile changes. In [121], silicone polymers compounds were found to retain all the necessary information to reconstruct both microscopic and macroscopic wear symptoms and image analysis techniques were combined to extract comprehensive high-resolution wear information of gear teeth.

The rapid development of machine learning, especially deep learning, offers a promising solution to address the first two issues mentioned above associated with manual classification of gear teeth images, which has succeeded in many similar applications such as pattern recognition and speech recognition [63, 132]. Conventional machine learning that starts with feature extraction has shown capabilities in image-based wear assessment combined with image processing [111, 112, 114, 126, 219–224]. For wear analysis, most of the studies extracted features from pre-processed images of worn surfaces or wear particles and used them in classifiers such as support vector machines (SVM) to classify images into different classes of wear mechanism and severity. In [6], three features obtained from SEM images of wear particles that underwent discrete wavelet transformation (DWT) were used in an SVM to classify wear particles into abrasive, adhesive, and fatigue particles. In [114], a descriptor called Histogram of Oriented Gradient (HOG) was extracted from images of abrasive wear of a cast iron specimen and used in various classifiers to classify images into mild and severe wear. Indeed, these models automated wear mechanism and severity assessment, but the extracted features were carefully engineered to work on specific wear mechanisms and severity levels. It can be imagined that it would be very difficult to design a complete set of features that cover all wear mechanisms and severity levels. Furthermore, it has been shown that the feature extraction method is limited in processing raw images, and a series of manually tuned preprocessing steps are often needed [63].

Deep learning, however, enables computers to learn intricate patterns in large datasets easily and match or outperform human experts in many tasks [63, 225]. These tasks are usually easy to perform by humans, but expressing the rules mathematically can be very difficult. Examples include the high accuracies of deep convolutional neural networks (CNNs) in large-scale image classification tasks that only well-trained human experts

could match [226], the low error rates of speech recognition achieved by deep recurrent neural networks (RNNs) that benefit many people every day [63]. Deep learning enables computers to learn representations of large datasets and requires significantly less engineering knowledge than conventional machine learning [63, 176]. The learning process of machines is essentially performed by iterative adjustments of multiple internal parameters (weights and biases) attached to multiple layers of information abstraction, with training information flowing from a set of known desired network outputs back towards the entry point of the network (backpropagation) [63, 225]. Deep learning is also increasingly being used in machine condition monitoring based on vibration [227–229], AE [230, 231] and WDA [128, 129, 134]. CNNs have also been used to classify [128, 129] or deblur [134] wear particle images. Wang et al. [128] integrated a back-propagation neural network and a CNN for particle classification and achieved an average classification accuracy of 83%. Peng et al. [128] achieved a higher classification accuracy (93.89%) for ferrography particle images using a 1D CNN with dropout operation. However, to the best of the authors' knowledge, no study so far has proposed the development of deep learning techniques to process raw surface images for automated wear mechanism and severity assessment.

Thus, this study aims to investigate the capabilities of deep CNNs in identifying gear wear mechanism and their severity based on raw image inputs obtained from moulds of gear teeth and wear particle samples. The rest of this chapter is organised as follow. Section 6.2 details the methodology of this study, in which the differences in handling mould images and wear particle images are made clear, and section 6.3 presents results and discussions regarding the wear assessment performance of the CNNs, followed by section 6.4 summarising the developments and key findings of this chapter.

6.2 Methodology

In this study, the capabilities of deep CNNs in processing raw images (i.e., mould images of gear teeth and wear particle images) for automated gear wear mechanism and severity assessment are investigated. As mentioned in section 6.1, a deep learning method is chosen

due to its strong capability to learn intricate patterns in large datasets without feature engineering. Among various deep neural networks, CNNs have shown better performance in image classification than other neural networks [63, 225] because they consider spatial correlations among image pixels and use shared weights to ensure that the classification result will not be affected by the feature' location.

In this study, deep CNNs are trained on large datasets of mould images to establish a two-level inference system in which the first level identifies the dominant gear wear mechanism(s) if any and the second level assesses the severity level of identified wear mechanism(s), as shown in Figure 6.1.

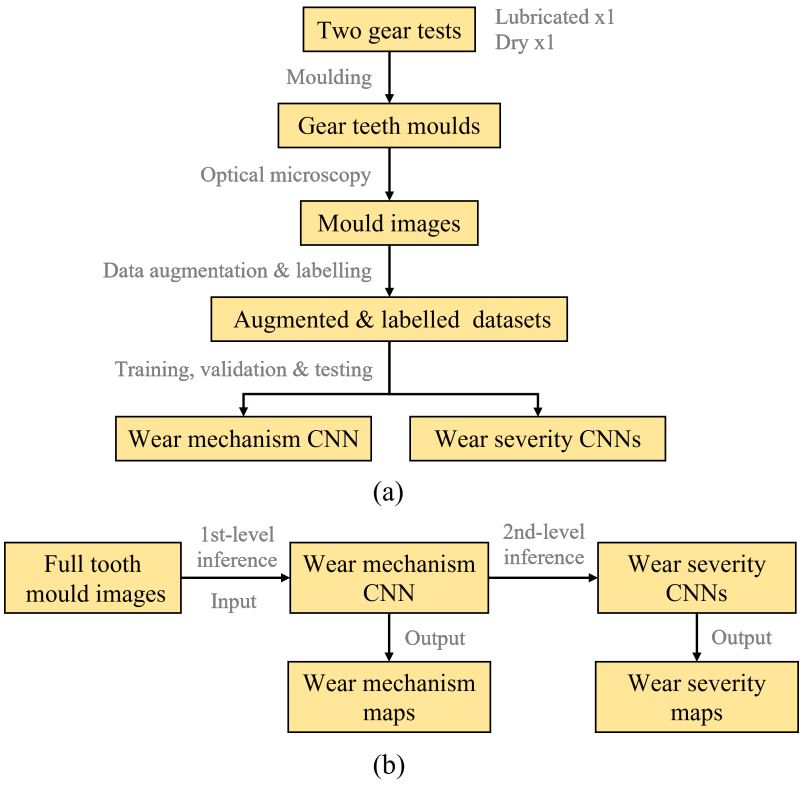


Figure 6.1: Diagram of the methodology for automated gear wear assessment using mould images. (a) Flow of gear wear data from gear tests to usable datasets for the wear assessment CNNs. (b) Integrated testing of the two-level inference system using full tooth mould images.

Then, deep CNNs with similar architecture are trained on wear particle images where

images of individual particles are used for particle classification, and images of group particles are used for wear severity assessment, as shown in Figure 6.2. The following sub-sections will provide methodological details on the acquisition of the datasets, the procedures implemented for data augmentation and labelling, the architecture, training, validation, and testing of the CNNs.

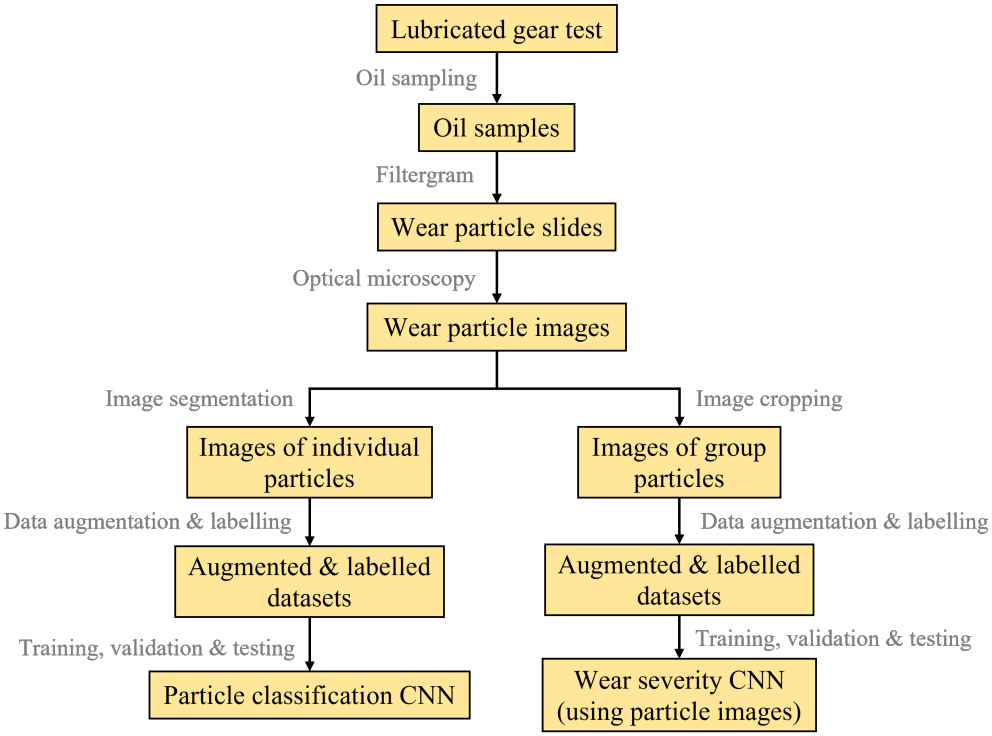


Figure 6.2: Diagram of the methodology for automated gear wear assessment using wear particle images.

6.2.1 Data acquisition from gear tests

In order to train the CNNs for automated gear wear mechanism and severity assessment, mould images of worn gear teeth were obtained from the two gear tests (the lubricated test and the dry test) of this research, and wear particle images were obtained from only the lubricated test. In the lubricated test that ran for 3.25 million cycles, gears were severely worn by macropitting and abrasive wear, while in the dry test that ran for 54.6 thousand cycles, gears were severely worn by scuffing. The lubricated test was paused 40

times for moulds making and oil sampling, and the dry test was paused 11 times only for moulds making. Detailed information on the tests, including the test rig, experimental parameters, and operating conditions, can be found in section 3.1.

The method of moulding and imaging developed in Chapter 4 was used to collect gear mould images from two gear tests. During each pause in the tests, six gear teeth moulds were obtained, corresponding to 6 marked teeth (3 on the driving gear and 3 on the driven gear). Then, the moulds were scanned along the face width using optical microscopy to get 2D images covering the entire dedendum or addendum. Figure 6.3 shows two sample mould images in the lubricated test. Each large mould image ($\sim$ 9400 pixels x 1000 pixels) covered a rectangle area of around 20 mm x 2 mm, which was obtained by stitching seven images (each has the size of 1360 pixels x 1024 pixels) captured using a 5x objective lens. Each pixel in the images is equivalent to 2.04 µm on gear moulds. In total, a sufficient number of 2D large mould images (165 images) were obtained for the training, validation, and testing of the wear mechanism CNN and three wear severity CNNs.

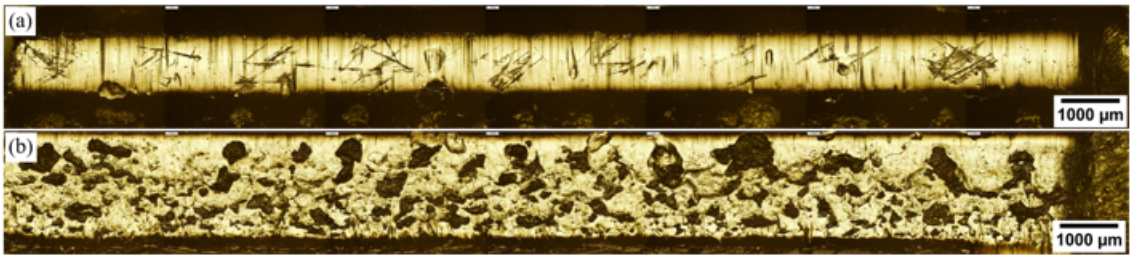


Figure 6.3: Examples of gear teeth mould images in the lubricated test (at 3.25 million cycles). (a) Addendum. (b) Dedendum.

The wear particle images were obtained from lubricating oil samples in the lubricated test (see details in section 3.2 and 3.3). Wear particles from the dry test were not used in CNN training because dry particles (many larger than 1 mm) are much larger than lubricated particles (usually smaller than 50 µm), which cannot be fitted using the image size for lubricated particles. Moreover, conventional particle classification approaches (e.g., image processing and feature extraction) can be used to classify those dry wear particles due to their distinct features (e.g., large sizes, rough surfaces). Due to the small sizes of wear

particles, a 20x objective lens was used to capture their images using optical microscopy. It was observed that wear particles were evenly distributed within a circular ring (outer diameter: 11 mm, inner diameter: 5 mm) on the membrane filter. Therefore, 40 images were captured at different locations within the ring for each slide, and the total imaged area accounted for around 20% of the ring area. In total, 10 wear debris slides were imaged, and 400 wear particle images were obtained via optical microscopy. The size of each image is 1360 pixels x 1023 pixels, with a scale of 0.51 µm/pixel, and Figure 6.4 (a) shows an example of captured wear particle images.

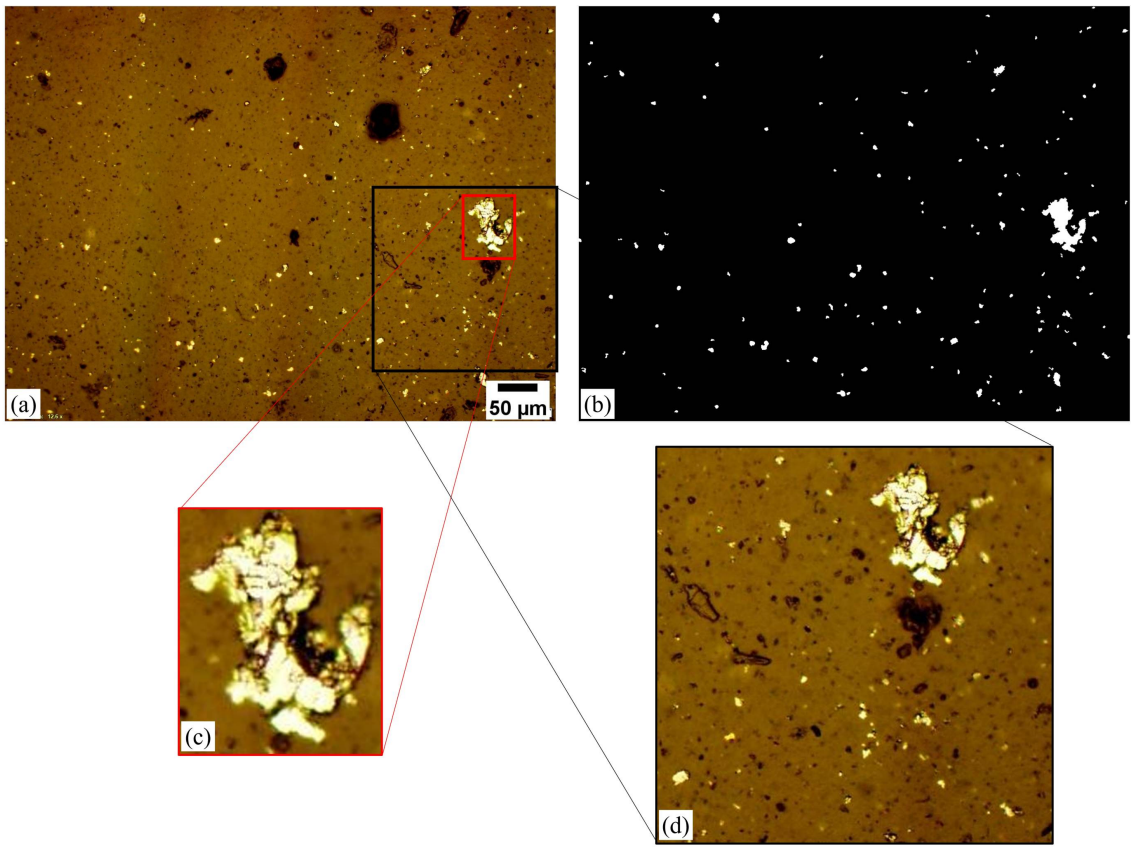


Figure 6.4: An example of wear particle images obtained by optical microscopy. (a) The raw image. (b) The segmented binary image. (c) The image of an individual particle cropped from (a). (d) The image of group particles cropped from (a).

The wear particle images were further processed into images of individual particles and group particles for particle classification and severity assessment, respectively.

- Training a CNN for image classification requires the inputs to be categorised according to their types (i.e., labelling). Therefore, the raw images that contain different types of wear particles need to be broken down into smaller images where each image only contains one particle. Image segmentation was firstly applied to the raw images (e.g., Figure 6.4 (a)) to separate wear particles from the background, outputting binary images, as shown in Figure 6.4 (b). A histogram-based method was used for segmentation thanks to the high contrast between wear particle and the background. The binary images enable wear particles to be indexed by their locations separately, and the location information was used to crop smaller images of individual particles from the raw images, as shown in Figure 6.4 (c). Then, the cropped images were padded with zeros to the same size (200 pixels x 200 pixels) to become valid inputs for the particle classification CNN. In total, 5,119 images of individual particles were obtained to train, validate and test the particle classification CNN.

- Collective features of wear particles such as concentration and size distribution are commonly used for wear severity assessment. Therefore, images of group particles (each image has a size of 450 pixels x 450 pixels) were cropped from the raw images without overlapping for the wear severity CNN, as shown in Figure 6.4 (d). The raw image was not used in its entirety due to its large size and rich details, and it requires a CNN with a significant capacity to process, which usually also needs a high number of training samples to avoid overfitting. Cropping the raw images into smaller images can address both issues simultaneously by reducing the image size and increasing the number of images. In total, 2,316 images of group particles were obtained for the training, validation, and testing of the wear severity CNN.

6.2.2 Data augmentation and labelling

This section details the augmentation and labelling of the image samples (section 6.2.2.1 for mould images and section 6.2.2.2 for wear particle images) acquired in the gear tests, which converted those samples to usable data for the CNNs.

6.2.2.1 Data augmentation and labelling of mould images

Deep neural networks require a high number of training samples (e.g., thousands to hundreds of thousands) to avoid the problem of overfitting. In order to generate such a large training dataset, the original mould image set (165 large mould images) was augmented by cropping and in-plane rotation. Despite the original images having a similar nominal size ($\sim$ 9000 pixels x 1000 pixels), the curvature of the addendum and dedendum at different stages of wear resulted in different sizes of the usable area (e.g. the illuminated portion of Figure 6.3). Therefore, the minimum height of such a usable area (450 pixels) was chosen to crop the original images into small samples (450 pixels x 450 pixels). Rather than splitting the image in non-overlapping square samples, a stride of 50 pixels was used both horizontally and vertically to obtain a higher number of samples. This choice, a commonly used method to augment image data in deep learning [176], also ensured that surface features were captured by at least one 450 x 450 sample in their entirety. Note that, out of the 165 large mould images, 16 were used to produce test data, while the remaining 149 were used to produce datasets for training and validation. Images in the dataset were then labelled, i.e. manually classified into different groups according to their dominant wear mechanism and severity level. Specifically, the small mould images were classified into four wear mechanism groups: normal, abrasive wear, macropitting, and scuffing; and each image in the three abnormal groups was then classified into three severity levels: mild, moderate, and severe. When the number of samples for some wear mechanisms or severity levels was notably smaller than others, in-plane rotations were applied to the corresponding image subsets to increase their numbers. The composition of the datasets after augmentation and the two-level classification is shown in Table 6.1.

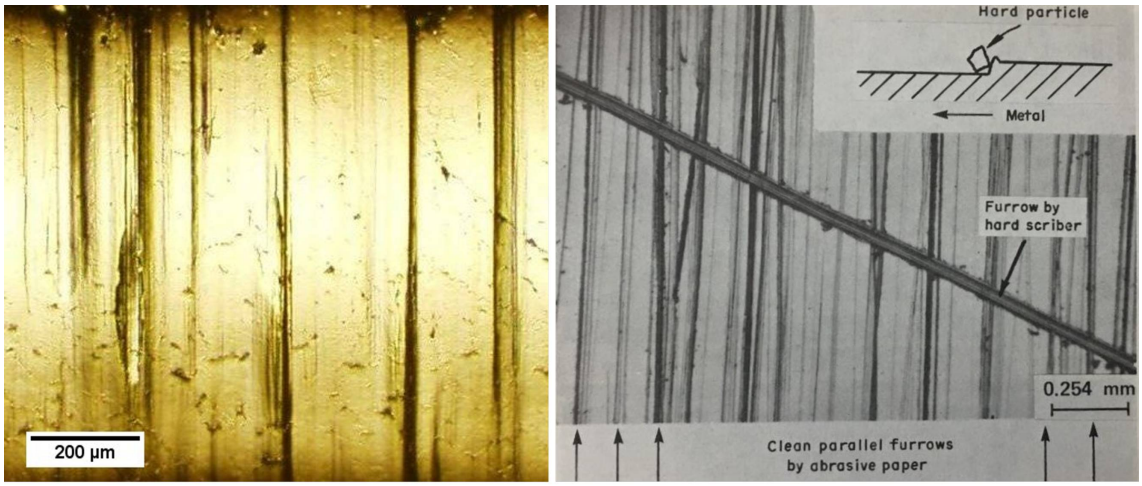
Table 6.1: Categories of the labelled mould image datasets (listed in (training data + validation data)/test data, in the number of images).

Wear mechanisms	Mild	Moderate	Severe	Total
Normal		N/A		25,941/2,150
Abrasive wear	10,304/1060	4,194/499	2,717/216	17,215/1775
Macropitting	10,234/1154	12,403/1,300	14,130/999	36,767/3,453
Scuffing	2,377/488	2,111/545	24,905/1,974	29,393/3,007
Total				109,316/10,385

In order to minimise subjectivity and ensure consistency, the labelling was performed according to established guidelines (both in the form of descriptions and images of worn surface appearance) taken from gear standards, peer-reviewed journal articles, and books. Firstly, the American standard ANSI/AGMA 1010-F14 [21] that provides general descriptions and sample images of different wear mechanisms was followed. It is worth noting for terminological accuracy that in this study macropitting and scuffing are referred to as wear mechanisms, even if they appear as "gear failures" in the nomenclature of the standard. To ensure an unbiased reference for the labelling, additional descriptions and images from peer-reviewed journal articles and books were also used [7,22,28,29,232–234]. Table 6.2 summarises the main features of the three wear mechanisms used in this work to label the image dataset [21]. In addition to comparison to literature cases, the knowledge of the historical development of the two tests generating the samples was used to validate and further correct the labelling of images. Figure 6.5, Figure 6.6, and Figure 6.7 show sample labelled mould images and their references.

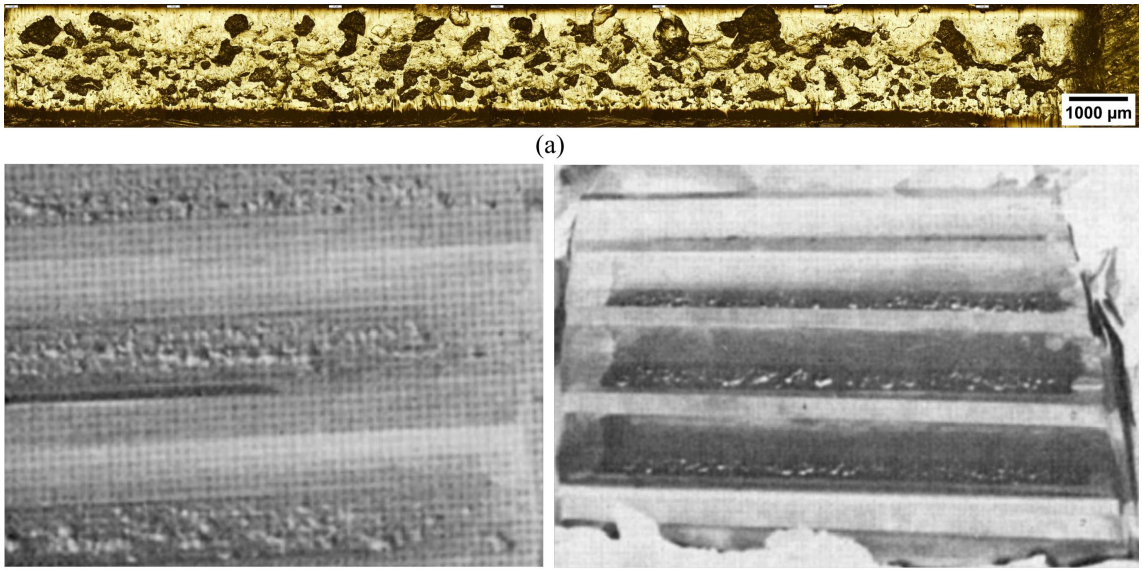
Table 6.2: Summary of key features associated with wear mechanisms according to the established literature [7,21,22,28,29,232–234].

Mechanism		Description
Abrasive wear		Parallel grooves in the sliding direction on the tooth surface under magnification
	- Mild	Fine scratches and machining marks are visible
	- Moderate	A few machining marks
	- Severe	Machining marks removed
Macropitting (fatigue wear)		Macroscopic pits on the tooth surface usually with sharp and angular edges
	- Mild	Small macropits occur in localised areas
	- Moderate	Macropits enlarge and propagate to a moderate part of the tooth surface
	- Severe	Macropits occupy a significant part of the tooth surface
Scuffing (adhesive wear)		A rough and torn surface, with plastic deformation and material transfer
	- Mild	Fine scuffed marks with an etched appearance and occurs on localised areas
	- Moderate	Occurs in patches that cover significant areas
	- Severe	Occurs on the entire addendum and/or dedendum and the surface is very rough



(a) (b)

Figure 6.5: (a) A mould image of a gear surface subjected to abrasive wear in the lubricated test, (b) the reference in [232].



(a)

(b) (c)

Figure 6.6: (a) A mould image of a gear surface worn by macropitting in the lubricated test, (b) and (c) the references in [29].

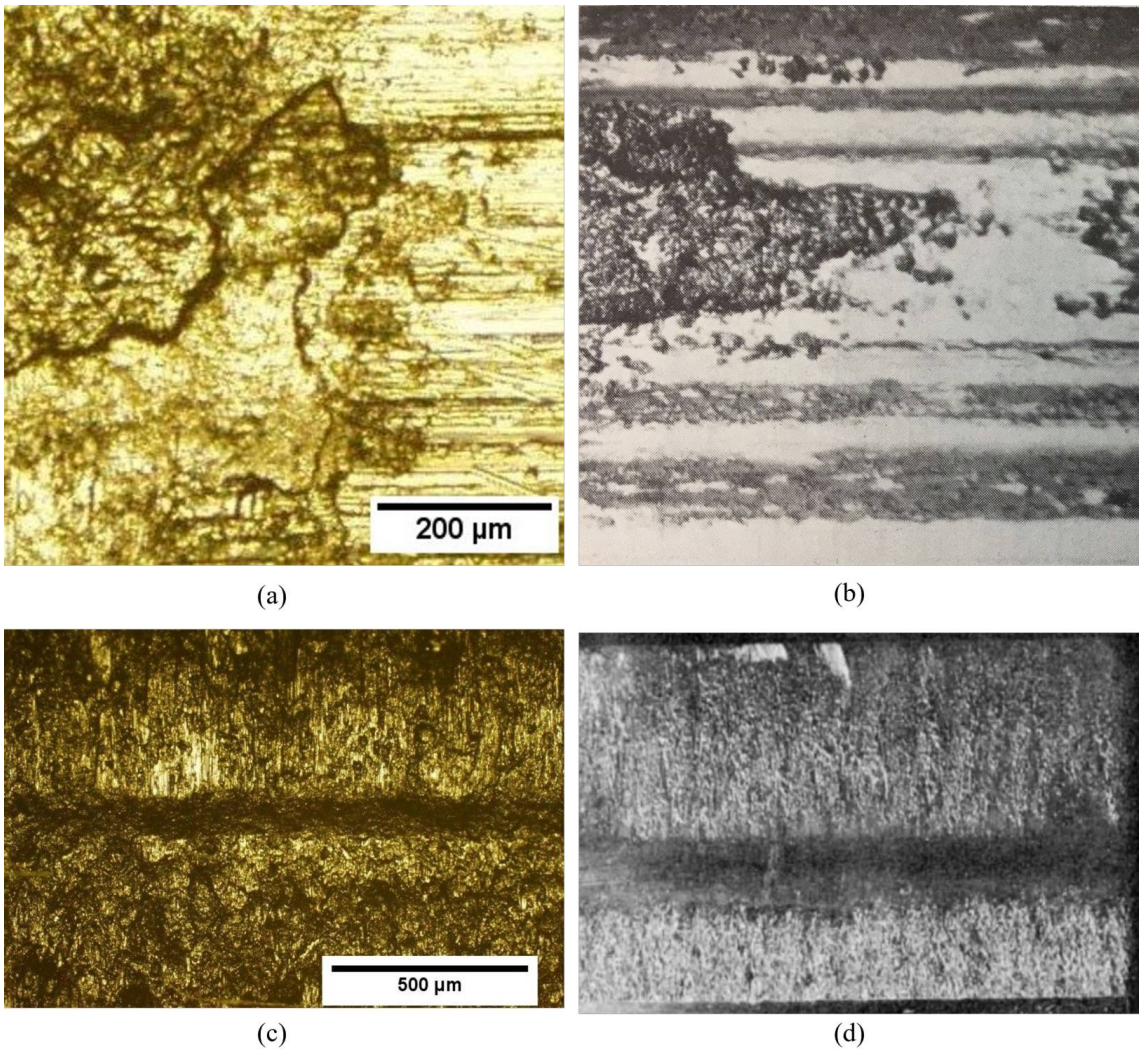


Figure 6.7: (a) A mould image of a moderate-scuffed gear surface in the dry test, (b) the reference in [232], (c) a mould image of a severe-scuffed gear surface in the dry test, (d) the reference in [29].

6.2.2.2 Data augmentation and labelling of wear particle images

For wear particle images, the labelling process was carried out before data augmentation because the contents in either images of individual particles or group particles were fixed in the data acquisition step. Established guidelines [25, 126–128] were also employed to ensure the accuracy and consistency of the labelling process. Wear particles are commonly classified into six types: rubbing, cutting, laminar, fatigue, spherical, and severe sliding

particles [25], and four of them were identified in the lubricated test using the criteria described in Table 6.3. Figure 6.8 shows labelled images of individual particles and their references. For images of group particles, they were classified into three severity levels, namely mild, moderate, and severe, based on the concentration and size of particles, which are widely used for wear severity assessment [25, 47, 80, 187] (see Table 6.4). It is worth noting that the concentration and size of wear particles can vary significantly across different machines, operating conditions, etc., and the specific values of T_m and T_s as described in the table can vary notably in different cases.

Table 6.3: Key features of different types of wear particles identified in the lubricated test [25, 126–128].

Particle type	Size range and morphological features
Rubbing	$<$ 20 µm, thin flakes with smooth surfaces
Cutting	Long and curved particles
Laminar	20 - 50 µm, thin and reflective particles with frequent occurrence of holes and uneven edges
Fatigue	$>$ 20 µm, chunky particles with rough surfaces

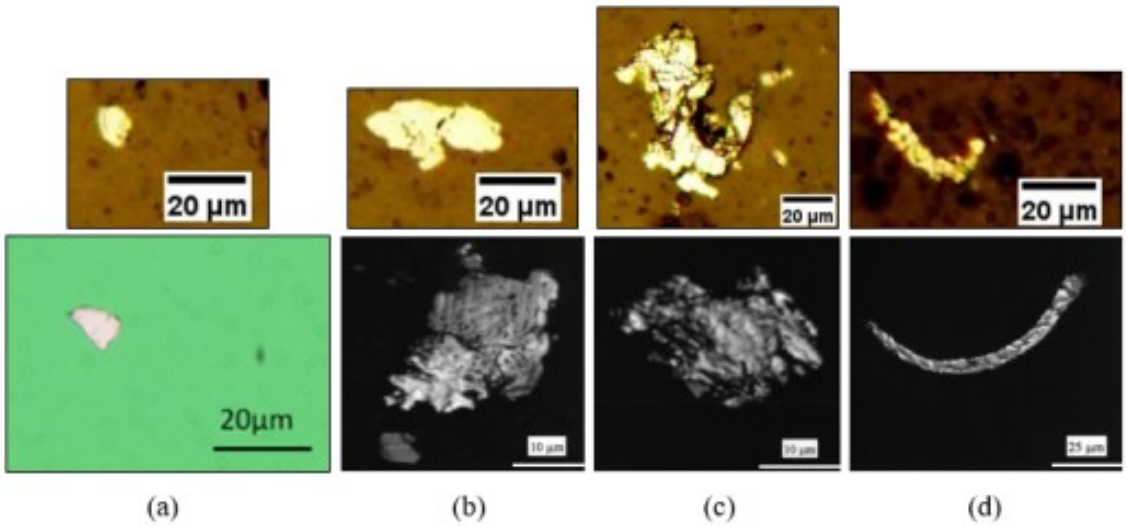


Figure 6.8: Images of different types of wear particles (upper) identified in the lubricated test and their references (lower). (a) Rubbing particles [128], (b) laminar particles [126], (c) fatigue particles [126], (d) cutting particles [126].

Table 6.4: Severity levels of images of group particles and their assessment criteria [25, 80, 102]. $0 < T_m < T_s$.

Severity level	Assessment criteria
Mild	The total quantity of wear particles is less than T_m with no particle larger than 20 µm
Moderate	The total quantity of wear particles is between T_m and T_s with no or limited particles larger than 20 µm
Severe	The total quantity of wear particles is more than T_s with multiple or more particles larger than 20 µm

After the labelling process, wear particle images were split into training and validation images (90%) and testing images (10%) before data augmentation in order to ensure that testing images were completely different from training and validation images. It was found in the labelling process that images of individual particles were unbalanced across different categories. The number of rubbing particle images was much higher than others, and the number of cutting particle images was around half of that of laminar and fatigue particles. Therefore, similar to mould images, in-plane rotation was applied to images of laminar, fatigue, and cutting particles to balance and augment the image datasets of individual particles. Table 6.5 shows the number of images of individual particles before and after data augmentation.

Table 6.5: The number of images of individual particles before and after data augmentation (listed in (training data + validation data)/test data).

Particle type	Number of images	
	Original	After augmentation
Rubbing	3,674/408	3,674/408
Laminar	359/39	3,590/390
Fatigue	363/40	3,630/400
Cutting	213/23	3,834/414

For the images of group particles, the issue of an unbalanced dataset was not evident, and data augmentation was applied to all categories of images, as shown in Table 6.6.

Table 6.6: The number of images of group particles before and after data augmentation (listed in (training data + validation data)/test data).

Severity level	Number of images	
	Original	After augmentation
Mild	626/69	2,504/376
Moderate	834/92	3,336/368
Severe	626/69	2,504/376

6.2.3 Construction, training and testing of the CNNs

This section details the architecture of the CNNs (section 6.2.3.1) and the strategy for training, validation, and testing of the CNNs (section 6.2.3.2).

6.2.3.1 Architecture of the CNNs

Convolutional neural networks were employed to classify mould images via the two-level inference (see Figure 6.1) and classify wear particle images via the method shown in Figure 6.2. A CNN contains convolutional layers that usually have multiple feature maps in parallel for different features and convolve their input data (e.g. a 2D image) with shared weights to extract higher-level features. In this study, different image datasets were used to train different CNNs for different classification tasks but the CNNs share similar architecture and hyper-parameters.

For mould images, the wear mechanism CNN (first-level inference) was designed to classify images against the four possible wear mechanism classes (including healthy), whereas the wear severity CNNs (second-level inference) were designed to classify images belonging to a specific wear mechanism within three wear-severity classes. There are two reasons for choosing such a two-level design rather than a single CNN that classifies both wear mechanism and severity. (i) The two-level design embeds the knowledge of the correlations among three severity levels of the same wear mechanism (e.g. both “severe pitting” and “mild pitting” must be “pitting”), while in a single CNN three severity levels are independent of each other. (ii) In the two-level design, a higher number of training samples for

the wear mechanism CNN can be used (i.e. the training dataset is split into 4 categories instead of 10), and it could result in more accurate identification of wear mechanism. Furthermore, the combination of the two effects is likely to improve the performance of severity CNNs because misclassifications between different wear mechanisms are less likely.

For wear particle images, the particle classification CNN was designed to classify images of individual particles against the four possible particle types (rubbing, laminar, fatigue, and cutting), whereas the wear severity CNN (using particle images) was designed to classify images of group particles against the three possible severity levels (mild, moderate, and severe).

The architecture used for the wear mechanism CNN is shown in Figure 6.9, and the other CNNs have similar architecture. There are 11 hidden layers in the CNN that comprises three stacks of convolutional layers plus max-pooling layers, followed by another two convolutional layers and three fully connected layers. The construction of the CNNs has drawn heuristics from other successful image classification CNNs [132, 235].

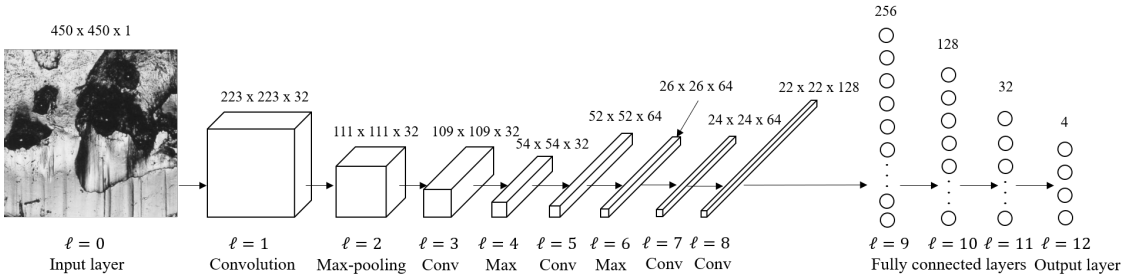


Figure 6.9: The architecture of the wear mechanism CNN. The other CNNs have similar architecture. The input layer (l=0) is a grayscale mould image (450 x 450 x 1). The output shape on top of convolutional and max-pooling layers stands for pixel x pixel x feature maps.

For the CNNs using mould images in this study, the input is a grayscale mould image (450 x 450 x 1), and the output is a four- or three-element score vector representing the probability of each wear mechanism or severity. For instance, an output of [0.1, 0.2, 0.6, 0.1] from the wear-mechanism CNN shows that the CNN has 10% confidence that the input image is normal, 20% confidence for abrasive wear, 60% for macropitting and 10%

for scuffing. For the CNNs using wear particle images, the input is an RGB image (200 x 200 x 3), and the output is a four- or three-element score vector.

The mathematical transformations from inputs to outputs of the wear mechanism CNN are briefly introduced here. The CNN was programmed to turn the input grayscale image, an intensity matrix, to a score vector (output) via 12 layers of nonlinear transformation. The lth layer of the CNN, A^l, can be denoted as

$$A^l = \{a^l_{i,j,k}\} \tag{6.1}$$

where a is an artificial neuron, $i \in \{0\,..\,N^l - 1\}, j \in \{0\,..\,M^l - 1\}, k \in \{0\,..\,D^l - 1\}$, N^l, M^l, and D^l are the height, width, and depth of the lth layer, respectively. For instance, the input layer has the dimension of $N^0 = M^0 = 450, D^0 = 1$, and the output layer has the dimension of $N^{12} = 4, M^{12} = D^{12} = 1$.

For convolutional layers ($l = 1, 3, 5, 7, 8$), convolutions between kernels and previous layers are performed, and the output will be as

$$a^l_{i,j,k} = f\left(b^l_k + \sum_{i'=0}^{S^l-1}\sum_{j'=0}^{S^l-1}\sum_{k'=0}^{D^l-1} w^l_{i',j',k'} \cdot a^{l-1}_{i\Delta^l+i',j\Delta^l+j',k'}\right) \tag{6.2}$$

where each kernel $W^l_k = w^l_{i',j',k'}$ has the width and height of S^l and the depth of D^{l-1}, and slides across its previous layer A^{l-1} horizontally and vertically, with a stride of Δ^l. In this study, $S^1 = 5, S^3 = S^5 = S^7 = S^8 = 3$, and $\Delta^1 = 2, \Delta^3 = \Delta^5 = \Delta^7 = \Delta^8 = 1$. In order to connect A^{l-1} and A^l, D^l kernels are needed to generate D^l feature maps. In this study, $D^1 = D^3 = 32, D^5 = D^7 = 64, D^8 = 128$. After the convolution operation between the kernel W^l_k and its previous layer A^{l-1}, a bias b^l_k is added. Finally, a nonlinear activation function is applied to the sum to generate the corresponding element $a^l_{i,j,k}$ in the next layer. Rectified linear units (ReLu) were used as the activation function in this

study

$$f(x) = max\,(\,0\,,\,x\,) \tag{6.3}$$

For max-pooling layers ($l = 2, 4, 6$), the output will be as

$$a^l_{i,j,k} = max(a^{l-1}_{i\Delta+i',j\Delta+j',k}) \tag{6.4}$$

where $i', j' \in \{0, 1\}$, and $\Delta = 2$ in this study. This operation effectively passes the largest value in a $2x2$ window in the input layer to the next layer, and the window slides across the input layer without overlapping.

For fully connected layers ($l = 9, 10, 11, 12$), the output will be as

$$a^l_{i,j,k} = f\left(b^l_i + \sum_{i'=0}^{N^{l-1}-1} w^l_{j,i',k} \cdot a^{l-1}_{i',j',k'}\right) \tag{6.5}$$

where $j = j' = k = k' = 0$. This operation effectively multiplies the input layer A^{l-1} by a weight matrix $W^l = \{w^l_{j,i'}\}$, adding a bias b^l_i, and applied an activation function of ReLu except for layer 12. For $l = 12$, the Softmax function is applied to normalise the output. The use of Softmax function enables the outputs to be interpreted as a probability distribution, as

$$f_{softmax}(x_i) = \frac{e^{x_i}}{\sum_{i=0}^{N^l-1} e^{x_j}} \tag{6.6}$$

where $i \in \{0\,..\,N^l - 1\}$. Also, note that the last convolutional layer ($l = 8$) is flattened to connect with the following fully connected layer in the above way.

The learning process of such a neural network is realised by iterative adjustments of internal parameters, i.e. weights and biases, to minimise the objective function, or cost

function, which is chosen to be the regularised cross-entropy as

$$C = \frac{1}{N_S} \sum_{x=1}^{N_S} \left(- \sum_{i=0}^{N^l-1} (y_i \cdot \ln a_i^l + (1-y_i) \ln(1-a_i^l)) + \frac{\lambda}{2} \sum_{l=1}^{L} \sum_{k=0}^{K^l-1} \sum_{j=0}^{J^l-1} \sum_{i=0}^{I^l-1} {w_{i,j,k}^l}^2 \right) \quad (6.7)$$

where N_S is the number of training samples, x refers to each input image, y_i is the ground truth value corresponding to the ith neuron in the output layer (either 0 or 1), a_i^l is the ith output ($l = 12$), λ is the regularisation parameter, and the last term represents the sum of square of all weights of the network. For convolutional layers ($l = 1, 3, 5, 7, 8$), $I^l = J^l = S^l$ and $K^l = D^l$. For fully connected layers ($l = 9, 10, 11, 12$), $I^l = N^l$,$J^l = N^{l-1}$ and $K^l = 1$. $L = 12$ in this study.

The learning process utilised the commonly-used backpropagation algorithm and the method of stochastic gradient descent (SGD) [63, 236]. The iterative learning process consists of three steps. (i) The network transformed input images (a batch of images) into output vectors through the feedforward path (Equation 6.1 to 6.6) and calculated costs (Equation 6.7). (ii) The gradients of costs with respect to weights in the output layer were calculated, and gradients were backpropagated from the output layer to the input layer. (iii) The SGD method was implemented to update weights and biases in the network. Then, these three steps were repeated on the remaining training samples, which were fed in the network batch by batch to finish an epoch of training. After each training epoch, the model was validated using the validation dataset and the model with the highest validation accuracy in finished epochs was saved. The training was stopped early if the validation accuracy had not improved in 20 epochs.

Note that the architecture of the networks (e.g. number of convolutional and fully connected layers) was tuned manually. In other words, the performance, i.e., validation accuracy, of an architecture was evaluated after training and then a new architecture was engineered based on the results of the previous trials. Compared to other common methods for model parameters tuning such as grid search, this method requires more manual work but considers history training and overall requires less computing time.

6.2.3.2 Strategy for training, validation and testing

The labelled mould images were split into three datasets: the training dataset with 87,571 images (73%), the validation dataset with 21,745 images (18%), and the test dataset with 10,385 images (9%). The training and validation dataset came from 149 large mould images (out of the 165 large mould images, see section 6.2.1), while the test dataset was from 16 entirely different large mould images. Using different images for testing ensures that the evaluation of the models is not biased by the use of similar training and test sets. The split ratio was selected to ensure the number of test images is sufficiently large enough to be statistically meaningful and to minimise overfitting. The split of labelled wear particle images was similar (see Table 6.5 and Table 6.6). Note that the training and validation dataset were shuffled together and then split into training and validation data in the training process. Finally, gear tooth images from gear standards and journal articles were also included in the test dataset of mould images. In the training process, the network learned from training samples, and at the end of each epoch of training, the performance of the networks was evaluated using the validation dataset. The program was written in Python using TensorFlow.

The testing of the two-level inference system included individual tests of each CNN on small mould images in the test dataset and integrated tests on full tooth mould images. Different from the large mould images, the full tooth mould images were stitched using 2 or 3 large mould images in order to show gear wear on the entire active tooth flank. Thus, full tooth mould images have a larger height ($\sim$ 9000 pixels x 1600 pixels) than large mould images ($\sim$ 9000 x 1000). The individual tests aimed to check the classification accuracy of each CNN, and the integrated tests aimed to evaluate the overall performance of the system in assessing wear mechanism and severity using full tooth mould images. Indicators of performance, including classification accuracy in training, validation and testing, were obtained in individual tests, and the corresponding confusion matrices were also built for detailed analyses on the performance of each CNN. The integrated tests generated wear mechanism maps fused with their original mould images showing the distribution of identified wear mechanisms on gear teeth, followed by wear severity maps for each wear

mechanism. The wear mechanism and severity maps were obtained through five steps.

1. Full tooth flank mould images ($\sim$ 9000 pixels x 1600 pixels) were cropped into small mould images with a square window (450 x 450) that translates vertically and horizontally with a step size of 50 pixels (or $\sim$ 102 µm). This step was necessary because the CNNs only accept inputs of the same size (450 x 450).

2. Cropped images (450 x 450) were fed into the wear mechanism CNN to assess the wear problem in each image.

3. Images identified as abnormal, i.e. abrasive wear, macropitting, or scuffing, were sent to the corresponding wear severity CNN to assess the wear severity level.

4. Results of wear mechanism assessment of small mould images were grouped to generate RGB wear mechanism maps, and the maps were fused with corresponding full tooth flank images to show the distribution of identified wear mechanism on gear teeth. Specifically, the output of the wear mechanism CNN was rounded to a one-hot encoded vector, i.e. $[0, 0, 0, 0]$ for normal, $[0, 1, 0, 0]$ for scuffing, $[0, 0, 1, 0]$ for abrasive wear, and $[0, 0, 0, 1]$ for macropitting. The last three elements were regarded as three colour channels to be used to construct a coloured wear mechanism map with the following colour code: black ($[0, 0, 0]$) for normal, red ($[1, 0, 0]$) for scuffing, green ($[0, 1, 0]$) for abrasive wear, and blue ($[0, 0, 1]$) for macropitting. The process of generating RGB wear mechanism maps is illustrated in Figure 6.10.

5. Results of wear severity assessment were grouped to generate grayscale wear severity maps to show the severity distribution of identified wear mechanisms on gear teeth. Specifically, the output of the wear severity CNNs for each 450 x 450 image was coded in grayscale as $[0.25, 0.5, 1]$, corresponding to mild, moderate and severe wear. Then, scores were grouped according to their wear mechanisms and grayscale wear severity maps for identified wear mechanisms were generated, as illustrated in Figure 6.10(b) A sample wear mechanism map is shown in Figure 6.11.

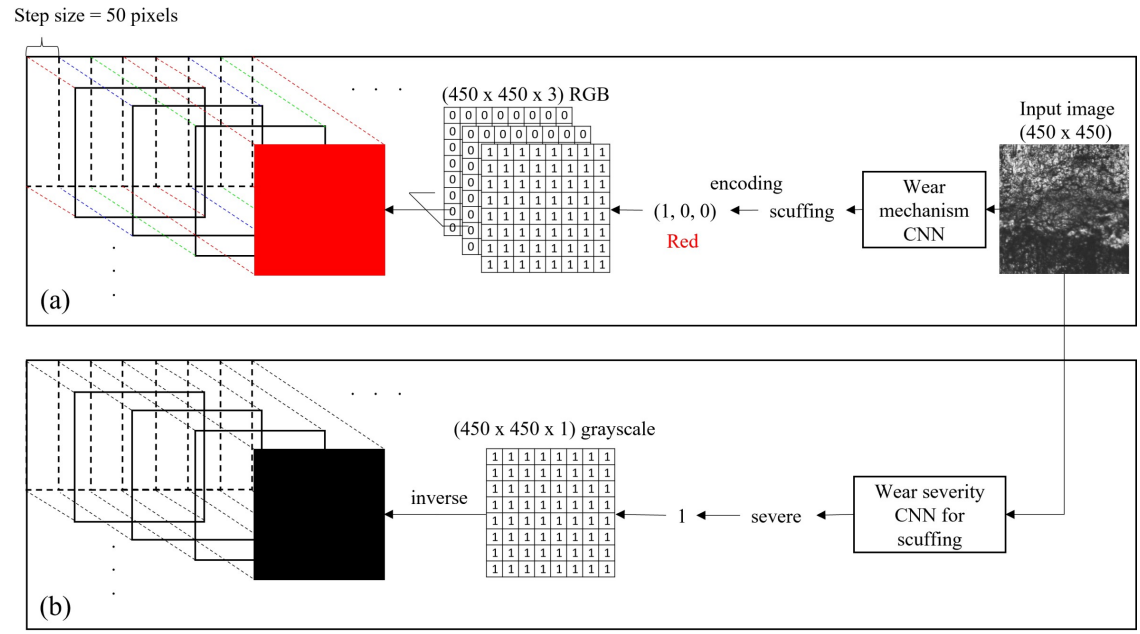


Figure 6.10: Schematic of construction of wear mechanism and severity maps out of small mould images: (a) wear mechanism map, (b) wear severity map. The two maps have the same size as the corresponding full mould image. A local average was applied to the maps after the cumulation of colour or grayscale blocks.

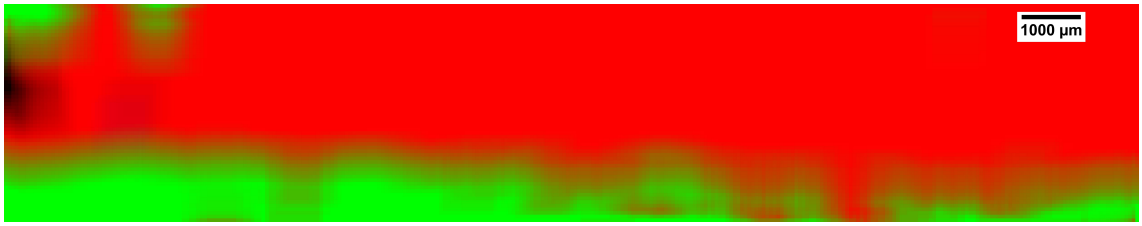


Figure 6.11: A sample wear mechanism map of a gear tooth flank. Red represents scuffing, and green represents abrasive wear, with black representing normal.

The testing of the CNNs using wear particle images measured the classification accuracy of the particle classification CNN and the severity CNN using their corresponding test samples. Confusion matrices of the testing results were also built for a detailed analysis of the CNNs' performance.

6.3 Results and discussions

This section presents training and testing results of the wear assessment CNNs using mould images (individual tests on small mould images in section 6.3.1 and integrated tests on full tooth mould images in section 6.3.2) and wear particle images (section 6.3.3). Discussions of the results are presented in section 6.3.4.

6.3.1 Training and individual tests of the wear assessment CNNs using mould images

6.3.1.1 Training and testing of the wear mechanism CNN

The model parameters of the wear mechanism CNN were manually tuned, and multiple models were trained and compared in terms of training performance. Figure 6.12 shows the training and validation accuracy of the top 5 models.

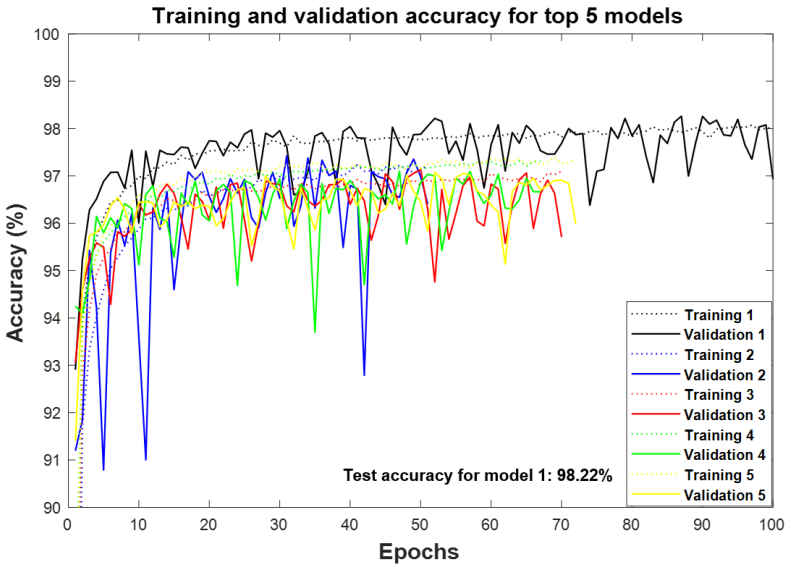


Figure 6.12: Training and validation accuracy for top 5 wear mechanism CNNs. Model 1 is the model employed as the wear mechanism CNN. Compared with model 1, model 2 has 4 convolutional layers and model 3 has 3 convolutional layers. Model 4 has the same hidden layers as model 1, except that it has 256 dense neurons in layer 11. Model 5 has only two fully connected layers.

Figure 6.12 shows that for the models, the accuracy in training and validation quickly rose to above 90% at the beginning of training and slowly increased with fluctuations. The classification accuracy was calculated based on the comparison between true labels and classification results (mechanism with the highest classification score). There is no apparent divergence between training and validation accuracy for all 5 models, which indicates that overfitting was effectively reduced. It can be seen that model 1 (in black) had the highest training and validation accuracy (around 98%). The test accuracy of model 1 was 98.22%.

Figure 6.13 shows sample test images and bar graphs of prediction vectors (relative likelihood of each wear mechanism) given by the wear mechanism CNN. The sample images show that the wear mechanism CNN can correctly distinguish with high confidence those mould images that contain single surface features such as parallel scratches of abrasive wear, irregular black pits of macropitting, and very rough surfaces of scuffing.

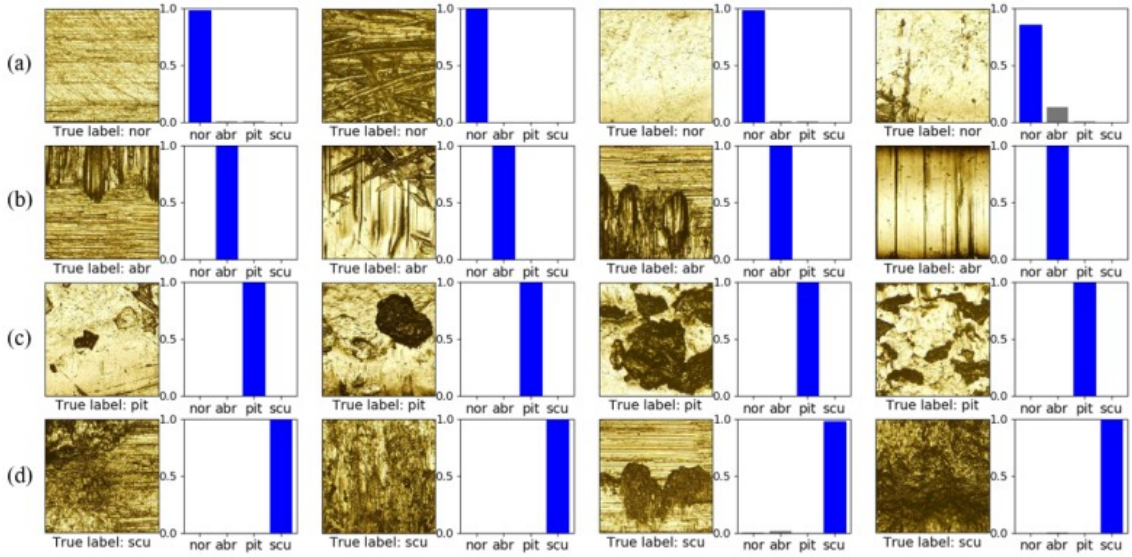


Figure 6.13: Sample test images and wear mechanism predictions. (a) Normal, (b) abrasive wear, (c) macropitting, and (d) scuffing. The identified dominant wear mechanism (classification result used for further analysis) is highlighted in blue.

Figure 6.14 shows the confusion matrix of the wear mechanism CNN on the test dataset and shows the detailed performance of the model. Rows in the matrix represent true

labels, whereas columns (in the same order as the rows) show classification results (i.e. the mechanism with the highest classification score) so that the ideal result would be a matrix with 100% on the diagonal terms (and 0% elsewhere). For example, the first row indicates that 98.09% of normal images were correctly classified as normal, whereas 0.65% were mistaken for abrasive wear, 0.19% for macropitting, and 1.07% for scuffing. In general, it is shown that the wear mechanism CNN attained a high identification accuracy across all wear mechanisms in this study. The wear mechanism CNN achieved the highest accuracy in identifying pitting (98.67%) and the lowest accuracy for abrasive wear (96.79%). Wrong predictions were also examined, and it was found that those who were wrongly identified as normal were mostly mild wear, i.e. mild abrasive wear, mild macropitting, and mild scuffing.

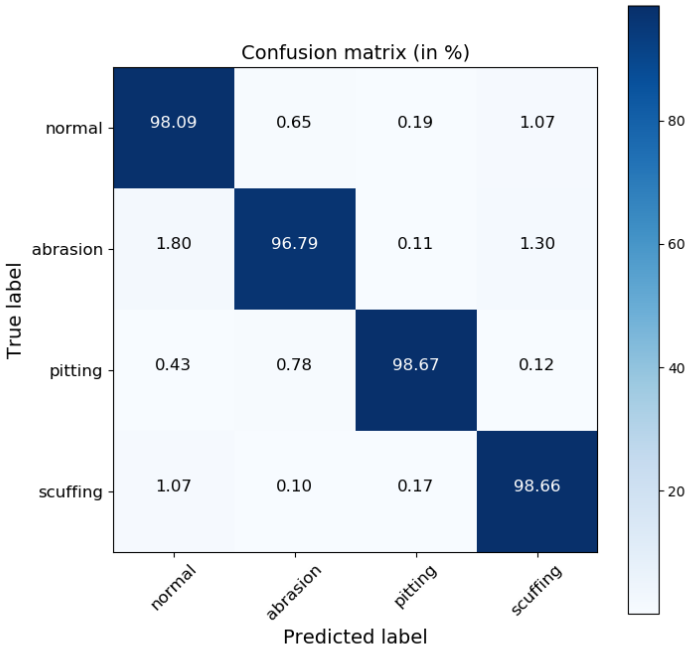


Figure 6.14: Confusion matrix of the wear mechanism CNN (percentage in the same row adds to 1).

The wear mechanism CNN was also tested on gear teeth images from other publications

and mould images captured by different cameras, and sample results are shown in Figure 6.15. The results show that the capability of the wear mechanism CNN is not limited to mould images with good illuminations. The model learnt general surface features to distinguish different wear mechanisms instead of 'memorising' specific features in mould images.

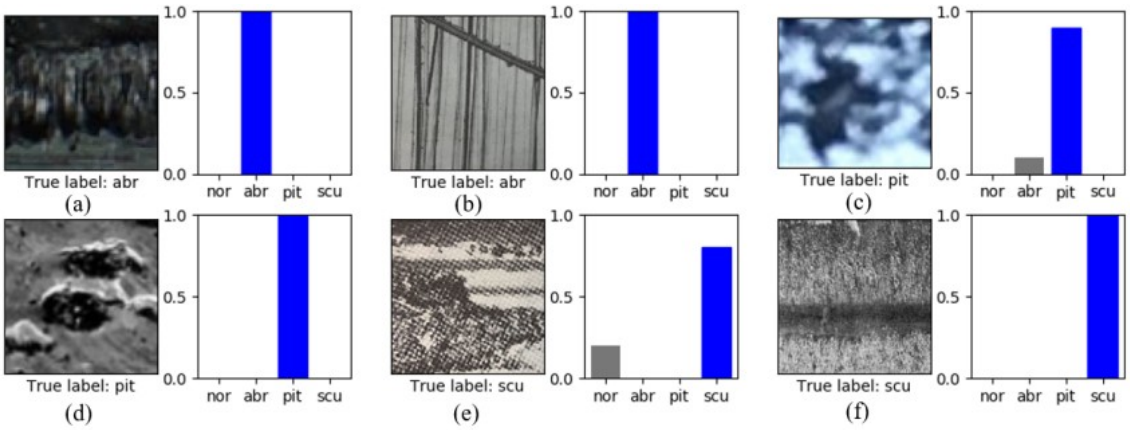


Figure 6.15: Sample external test images and predictions by the wear mechanism CNN. (a) A mould image of abrasive wear in the dry test captured by a normal camera. (b) An abrasive wear image in [232]. (c) A mould image of macropitting in the lubricated test captured by a mobile phone camera. (d) A macropitting image in [237]. (e) A scuffing image in [232]. (f) A scuffing image in [29].

6.3.1.2 Training and testing of the wear severity CNNs

Figure 6.16 shows the accuracy in training, validation and testing of three wear severity CNNs. In general, wear severity CNNs achieved a lower identification accuracy than the wear mechanism CNN, and there were variations among the three models. The scuffing CNN achieved the highest test accuracy of 97.61%, followed by macropitting and abrasive wear. The main reasons for the general lower test accuracy of wear severity CNNs could be (i) the smaller number of training samples (1/4 of the wear mechanism training samples) and (ii) less distinct boundaries between two severity levels. For instance, the difference between macropitting occupying 'a moderate' and 'a significant' part of the tooth surface is harder to be distinguished than the difference between macropits and parallel scratches. Figure 6.17 shows sample test images and bar graphs of predictions of wear severity CNNs.

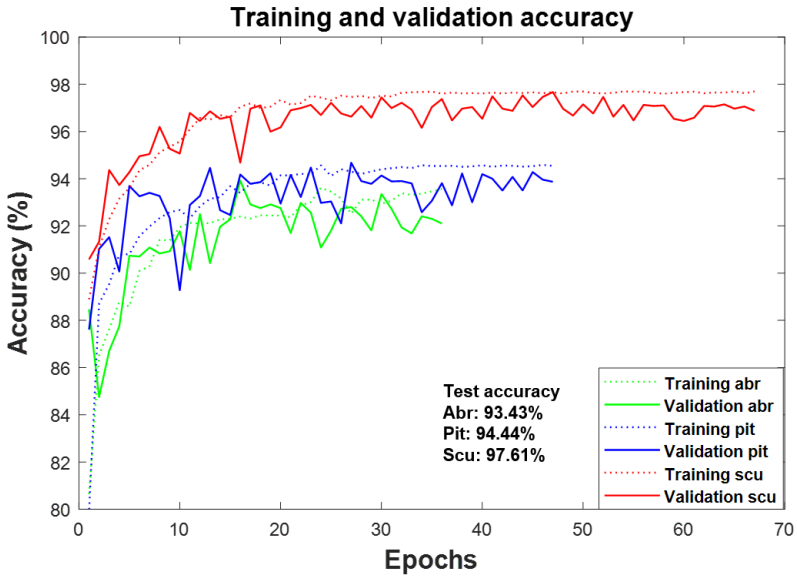


Figure 6.16: Training and validation accuracy of the three wear severity CNNs using mould images.

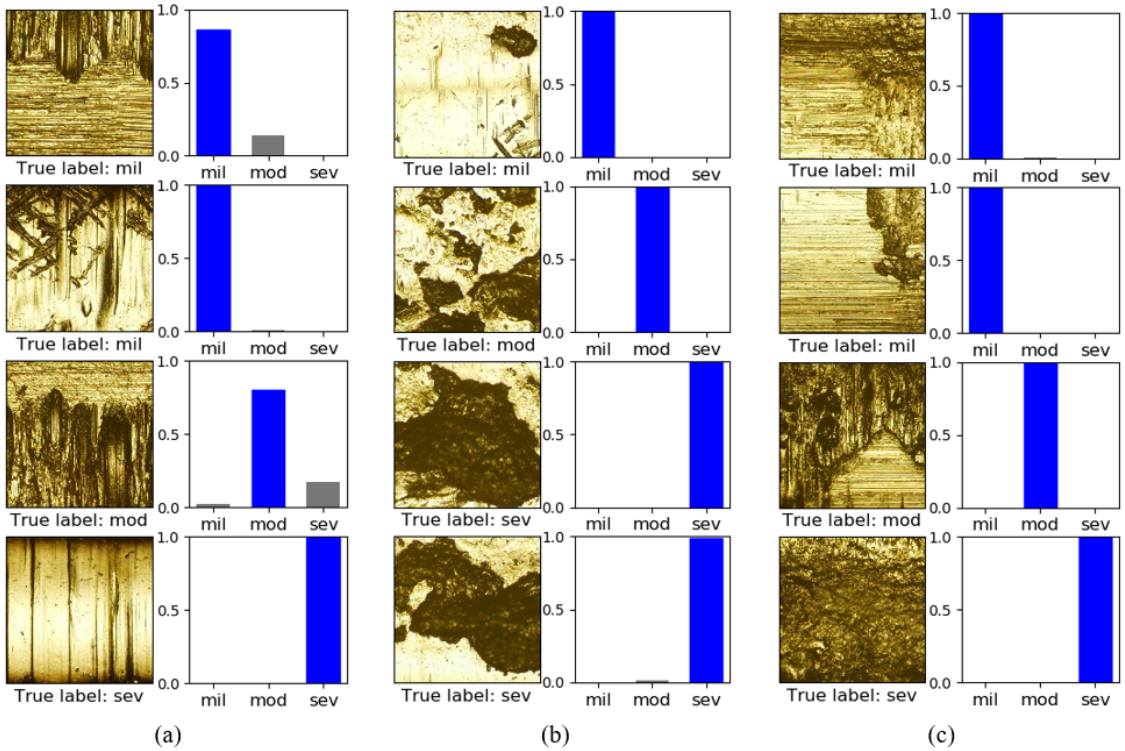


Figure 6.17: Sample test images and their wear severity assessment: (a) abrasive wear, (b) macropitting, (c) scuffing.

Figure 6.18 shows confusion matrices of three wear severity CNNs on the test dataset. The results show that all three CNNs achieved a classification accuracy of higher than 90% for all severity levels, and the scuffing had the best performance with an accuracy of at least 96%. The abrasive wear and macropitting CNNs had their highest accuracy in identifying mild wear while the scuffing CNN was best in identifying severe wear. On the other hand, they all had the lowest accuracy in identifying moderate wear.

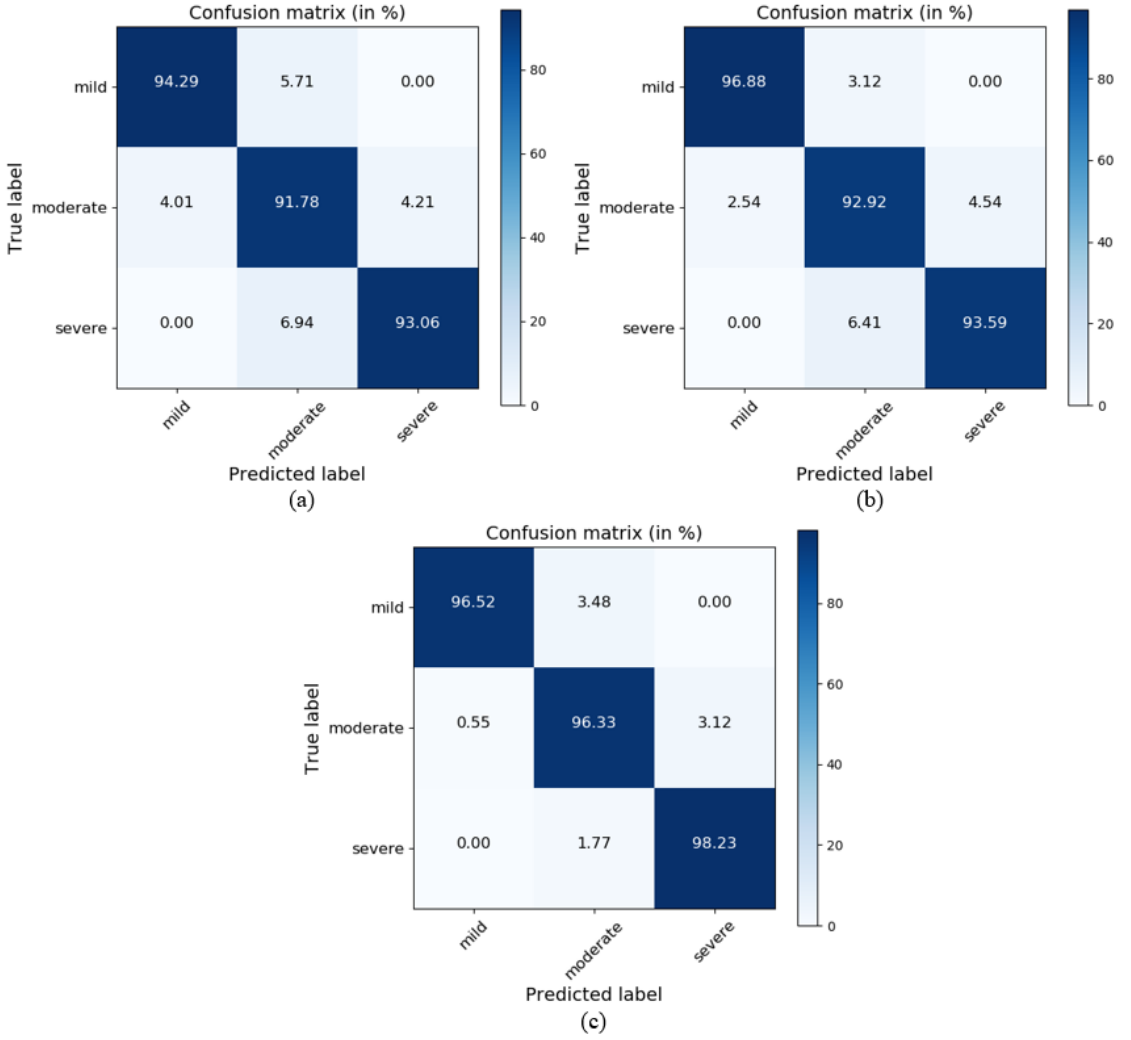


Figure 6.18: Confusion matrix of the three wear severity CNNs: (a) abrasive wear, (b) macropitting, (c) scuffing.

The different strengths and weaknesses of the three severity CNNs may be related to

two reasons, that is, (i) the unbalanced training dataset for abrasive wear and scuffing and (ii) the difference in the labelling accuracy of the training data. Regarding reason (i), mild wear makes up nearly 60% of the training samples for abrasive wear while its ideal ratio is 33% (due to three wear severity levels). In this case, even if moderate and severe wear have low accuracy, the overall accuracy, which is averaged across all training samples, could be high. Thus, it is plausible that abrasive wear CNN scored highest in mild wear and similarly, scuffing CNN scored highest in severe wear. Regarding reason (ii), as shown in Table 6.2, moderate wear is determined by the presence of a moderate amount of specific features, i.e., a few machining marks for moderate abrasive wear, enlarged macropits for moderate macropitting, and patch occurrence for moderate scuffing. In the labelling process, it is difficult to precisely distinguish images around the boundaries (i.e., mild/moderate and moderate/severe) every time. Thus, compared to mild and severe wear, which can be labelled and classified with high confidence, there was more uncertainty in labelling the training data of moderate wear. As a result, the prediction of the moderate wear of all three CNNs has the lowest probability on average. However, the three severity CNNs never misclassified severe wear as mild wear or the other way. In the future, the performance of severity CNNs could be further improved by two approaches. (i) Introduce new training samples from gear tests that experience different wear mechanisms and severity to increase the number of training samples and also to balance their distribution in the datasets. (ii) Employ automated tuning methods for models' parameters such as random search to optimise the models' configurations.

6.3.2 Integrated wear mechanism and severity assessment using full tooth images

Mould images that recorded the evolution of an entire tooth flank of the driving gear in the lubricated test are shown in Figure 6.19. The main observations drawn from the figure are as follows. (i) Initial roughening marks were effectively removed in a short period (0.17 million cycles), especially on the dedendum. (ii) The addendum and dedendum exhibited drastically different wear patterns in which the dedendum was severely damaged

by macropitting while the addendum was damaged at a slower rate by abrasive wear. (iii) Macropitting initiated near the root and propagated towards the pitch line.

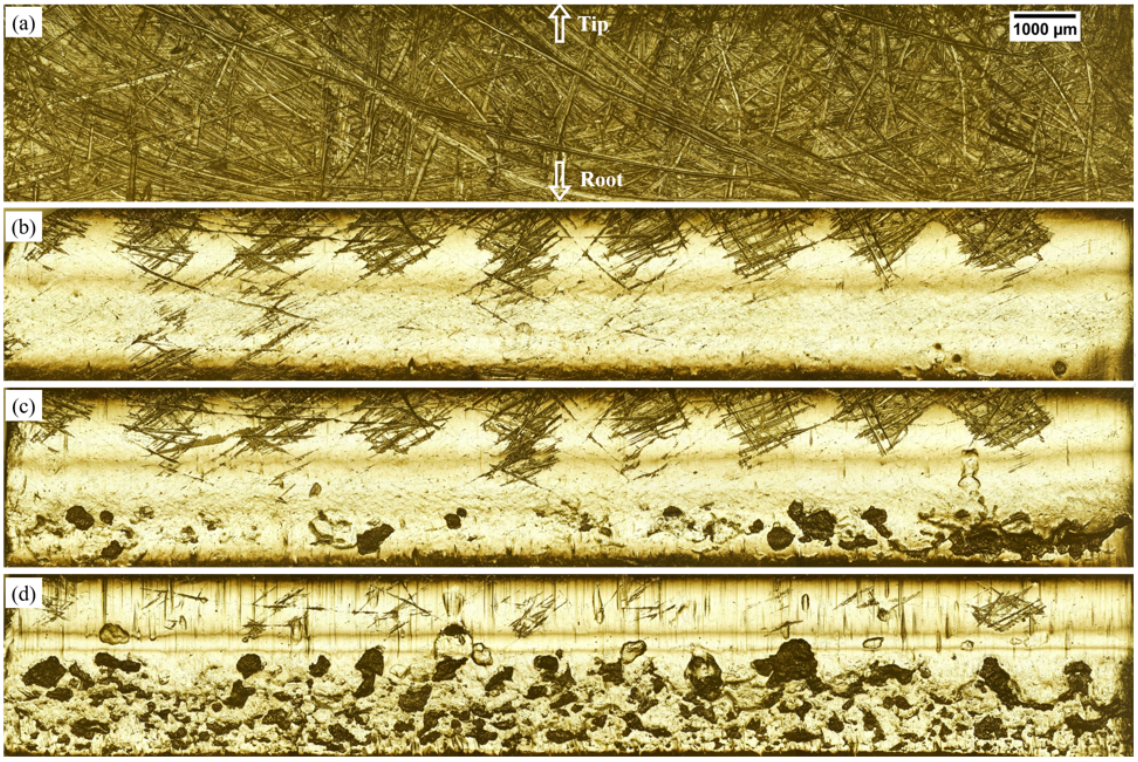


Figure 6.19: Surface evolution of a driving gear tooth flank in the lubricated test at (a) 0 cycle, (b) 0.17 million cycles, (c) 0.51 million cycles, and (d) 3.25 million cycles.

Figure 6.20 shows the evolution of a driven gear tooth flank in the dry test. The main observations drawn from the figure are as follows. (i) The tooth flank was severely damaged first at both edges, and then the damage quickly spread to the entire flank. (ii) Severe abrasive wear occurred at both edges of the tooth flank at 6.6 k cycles, especially on the dedendum, became the dominant wear mechanism on the dedendum at 12.6 k cycles but was replaced by scuffing later. (iii) Scuffing patches on the dedendum at 6.6k cycles were replaced by abrasive wear at 12.6 cycles. It is worth noting that a similar phenomenon, i.e. initial scuffing replaced by abrasive wear, was reported in [238].

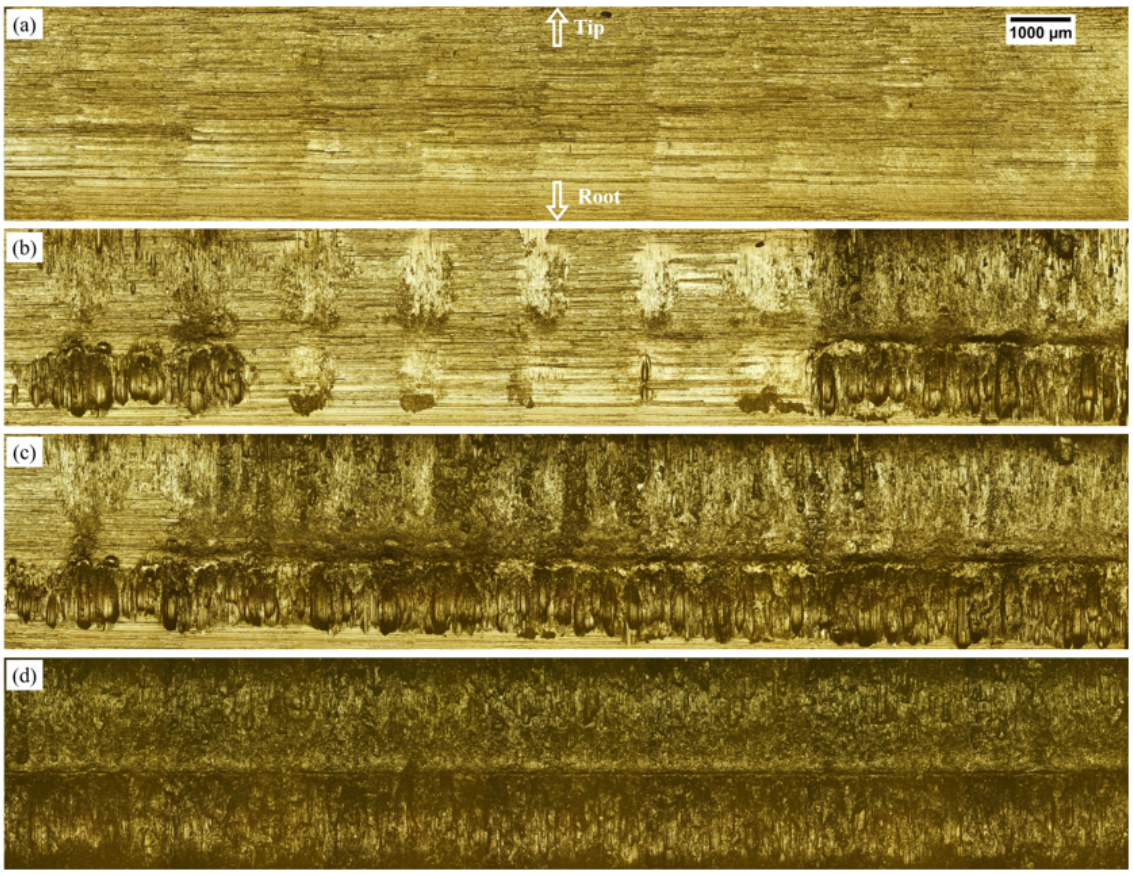


Figure 6.20: Surface evolution of one driven gear tooth in the dry test at (a) 0 cycle, (b) 6.6 k cycles, (c) 12.6 k cycles, and (d) 30.6 k cycles.

The two-level inference system was applied to the eight full tooth flank mould images of Figure 6.19 and Figure 6.20 and resulted in the wear mechanism and severity maps and their labelled maps of Figure 6.21 and Figure 6.22. In general, wear maps assessed by the CNNs are very close to those labelled manually. The first column of Figure 6.21 shows wear mechanism maps superimposed to original images taken on the pinion at different stages of the lubricated test. These maps clearly reflect the initiation and propagation of the two major wear mechanisms in the lubricated test, i.e. macropitting (blue) and abrasive wear (green). Furthermore, wear severity maps in the second and third columns of Figure 6.21 show the severity of the two wear mechanisms, respectively, and their spatial variations. It can be seen from Figure 6.21 (d2-d3) that the severity of abrasive wear was much lower than that of macropitting, although they both occupied a significant part of

the tooth flank. In addition, wear severity of either wear mechanism was not uniformly distributed across the tooth flank, especially in the early stages of wear.

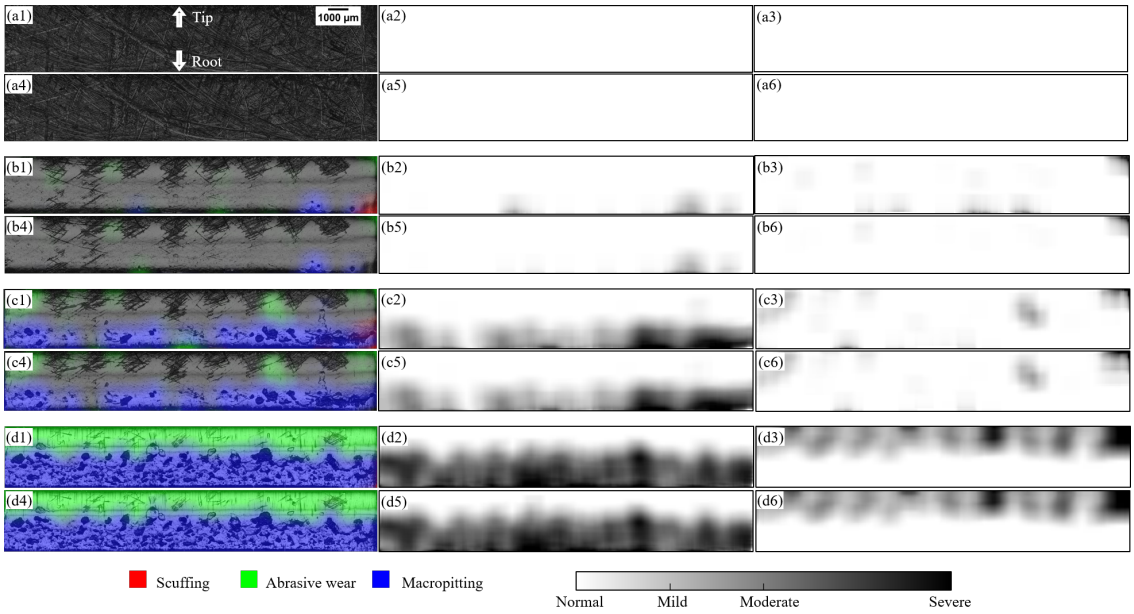


Figure 6.21: Wear mechanism and severity maps in the lubricated test and their labelled maps. (a1), (b1), (c1) and (d1) are wear mechanism maps at running cycles of 0, 0.17 million, 0.51 million, and 3.25 million, respectively. (a2) - (d2) are corresponding wear severity maps of macropitting and (a3) - (d3) for abrasive wear. (a4) - (d4) are labelled wear mechanism maps for (a1) – (d1). (a5) – (d5) and (a6) – (d6) are labelled wear severity maps for (a2) – (d2) and (a3) – (d3), respectively.

Similarly, Figure 6.22 shows the evolution of the distribution of scuffing and abrasive wear and their severity in the dry test. It can be seen that in the dry test, the tooth surface was worn more quickly than in the lubricated test, with a significant area severely damaged in just 12.6 k cycles. In addition, the severity maps of scuffing show that the spread of this wear mechanism started from the right edge instead of both edges, indicating the potential existence of minor shaft misalignment in the dry test.

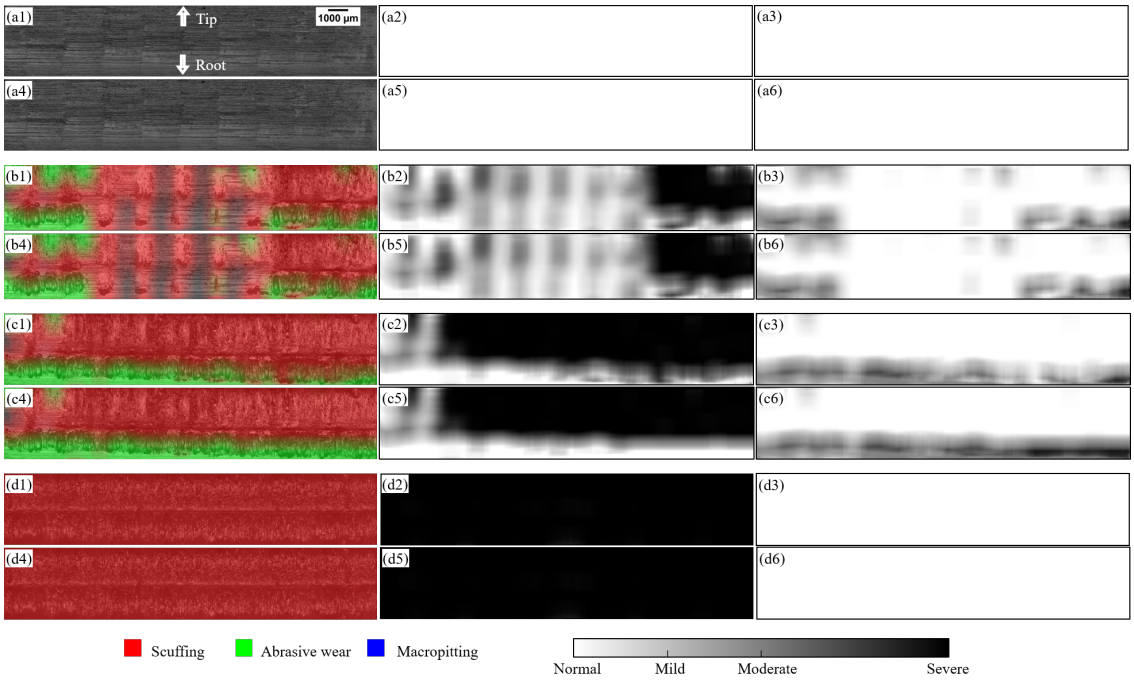


Figure 6.22: Wear mechanism and severity maps in the dry test and their labelled maps. (a1), (b1), (c1) and (d1) are wear mechanism maps at running cycles of 0, 6.6 k, 12.6 k, and 30.6 k, respectively. (a2) - (d2) are corresponding wear severity maps of macropitting and (a3) - (d3) for abrasive wear. (a4) - (d4) are labelled wear mechanism maps for (a1) – (d1). (a5) – (d5) and (a6) – (d6) are labelled wear severity maps for (a2) – (d2) and (a3) – (d3), respectively.

6.3.3 Training and testing of the wear assessment CNNs using particle images

6.3.3.1 Training and testing of the particle classification CNN

Figure 6.23 shows the training, validation and testing accuracy of the particle classification CNN. It can be observed that the training accuracy stabilised around 93% while the validation accuracy stabilised around 91%, with the highest validation accuracy of 91.65%. The stable training accuracy being slightly higher than the stable validation accuracy suggests the existence of minor overfitting. The particle classification CNN scored 88.97% in testing, which was 2.68% lower than the best performance in validation. Considering

the low number of training samples before data augmentation (see Table 6.5), the particle classification CNN achieved reasonable performance in testing. Figure 6.24 shows sample test images of individual particles and their classification results.

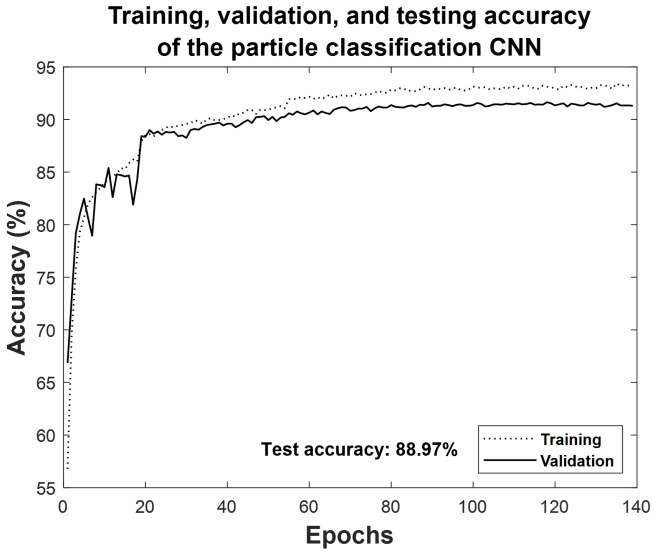


Figure 6.23: Training, validation, and testing accuracy of the particle classification CNN.

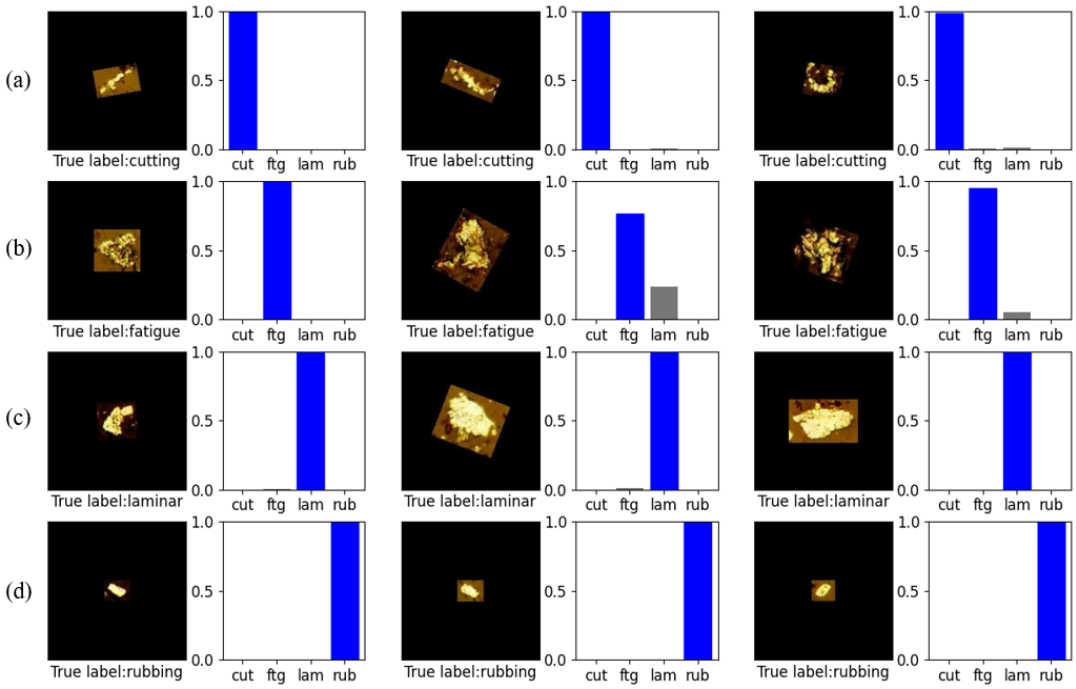


Figure 6.24: Sample images of individual particles and their classification results. (a) cutting, (b) fatigue, (c) laminar, and (d) rubbing particles.

Figure 6.25 shows the confusion matrix of the particle classification CNN, which achieved a significantly higher accuracy in identifying rubbing particles (97.79%) than others and showed the lowest accuracy in identifying fatigue particles (82.25%).

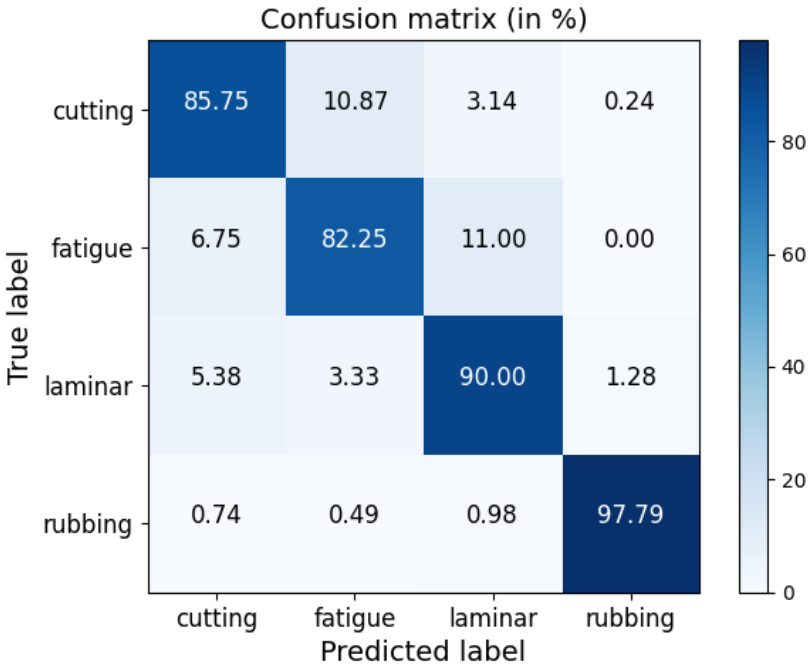


Figure 6.25: Confusion matrix of the particle classification CNN.

The fact that rubbing particles are usually smaller than the other three types of particles (see Table 6.3) could account for the high accuracy to a large extent. Moreover, the abundance of original training samples for rubbing particles, which did not require data augmentation, could also contribute to the result. For fatigue particles, the confusion matrix shows that the CNN had 11% of chance to classify them as laminar particles. It suggests that surface morphology could be harder for the CNN to capture than the size or shape information since fatigue particles usually have rough surfaces while laminar particles usually have smooth surfaces. In addition, the confusion matrix shows that the particle classification CNN rarely misclassify normal particles (rubbing particles) as abnormal particles (the others) or vice versa.

6.3.3.2 Training and testing of the wear severity CNN

Figure 6.26 shows the training, validation, and testing accuracy of the wear severity CNN using images of group particles. The highest training accuracy was 95.46%, while the highest validation accuracy was 92.69%. The wear severity CNN achieved a testing accuracy of 90.11%. Similar to the particle classification CNN, minor overfitting also occurred to the wear severity CNN. Overall, the wear severity CNN had higher classification accuracies than the particle classification CNN, which could be related to the higher number of original training samples for the severity CNN (see Table 6.6). Figure 6.27 shows sample test images of group particles and their classification results.

Figure 6.28 shows the confusion matrix of the wear severity CNN, which achieved the highest classification accuracy for mild wear (95.29%) and the lowest accuracy for moderate wear (85.82%). The misclassifications of moderate wear images as other types could be related to the difficulty in manually labelling those images that are around the boundaries between different categories (i.e., mild/moderate and moderate/severe).

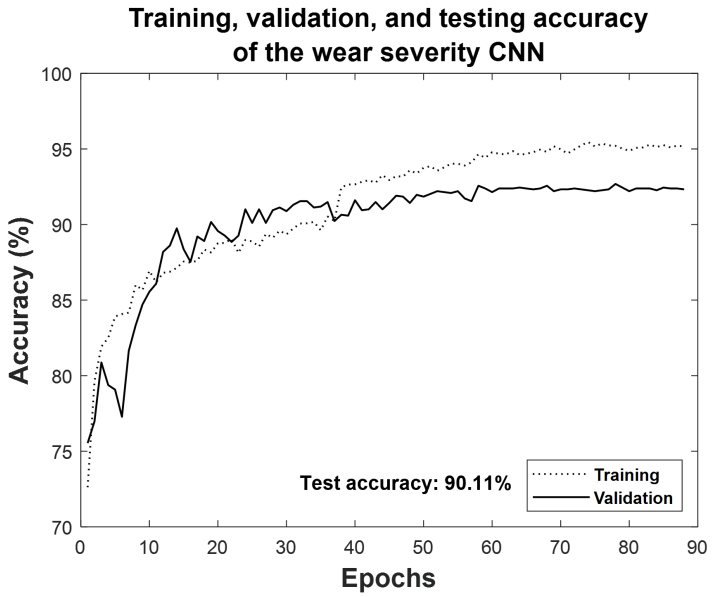


Figure 6.26: Training, validation, and testing accuracy of the wear assessment CNN using particle images.

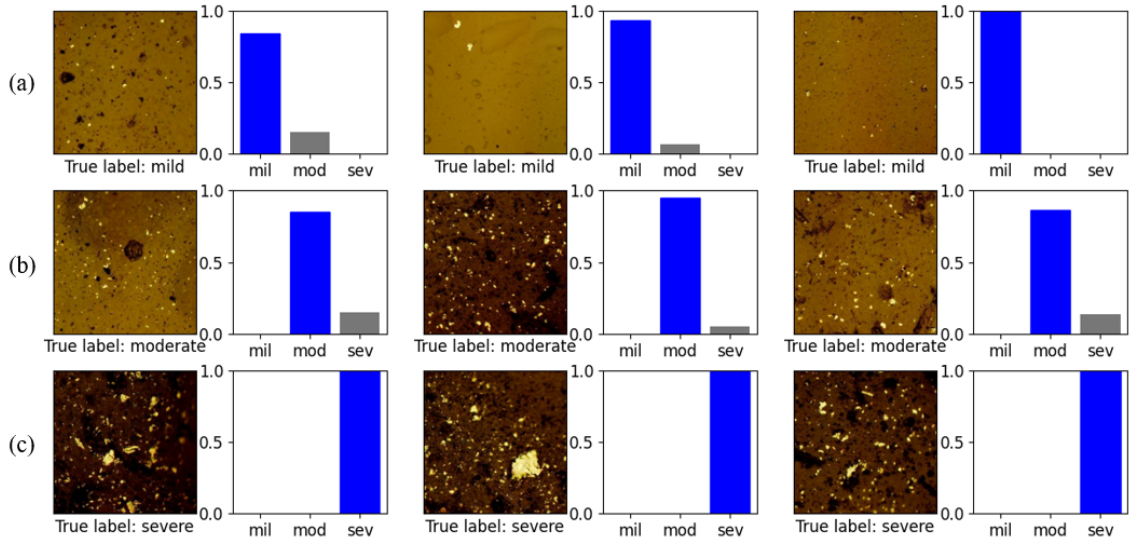


Figure 6.27: Sample images of group particles and their wear severity assessment results. (a) Mild, (b) moderate, and (c) severe wear.

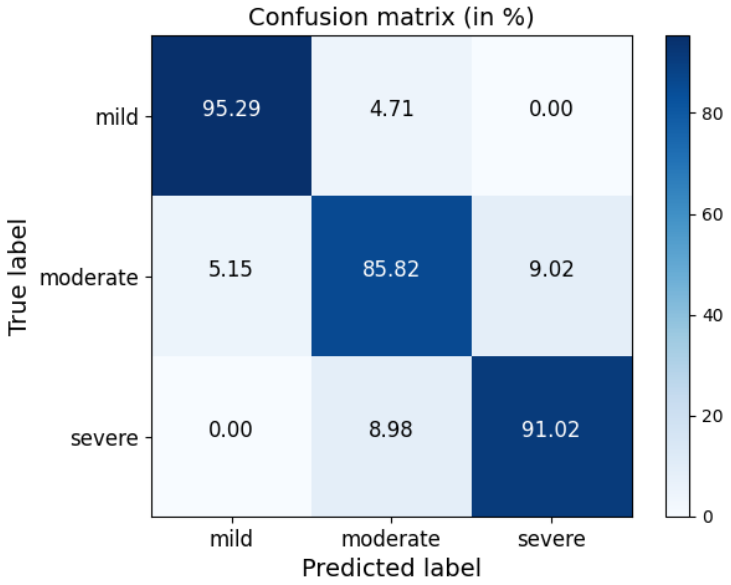


Figure 6.28: Confusion matrix of the wear severity CNN using wear particle images.

6.3.4 Discussions

The CNNs showed overall reasonable capabilities in wear mechanism and severity assessment in various tests above, and they also demonstrated different performance for different

image datasets. For mould images, the wear mechanism CNN achieved an overall higher classification accuracy (> 98%) than the three wear severity CNNs (93% - 98%), while for wear particle images, the wear severity CNN achieved a slightly better performance (90%) than the particle classification CNN (89%). This difference could be related to multiple factors, and one of them could be the number of original training samples. The original training samples for the wear mechanism CNN (> 100 thousand, see Table 6.1) were several times more than those for each of the three wear severity CNNs (15 - 40 thousand), while the numbers for the particle classification CNN (200 - 400, except for rubbing particles, see Table 6.5) were notably smaller than those for the wear severity CNN (600 - 900, see Table 6.6). This effect could also partly explain the better performance of the CNNs using mould images than the CNNs using wear particle images. Except for the number of training samples, the sharpness of surface features could also affect the performance of the CNNs. In this study, key surface features of mould images (see Figure 6.13 and Figure 6.17), e.g., pits, scratches, and cracks, are generally sharper than those of wear particle images (see Figure 6.24 and Figure 6.27), e.g., smoothness of the particle surface, which could contribute to the difference between the two groups of CNNs.

Compared to the few existing studies on wear assessment using worn surfaces images [114, 222, 223], the CNNs using mould images achieved higher classification accuracies (e.g., $\sim$ 80% in [114] in classifying mild and severe abrasive wear). More importantly, the deep learning approach used in this study did not require the complicated process of feature extraction, and it could be easier for the CNNs to extend to other wear mechanisms and severity levels when new images of worn surfaces are available. In addition, the wear mechanism and severity maps constructed using the classification results of the CNNs are much easier to use for maintenance engineers. For wear assessment using wear particle images, there are more existing works [109, 112, 126, 128, 129], and a few of them [128, 129] also employed deep learning techniques for image classification. The CNNs using wear particle images in this study showed better performance ($\sim$ 90% classification accuracy) than the integrated model (83%) in [128] and comparable performance to the CNNs (93.89%) in [129, 239].

The test results demonstrated the advantages of combining wear images and deep learning approaches, including no feature engineering required, high classification accuracy, etc. However, there could also be limitations for the deep learning approaches in tasks such as wear severity assessment using wear particle images. Specifically, the concentration, size and size distribution of wear particles, which are the main parameters used for wear severity assessment, can be extracted from wear particle images using image processing techniques. In addition, the wear severity CNN could have difficulty generalising to other wear particle images from different machines or operating conditions, which could result in significantly different particle concentrations and sizes. In such cases, an algorithm that can output quantitative features (e.g., actual values of particle concentration) and allows analysts to adjust severity thresholds to adapt to different cases could be more favourable.

6.4 Summary

A two-level inference system made up of one wear mechanism CNN and three wear severity CNNs was proposed in this study for a novel method of automated assessment of gear wear mechanism (first-level inference) and severity (second-level inference) using gear mould images. The deep learning approach was also applied to wear particle images, including a particle classification CNN and a wear severity CNN automated wear assessment.

For training, validating and testing the CNNs, a high number of mould images were obtained from the lubricated and the dry gear wear test via moulding and imaging, and a lower number smaller of wear particle images were obtained only from the lubricated test via oil sampling, filtergram, and imaging. The wear particle images were converted to images of individual particles (used by the particle classification CNN) via image segmentation and images of group particles (used by the wear severity CNN) via image cropping. Then, the labelling of the data strictly followed gear standards and other established guidelines to minimise subjectivity and data augmentation was applied to the datasets to generate nearly 120,000 samples for the wear mechanism CNN and three severity CNNs, nearly 15,000 samples for the particle classification CNN, and around 8,000 samples for

the wear severity CNN, to reduce overfitting. The CNNs shared similar architecture that contained 11 hidden layers and were trained using the backpropagation algorithm with the SGD method. The wear mechanism CNN and three severity CNNs took grayscale mould images (450 pixels x 450 pixels x 1 channel) as inputs and output four- or three-element score vectors indicating the wear mechanism or severity, while the particle classification CNN and the severity CNN took RGB particle images (200 x 200 x 3) as inputs and outputs four- or three-element score vectors indicating the particle type or severity level.

In general, this study presents a new methodology to perform automated wear mechanism and severity assessment combining deep learning with mould images of gear surfaces, making it possible to monitor and effectively communicate the evolution of a gear degradation process, with improved details and massively reduced manual operations if compared to previous approaches. The significance of the proposed method is mainly in its automated and repeatable nature, requiring little or no expert intervention after its initial synthesis. This method could provide important assistance to maintenance engineers without expertise in gear surface analysis or particle analysis to find the root cause of gear wear and carry out informed maintenance tasks. The findings of this chapter are summarised as follow.

- The wear mechanism CNN achieved a high classification accuracy of more than 98% among normal, abrasive wear, macropitting and scuffing, while three wear severity CNNs achieved classification accuracies of 93% (for abrasive wear), 94% (macropitting), and 98% (for scuffing) among mild, moderate, and severe wear images.
- The two-level inference system can be applied to full tooth flank mould images to generate an overall assessment and a clear display of the distribution of different wear mechanisms and severity levels of identified wear mechanisms across the entire tooth flank.
- The particle classification CNN achieved a reasonable (considering the low number of training samples) classification accuracy of around 89% among rubbing, fatigue, laminar, and cutting particles, while the wear severity CNN (using particle images)

achieved a similar accuracy of around 90% among mild, moderate, and severe wear images.

The capabilities of the two-level inference system can be extended to cover other gear wear mechanisms by incorporating image data from other gear tests. Similarly, more wear particle images can enhance the performance of the particle classification CNN. An open database will be built in the future in which gear experts can utilise existing data and contribute new data for a continuously improving dataset for gear wear assessment.

Chapter 7

Image-based Gear Wear Prediction: A Deep Learning Approach

This chapter addresses the fourth and the last objective of this research, i.e., **To develop a gear wear prediction model using direct wear information**, by developing a deep learning model to predict the future gear surface appearance using history gear mould images. This work is built upon the the moulding-imaging method developed in Chapter 4, and it investigates the potential of utilising mould images for gear prognostics using a deep learning approach.

This chapter is a modified version of the below journal manuscript that will be submitted to the journal of Wear in Jan/Feb 2022, and the main modifications include a condensed methodology section of which the details about the experimental test were moved to Chapter 3.

- **H. Chang** and Z. Peng, "Image-based gear wear prediction: A deep learning approach," *Wear*, 2022.

Abstract

Gear wear prediction is vital to optimise maintenance policy and minimise machine downtime. Due to increasing data availability and fast development of deep learning, data-driven methods have emerged as promising solutions for gear wear prediction. However, most data-driven models adopt gear health data that contain limited wear information and can be easily impacted by various factors. To achieve efficient and accurate gear wear prediction, this chapter proposes a novel framework in which a convolutional long short-term memory (ConvLSTM) model learned gear wear dynamics from direct gear health data, i.e., gear tooth images. Given a sequence of history images and time information, the model is able to generate a future image of the same gear tooth. Qualitative evaluations showed that the prediction is realistic and preserves important tribological features for wear assessment. Quantitative evaluations showed that the model achieved high scores for both realism-related metric (e.g., 0.994±0.002 for VGG cosine similarity) and accuracy-related metrics (e.g., -2.56%±4.21% for pitted region error).

7.1 Introduction

Gear wear is an inevitable surface degradation process occurring on gear teeth, and it is often the initial stage of gear failure, which is a main cause of machine failure [20, 21]. To minimise the downtime, maintenance costs, and safety risks associated with machine failures, condition-based maintenance (CBM) is widely used for important machines [240]. A critical part of CBM is prognosis, and it is to monitor the degradation of a component (gear in this study) from its initial stage and predict its remaining useful life (RUL) [14]. Gear prognosis is a proactive process that enables maintenance scheduling to be optimised and prevents unplanned downtime [14]. Gear wear prediction is a core task in gear prognosis and has received considerable research interests [15, 152, 154, 241].

However, it is still difficult to accurately predict the development of gear wear in an efficient manner, mainly due to the large number of impacting factors such as load, speed,

gear materials, lubrication, etc [146]. As reviewed in section 2.4, existing studies on gear wear prediction usually adopt two methods, namely physics-based methods and data-driven methods [136]. Physics-based models, using a bottom-up approach, consider main impacting factors and theoretically simulate the gear wear process. In contrast, data-driven models attempt to learn the characteristics of gear wear solely from condition monitoring (CM) data such as vibration signals [40, 135]. However, both methods have certain limitations that restrict their applications. For physics-based models, current modelling efforts can only include limited impacting factors, and complex models are too computationally expensive [15, 242]. For data-driven methods, despite their traditional problem of feature engineering being solved by deep learning, the quality of CM data can hinder the development of data-driven methods [15, 135]. Firstly, most CM data do not directly reveal changes in the appearance or dimension of gear teeth, and therefore, they contain limited gear wear information [7]. Then, most CM data can be easily impacted by factors such as inherent imperfections of the gear, noises from the signal transmission path, and varying operating conditions [5, 15, 17].

Compared with other CM data, images of gear tooth flank contain direct wear information and are often captured during inspections for basic wear assessment. However, they are rarely used for gear wear prediction in the research field. The reasons for their under-utilisation for wear prediction could be twofold: (i) high-resolution gear tooth images are difficult to obtain without disassembling the gearbox; (ii) the trend of gear wear evolution in the form of images is difficult to capture, especially with multiple wear mechanisms. The first issue, i.e., acquisition of gear tooth images without disassembling the gearbox, has been addressed by the combination of surface replication and image analysis [121]. The authors showed that the wear evolution on gear teeth could be recorded on gear moulds with micro-scale precision. Through analysing the mould images captured using a microscope, valuable and direct wear information is obtained. Then, the problem of learning gear wear dynamics could be tackled by deep learning, which has shown great capabilities to learn various high-level representations from raw data such as raw images and videos [63, 243]. Similar to the image sequence in a video, images of a gear tooth flank at different times have spatiotemporal correlations that characterise the gear wear

process. Thus, the deep learning techniques for video prediction can also be applied to image sequences of gear teeth for wear prediction. Essential background information on video prediction is provided below.

Video prediction using a deep learning approach is a relatively new research topic. Given a sequence of video frames, a video prediction model generates subsequent frame(s). A video, which is a time series of images, has spatial dimensions (width, height, and channels of each frame) and a temporal dimension (i.e., time). Both the spatial and temporal information should be modelled to extract meaningful patterns of a video. Current research works [171, 244, 245] usually integrate convolutional models that are famous for handling images, and recurrent models that are widely used for time-series data, to model the spatiotemporal dimensions of a video. Among various recurrent neural networks (RNNs), long short-term memory (LSTM) networks proved to be exceptionally effective in learning and storing information for a long term, which are widely used in tasks such as speech recognition [63]. To allow LSTM to model the spatial dimensions, Shi et al. [171] replaced key matrix multiplications with convolutions in a LSTM layer, and trained the proposed ConvLSTM model for precipitation nowcasting. The ConvLSTM model proved to have substantially better performance than normal LSTM models in video prediction [171]. A common problem many video prediction models encounter is blurry predictions, which is mainly due to the deterministic loss that tends to average equally probable predictions. Many attempts were made to solve this problem. For example, Mathieu et al. [172] used an adversarial training strategy for the first time in video prediction and obtained much sharper predictions than previous deterministic models. Lee et al. [173] introduced stochasticity to GAN models and proposed VAE-GANs, a combination of variational autoencoders (VAEs) and generative adversarial networks (GANs), to generate diverse (thanks to VAEs) and realistic (thanks to GANs) predictions.

This study is conducted to demonstrate that efficient and accurate gear wear prediction is achievable using deep learning techniques and based on limited history gear tooth images. The novelty of the study is that automated wear prediction is performed using the state-of-art AI method and raw image sequences of gear tooth flank that contains direct and

detailed information of gear wear evolution.

7.2 Methodology

As introduced in section 3.5, gear wear prediction was achieved through training a deep learning model on image sequence datasets, in which each sequence is a time series of gear mould images for the same tooth. The model automatically learns gear wear dynamics through mould image sequences and generates images that predict future gear wear. This section details the acquisition of the image sequence datasets of gear wear evolution in section 7.2.1, the deep learning model for wear prediction in section 7.2.2, and the evaluation of predicted images in section 7.2.3.

7.2.1 Image sequence datasets of gear wear evolution

The development of a deep learning model for gear wear prediction needs datasets of image sequences of gear tooth for training, validation, and testing. The datasets used in this study came from the lubricated test (see section 3.1), during which the moulding method (see section 3.2) was applied at different times to obtain 41 batches of gear moulds (6 moulds per batch). The moulds were imaged using optical microscopy with a 5x objective lens. Due to the large dimension of the active gear tooth flank (20 mm x $\sim$ 3 mm), multiple images (21 images: 3 rows along the tooth profile and 7 columns along the face width) were captured to cover the entire active flank. Then, the images were stitched to form a full-flank image (see Figure 7.1 as an example).

Figure 7.1: A stitched mould image covering an entire active tooth flank of the driving gear in the gear test (the black pits are macropits and the non-parallel scratches are manual roughening marks).

Then, the stitched full-flank images were converted into ready-to-use datasets of image sequences for the wear prediction model through three steps, as illustrated in Figure 7.2.

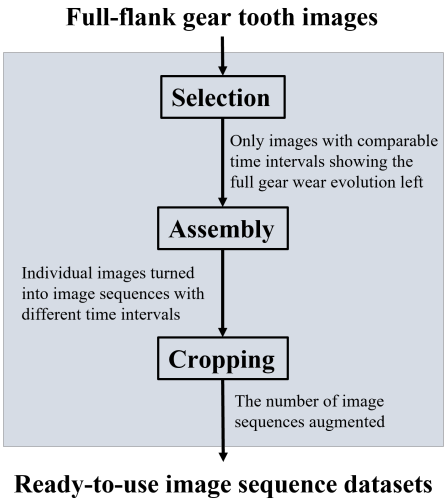


Figure 7.2: The three main steps, namely, selection, assembly and cropping, to prepare image sequence datasets.

(i) Selection of gear tooth images for datasets building (see Figure 7.3). The image sequence datasets were built using 36 batches (out of 41) of full-flank images of the driving gear. Firstly, the driven gear's images were not used because the driven gear was worn at a much slower rate than the driving gear. There was only mild to moderate macropitting on the driven gear at the end of the test while the driving gear was already severely worn. Then, 5 batches of full-flank images from the initial stage of the test were excluded because the time intervals were much shorter in the initial stage than the rest of the test. This ensured comparable time intervals between image batches. Therefore, 108 (3x36) full-flank images were used to build the datasets.

The training and validation datasets were made using only images from tooth No.2 and No.3 (see Figure 7.3 for the numbering) on the driving gear, while the test dataset was made using only images from tooth No.1. This split of dataset sources ensured that the image sequences in the test set were different from those in the training and validation sets and the model's generalisation capability could be properly tested.

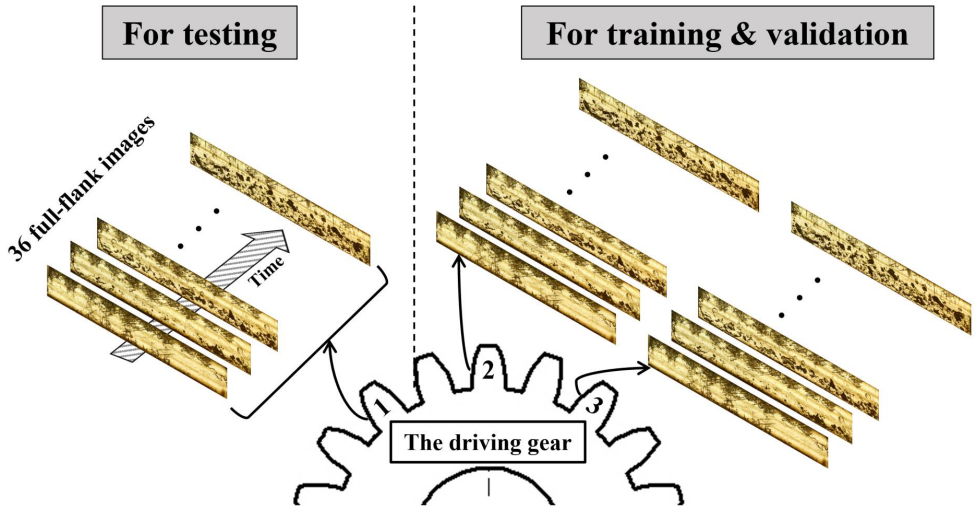
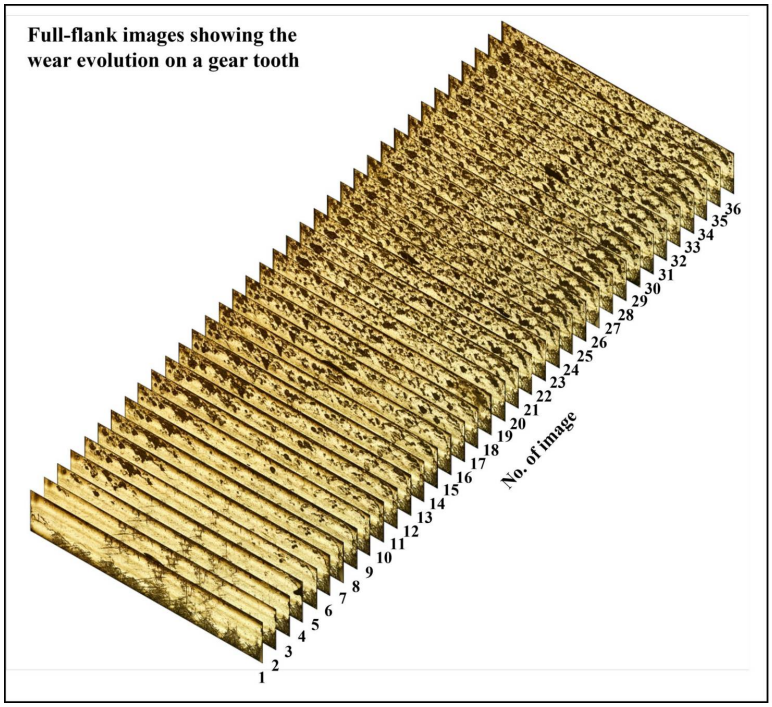


Figure 7.3: Selection of gear tooth images for datasets building.

(ii) Assemble individual images into image sequences (see Figure 7.4). This step was to create image sequences with different time intervals to make the model compatible with different gears in practice. In this study, sequence-to-one (5-to-1) prediction was adopted. That is, each sequence contains 6 gear tooth images, in which the first 5 images are provided as history images to the prediction model, and the last image is the ground truth. The number of history images (i.e., 5) was chosen to ensure the prediction reliability by providing sufficient history information and considering the restriction of moulds making in practice. Therefore, 6 chronologically ordered images from the same gear tooth were needed to assemble an image sequence.

Unlike image sequences of common videos with a constant frame rate and thus an even time interval between consecutive frames, the time interval between consecutive gear tooth images varies and depends on inspection times. In practice, inspection times are specific to equipment and many other factors, and a fixed time interval might not be reasonable for different gears. Therefore, in this study, image sequences with different time intervals were assembled, and the time information for each image was also provided to the model in training, validating, and testing. Specifically, there were 6 levels of time intervals between consecutive images in a sequence, from interval 1, corresponding to the shortest intervals,

to interval 6, corresponding to the longest intervals, as shown in Figure 7.4. For example, sequences with interval 1 consist of 6 consecutive images (e.g., image No. 1 to 6), and sequences with interval 6 consist of 6 images that were taken every other 5 images (e.g., images No.1, 7, 13, 19, 25, 31). For each gear tooth, there can be at most 31 sequences with interval 1, and 6 sequences with interval 6, as illustrated in Figure 7.4. In total, there were 333 (3 x 111) sequences of full-flank images.



Sequences with interval 1

No. of sequence	No. of image					
1	1	2	3	4	5	6
2	2	3	4	5	6	7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	31	32	33	34	35	36

Sequences with interval 2

No. of sequence	No. of image					
1	1	3	5	7	9	11
2	2	4	6	8	10	12
⋮	⋮	⋮	⋮	⋮	⋮	⋮
26	26	28	30	32	34	36

⋮

Sequences with interval 6

No. of sequence	No. of image					
1	1	7	13	19	25	31
2	2	8	14	20	26	32
⋮	⋮	⋮	⋮	⋮	⋮	⋮
6	6	12	18	24	30	36

Figure 7.4: Selection of individual images to assemble image sequences using different time interval levels. For each gear tooth, the selection was made among its 36 full-flank images (chronologically ordered, shown on the left).

(iii) Cropping of assembled full-flank image sequence for data augmentation (see Figure 7.5). The 333 assembled full-flank image sequences were presumably insufficient for training a deep learning model, and therefore, they were cropped to augment the datasets. Firstly, a rectangular window of 8850 x 1330 (pixel x pixel, equivalent to 18.06 mm x 2.71 mm) window was used to crop the full-flank images to remove unclear image borders (original image size was around 9000 x 1500). Then, a square window of 1330 x 1330 (pixel x pixel) moving along the face width direction was used to crop the border-cleared

images into smaller images. The step size of the moving window varied according to the interval level because of the different maximum number of sequences for different interval levels, as shown in Figure 7.4. This was for the balance of datasets across different interval levels so that the training of the wear prediction model was not skewed to a specific level. Moreover, the generated image sequences were manually inspected, and those with significant contamination (from moulds making) were removed. In the end, there were 5,183 image sequences (all from tooth No.2 and No.3 of the driving gear) for training and validation and 2,131 sequences (all from tooth No.1 of the driving gear) for testing. Each image sequence has the shape of 1330 x 1330 x 6 (width (in pixel) x height (in pixel) x frames).

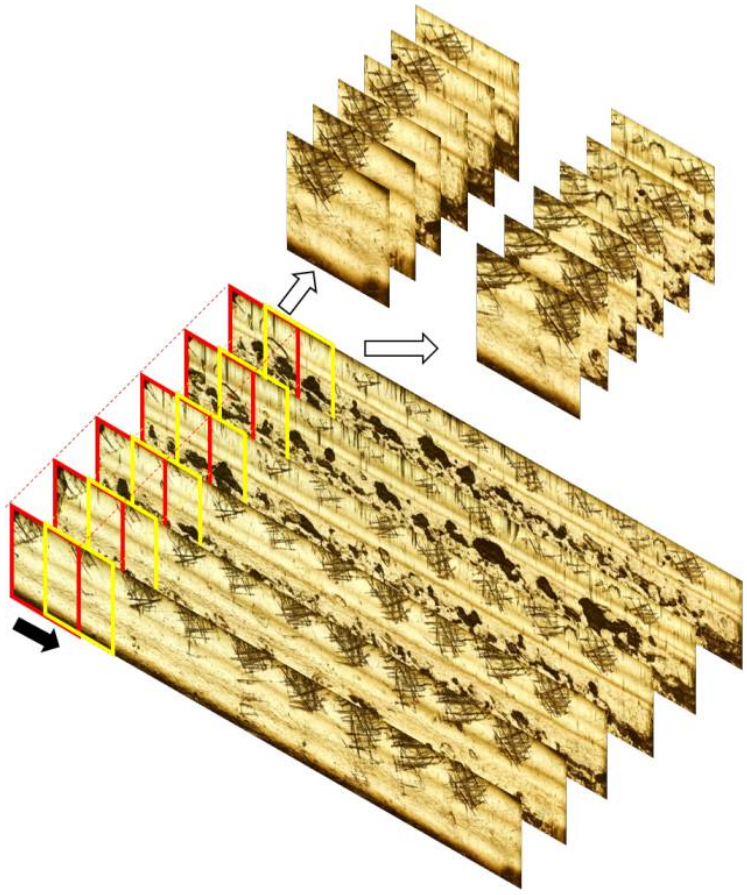


Figure 7.5: Cropping full-flank image sequences using a square moving window. The full-flank images were border-cleaned and had a size of 8850 x 1330. The size of the moving window was 1330 x 1330, and the step size varied according to the interval level.

7.2.2 Deep learning model for wear prediction

This section details the inputs, outputs, loss functions, and the architecture of the wear prediction model and its adversarial training procedures. Overall, the wear prediction model is a generative adversarial network (GAN) that consists of a generator and a discriminator, and the adversarial relationship between the two components is utilised to produce sharp and realistic predictions. The generator is a ConvLSTM network that models the spatiotemporal dimensions of image sequences to generate accurate predictions, while the discriminator is a simple convolutional neural network (CNN) used to distinguish generated images from real images.

7.2.2.1 Inputs, outputs and overall structure of the model

The wear prediction model is GAN consisting of a generator and a discriminator. The generator generates a future gear tooth image given a sequence of history images, and the discriminator discriminates generated images, or fake images, from real images. During training, the generator's objective is to ensure that the generated image is as close to the ground truth as possible and realistic enough to deceive the discriminator. The discriminator's objective is to correctly classify real and fake images, i.e., not to be deceived by the generator. Therefore, the generator and the discriminator will compete and ideally improve their capability to eventually reach a dynamic balance. Figure 7.6 illustrates the structure of the wear prediction model. The generator's inputs consist of 5 past images (the first 5 of an image sequence, T=6 in this study) and their corresponding time information (running cycles), and its output is a predicted gear tooth image. For the discriminator, its inputs consist of a fake image, i.e., the predicted image generated by the generator, and a real image, which can be randomly sampled from the image sequence dataset, and its output is the classification results, either '0' (fake) or '1' (Real).

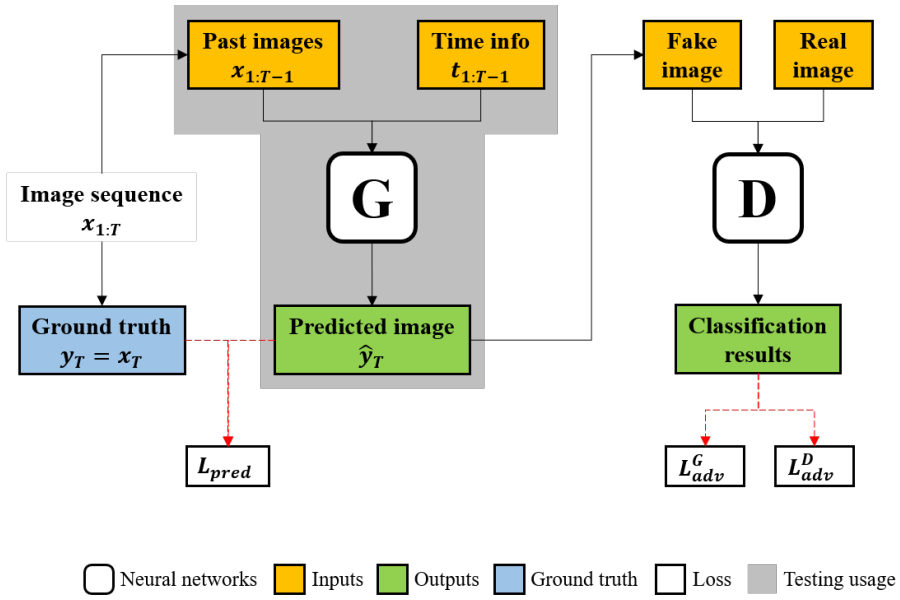


Figure 7.6: The GAN structure and its inputs, outputs, and losses. The 'G' block represents the generator and 'D' represents the discriminator. The predicted image is the fake image and acts as one of the inputs for the discriminator.

7.2.2.2 Losses of the generator and the discriminator

The loss function for the generator (see Equation 7.1), L^G, is a weighted combination of two parts: (i) prediction loss, L_{pred}, which measures the closeness of the predicted image to the ground truth, and (ii) adversarial loss, L^G_{adv}, which measures the chances of the discriminator misclassifying fake images as real images, i.e., the realism of predicted images. Thanks to the adversarial loss, the balance between realistic predictions and predictions of high similarity to the ground truth, can be adjusted by the parameter, λ_{adv}.

$$L^G = L_{pred} + \lambda_{adv} * L^G_{adv} \tag{7.1}$$

The prediction loss for the generator (see Equation 7.2), L^G_{pred}, can be broken down into: (i) mean absolute error (MAE, also known as l_1 loss)(see Equation 7.3), L_{MAE}, which is commonly used in video prediction but would cause blurry predictions [172, 243], and (ii) image gradient difference loss (GDL)(see Equation 7.4), L_{GDL}, which is used to sharpen

the predicted images [172]. The weight of the image gradient different loss can be adjusted by the parameter, λ_{GDL}.

$$L_{pred} = L_{MAE} + \lambda_{GDL} * L_{GDL} \tag{7.2}$$

$$L_{MAE} = \frac{1}{N} \sum_{i,j} \left| |Y_{i,j} - \hat{Y}_{i,j}| \right| \tag{7.3}$$

$$L_{GDL} = \frac{1}{N} \sum_{i,j} \left| |Y_{i,j} - Y_{i-1,j}| - |\hat{Y}_{i,j} - \hat{Y}_{i-1,j}| \right| + \left| |Y_{i,j} - Y_{i,j-1}| - |\hat{Y}_{i,j} - \hat{Y}_{i,j-1}| \right| \tag{7.4}$$

In Equation 7.3 and 7.4, $Y_{i,j}$ and $\hat{Y}_{i,j}$ is the pixel (i, j) of the ground truth and the predicted image, respectively, and N is the total number of pixels in either image (the two images must have the same size).

The adversarial loss for the generator (see Equation 7.5), L^G_{adv}, measures the chances of the predicted images successfully deceiving the discriminator and is calculated by comparing targets '1' (the generator aims to deceive the discriminator into classifying fake images as '1') and the discriminator's classification result on the predicted image.

$$L^G_{adv} = L_{BCE}(D(\hat{Y}), 1) \tag{7.5}$$

In Equation 7.5, L_{BCE} is the binary cross-entropy loss given by Equation 7.6, and $D(\hat{Y})$ represents the output of the discriminator (D) given the input of a predicted image $\hat{Y}$.

$$L_{BCE}(x, y) = -y * \ln(x) - (1 - y) * \ln(1 - x) \tag{7.6}$$

The discriminator's aim is to classify all real images as '1' and all fake images as '0', and therefore, its loss, L^D, is calculated by comparing '1' with classification result on a real

image or '0' with classification results on a fake image, as given by Equation 7.7

$$L^D = \begin{cases} L_{BCE}(D(R), 1) & for\ real\ images \\ L_{BCE}(D(\hat{Y}), 0) & for\ fake\ images \end{cases} \tag{7.7}$$

7.2.2.3 Architecture of the generator and the discriminator

The generator is a ConvLSTM network with an encoder-decoder structure and skip connections, as shown in Figure 7.7. LSTM is a type of recurrent neural network (RNN) that can learn long-term dependencies, i.e., utilising history information, even that came from a long term ago, to inform current processing [176]. When convolution is incorporated in LSTM, the modified version, i.e., ConvLSTM, can process image sequences and found suitable for video prediction [171, 173]. Detailed mathematical transformations inside a ConvLSTM layer can be found in [171]. After each ConvLSTM layer, layer normalisation normalises the outputs from the previous layer separately within a batch, and it was employed to stabilise training and improve convergence [176, 246]. As there are multiple convolutions in a ConvLSTM layer and its recurrent nature, the training of ConvLSTM requires much more computing power than the training of normal convolutional networks. Therefore, the input sequence was downsampled to 200 x 200 for each image, and grey-scale inputs were used.

The encoder-decoder structure firstly compresses the image sequences from 200 x 200 to 50 x 50 through the encoder and then reconstructs the sequences back to the original size (200 x 200) through the decoder. This structure was employed to create a bottleneck (the four 50 x 50 blocks in Figure 7.7) to force the network to learn the most important features that characterise the spatiotemporal dynamics in an image sequence. However, a substantial amount of information is lost during the max-pooling process, which could be important in synthesising local features in the output image. Therefore, skip connections (dashed lines in Figure 7.7) were used to preserve the information. In the middle of the bottleneck, another input of time information was provided to the network.

The discriminator is a simple CNN similar to AlexNet [132]. Note that the input of the discriminator is individual images instead of image sequences. The image size was reduced via convolutional layers with a stride of 2 along both the width and height. The leaky rectified linear unit (ReLU) was used as the activation function for the convolutional layers based on its superior performance to the standard ReLU [247].

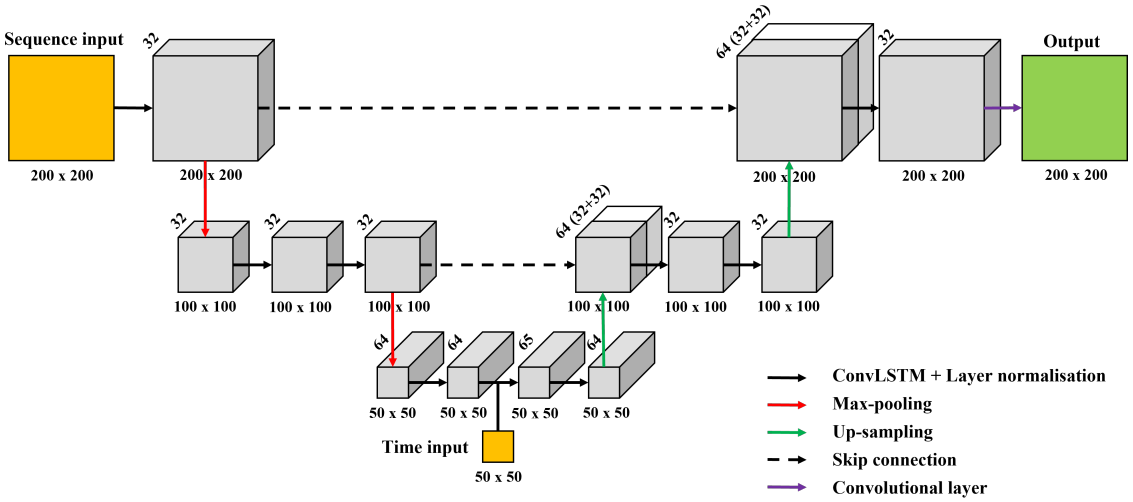


Figure 7.7: The architecture of the generator. The blocks only illustrate the width, height (below each block) and number of feature maps (along the depth of each block), and there are 2 other dimensions: batch size and number of past frames. For instance, the full shape of the sequence input is (batch size, 5, 200, 200, 1), and the first hidden layer has the shape of (batch size, 5, 200, 200, 32). The last hidden layer has the shape of (batch size, 1, 200, 200, 32).

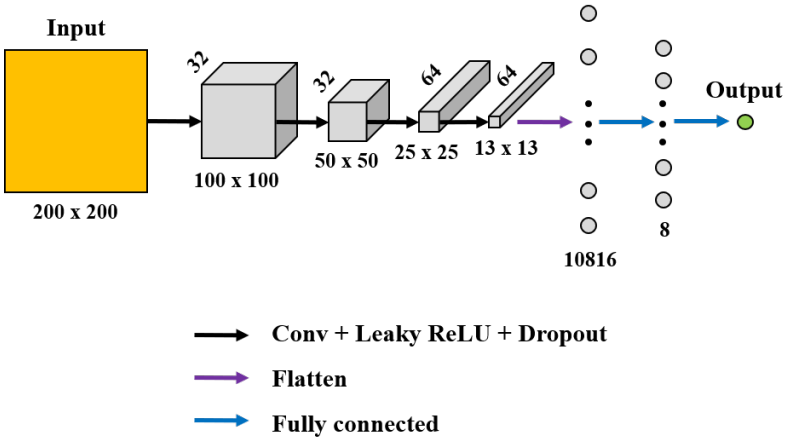


Figure 7.8: The architecture of the discriminator.

7.2.2.4 Adversarial training procedures

As illustrated in Figure 7.9, the adversarial training procedures started from taking a batch (M) of image sequences and time tensors to the generator. Then, the M predicted images and M real images were put in the discriminator, and classification results ('0' or '1') for fake images and real images were obtained, respectively. Losses for the generator and discriminator were then calculated based on the functions described in section 7.2.2.2. The gradients were calculated based on the losses using Adam, a method of stochastic gradient descent (SGD), and the trainable parameters of the generator and discriminator were updated, respectively, thus finishing a training step. The training step was repeated until the training datasets were exhausted.

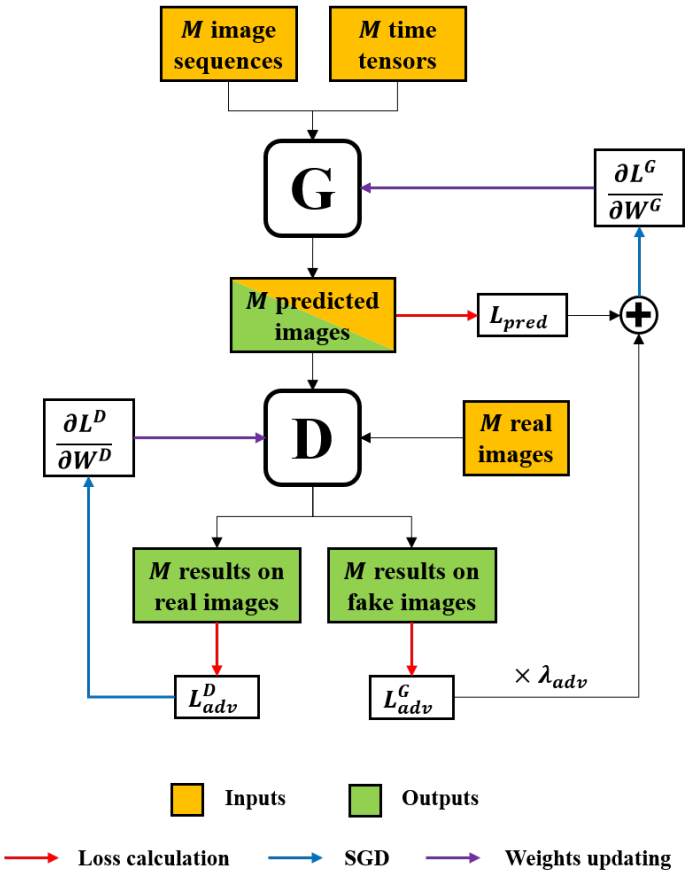


Figure 7.9: The adversarial training procedures.

Then, validation started, and it was similar to training, except that gradients were not calculated, and the parameters of the generator and discriminator were not updated. Validation was finished when the validation dataset was exhausted. A training epoch was finished when both training and validation datasets were exhausted, and it was repeated until convergence. Convergence was considered achieved when the losses for the generator and discriminator reached a dynamic balance and predicted images were reasonably close to the ground truth.

7.2.3 Evaluation of predicted images

Several important aspects need to be considered when evaluating outputs of a video prediction model such as their realism and accuracy [243]. Regarding the specific task of gear wear prediction, the predicted images should be realistic and sharp enough for engineers to recognise various wear features on the tooth flank and also accurately capture the dynamics of wear propagation.

Therefore, both qualitative and quantitative evaluations were conducted to assess the realism and accuracy of predicted images in the context of gear wear prediction. Qualitative evaluation was included because common quantitative metrics often produce different results from qualitative evaluation made by humans [172, 173, 243]. Recent studies on video prediction have found that traditional metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are biased towards blurry predictions instead of sharp ones with minor errors [173, 243, 248].

For quantitative evaluation, 4 metrics, namely multi-scale structural similarity index measure (MS-SSIM), VGG cosine similarity, pitted area error (PAE), and pitted region error (PRE), were employed. MS-SSIM was used to evaluate the closeness of the predicted images to the ground truth, and it has shown equal or better performance to SSIM [249], which is commonly used pixel-wise metric. VGG cosine similarity is the cosine similarity between two feature vectors extracted from the VGGNet [250], given the predicted image and its ground truth as the input, respectively. Studies have shown that such a metric

based on deep features is better at measuring the perceptual similarity of two images than traditional pixel-wise metrics [172, 248].

For the task of gear wear prediction (prediction of pitting propagation in this study), PAE and PRE were developed to measure the error of pitted area/region in a predicted image compared to that in the ground truth. The two metrics were used to examine if the model accurately captured the dynamics of wear propagation. The method to extract macropits from gear mould images and calculate pitted area and pitted region can be found in Chapter 5. It is also suggested that PRE could be a more accurate measure of pitting propagation than PAE when there is a shrinking of pits [251]. Equation 7.8 is the formula to calculate PAE (%), where $\hat{c}_{pits}$ (unitless) is the area coverage of macropits (see Equation 5.1) calculated using predicted images, while c_{pits} was calculated using the ground truth. PRE was obtained using a similar approach, where c_{pits} in Equation 7.8 was replaced with $c_{pitted\,region}$ (see Equation 5.2), which measured the area coverage of pitted region.

$$PAE = \frac{|\hat{c_{pits}} - c_{pits}|}{c_{pits}} \times 100\% \tag{7.8}$$

7.3 Results and discussion

This section provides qualitative and quantitative evaluation results using test data, from which the model's capability in predicting gear wear, in terms of the realism and accuracy of predictions, can be assessed. The results also demonstrate the effect of different time intervals of the image sequence on the prediction quality.

7.3.1 Qualitative evaluation of predicted gear wear images

This section shows the results of qualitative evaluation of predicted gear wear images, as shown in Figure 7.10, Figure 7.11, and Figure 7.12 that demonstrate the capability of the

wear prediction model to predict the gear surface change in the wear process. The effect of time interval on the prediction quality is also presented by comparing the three figures. Figure 7.10 shows a long-term prediction based on an image sequence with interval 6 (the longest interval level), and Figure 7.11 shows a mid-term prediction based on a sequence with interval 3, with Figure 7.12 showing a short-term prediction based on a sequence with interval 1.

In terms of the realism of predicted images, it can be seen that the wear prediction model is able to generate realistic images of gear tooth flanks and clear surface features for wear assessment, such as macropits and scratches. However, minor features such as minor cracks and dents are not generated in the predictions. By comparing the three figures, it was found that the model tends to generate more realistic images for short-term predictions. In Figure 7.12, the prediction includes clear surface features of different sizes and little blurriness, making it difficult to be distinguished from real images. In contrast, the prediction in Figure 7.10 showed relatively blurry scratches and is easier to be identified as a generated image.

In terms of prediction accuracy, the examples demonstrate that the model made plausible predictions on wear propagation, which were reasonably aligned with the ground truth. By comparing the predicted images and their last frames (the last frame of the five past frames), it can be found that the model was not simply copying the last frames but creating plausible changes on the tooth flank, e.g., new pits generated and old pits shrunk. It is worth noting that the predicted changes of surface features might not be the same as those in the ground truth, e.g., the locations and sizes of new pits. However, the overall trend of wear propagation was not misjudged, e.g., the total pitted area and pitted region (see Figure 7.10). Moreover, the effect of the time interval on the prediction accuracy was not obvious.

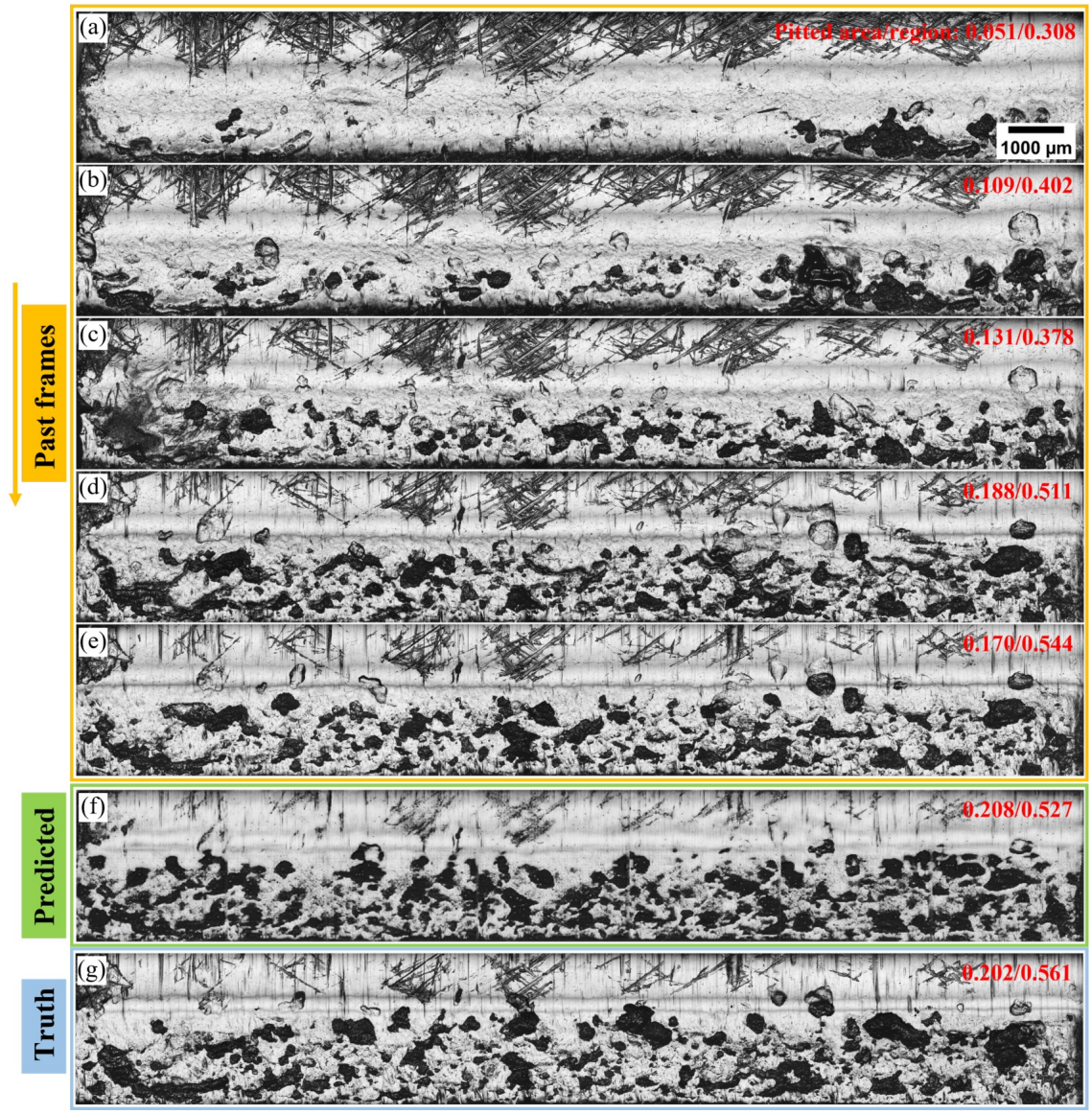


Figure 7.10: (a-e) Five past frames with the longest interval level (interval 6, see Figure 7.4), (f) the predicted image, and (g) the ground truth of a gear tooth flank in the lubricated test. The running cycles for the past frames and the ground truth are 0.57, 0.82, 1.28, 2.14, 2.64, and 3.25 million cycles, respectively. c_{pits} and $c_{pitted\,region}$ (both unitless) are also indicated on the top right of each image.

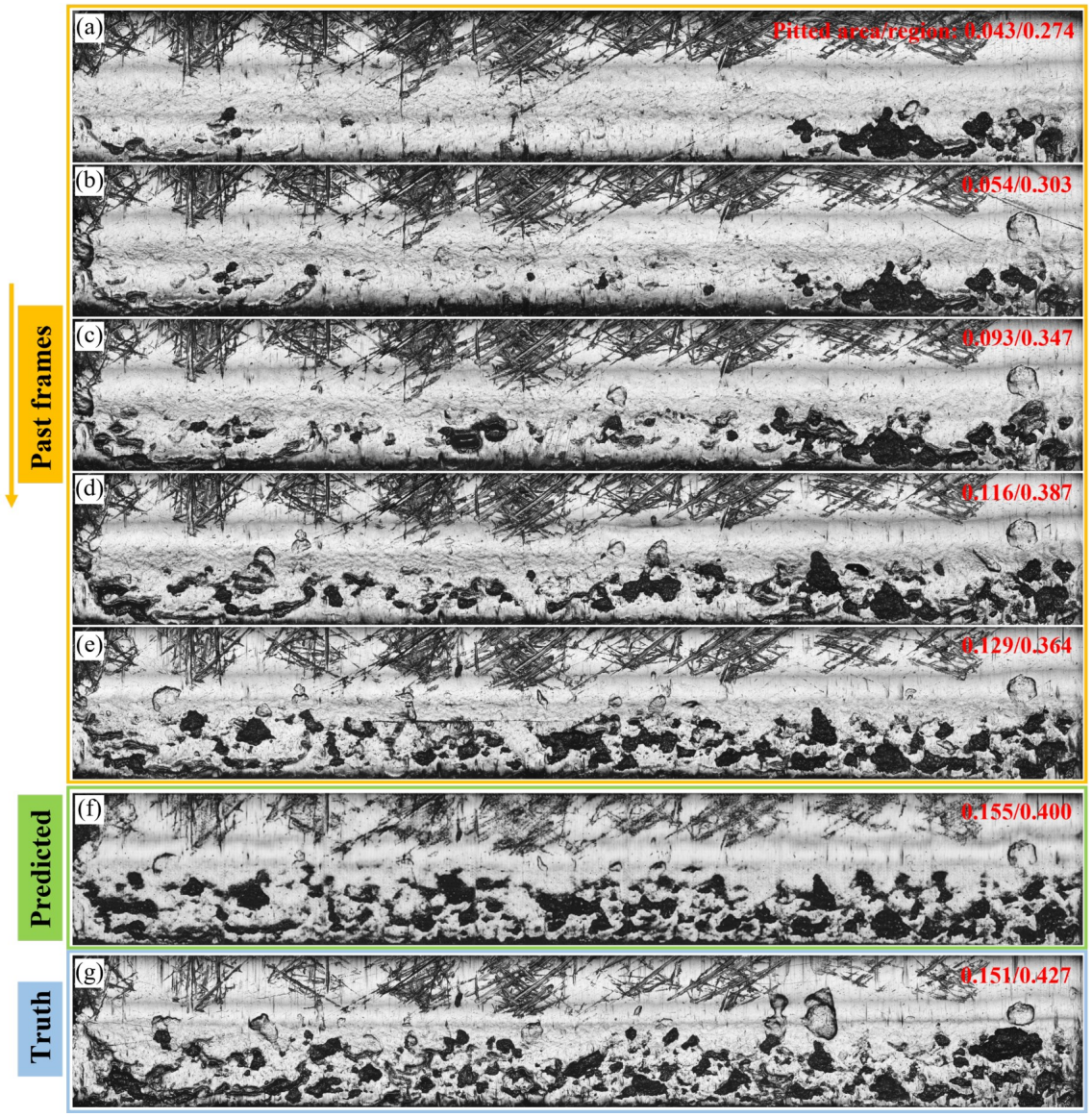


Figure 7.11: (a-e) Five past frames with a medium interval level (interval 3, see Figure 7.4), (f) the predicted image, and (g) the ground truth of a gear tooth flank in the lubricated test. The running cycles for the past frames and the ground truth are 0.51, 0.65, 0.75, 0.91, 1.16, and 1.69 million cycles, respectively. c_{pits} and $c_{pitted\,region}$ (both unitless) are also indicated on the top right of each image.

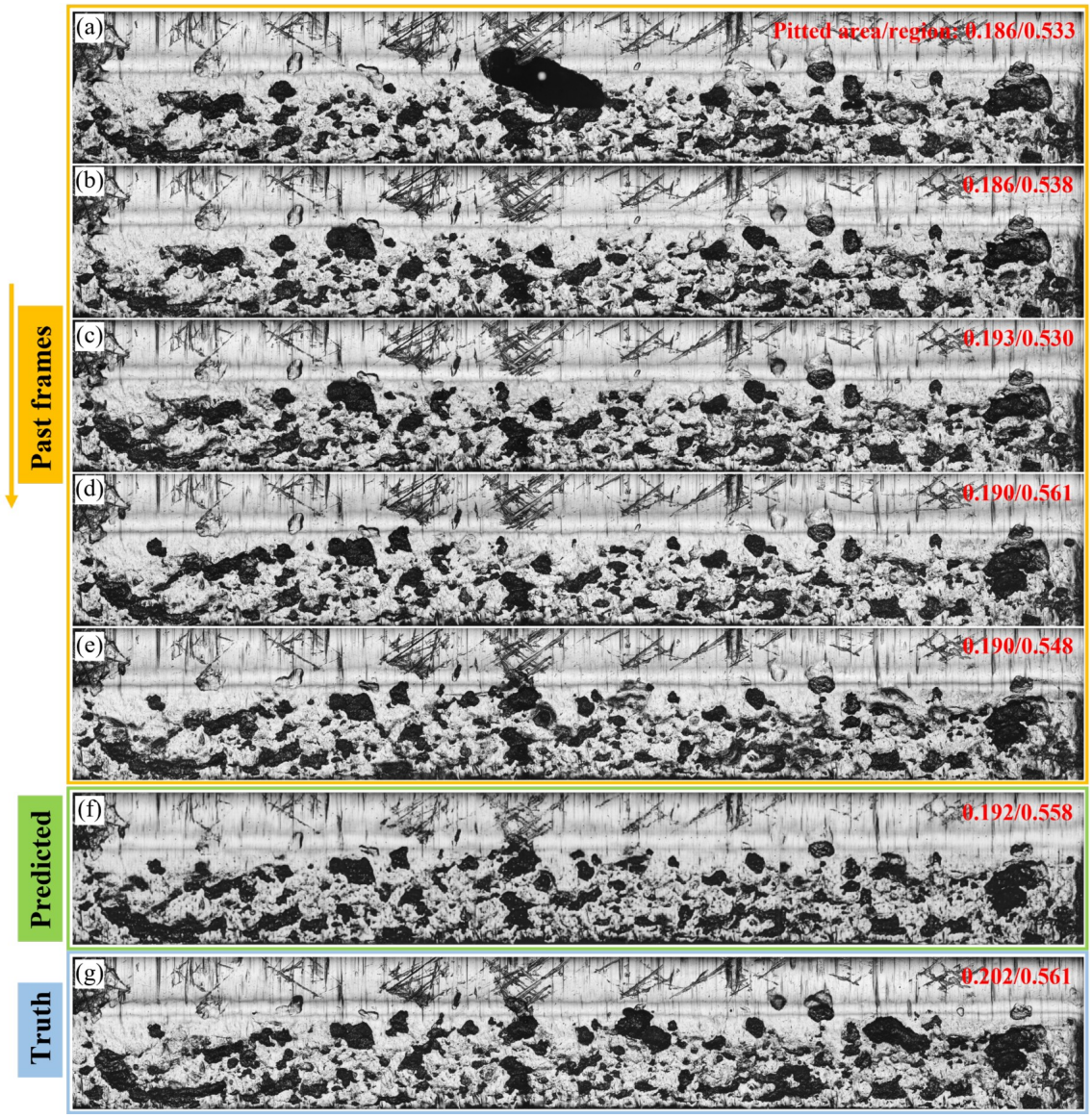


Figure 7.12: (a-e) Five past frames with the shortest interval level (interval 1, see Figure 7.4), (f) the predicted image, and (g) the ground truth of a gear tooth flank in the lubricated test. The running cycles for the five past frames and the ground truth are 2.72, 2.80, 2.88, 2.97, 3.14, and 3.25 million cycles, respectively. c_{pits} and $c_{pitted\,region}$ (both unitless) are also indicated on the top right of each image.

7.3.2 Quantitative evaluation of predicted gear wear images

This section shows the quantitative evaluation results, as shown in Figure 7.13 that quantifies the discrepancy between generated predictions and their ground truth using the four metrics mentioned in section 7.2.3. The effects of time intervals on the four metrics are also presented. For each interval level, the metric score is the average of all test sequences with the same interval level. The first two metrics, MS-SSIM and VGG cosine similarity, measure the similarity between a predicted image and its ground truth, and the higher score corresponds to the better model performance. For MS-SSIM, the model achieved the highest score of 0.605 when making short-term predictions (interval 1, see Figure 7.4) and the lowest score of 0.367 when making long-term predictions (interval level 6), with an average of 0.464±0.116 across all interval levels. Overall, an inverse relationship between interval level and MS-SSIM can be observed. For VGG cosine similarity, the model achieved high scores across all interval levels, with an average score of 0.994±0.002, suggesting a high similarity between the predictions and ground truth in terms of their underlying features. Moreover, a decreasing trend of the VGG cosine similarity score with respect to increasing interval levels could be observed, but the relationship is not as clear as that of MS-SSIM.

The last two metrics, PAE and PRE, focus on the main tribological features (macropits) on gear teeth and measure the model's capability to capture the dynamics of wear propagation. Positive scores correspond to overestimation by the model, and negative scores correspond to underestimation. For PAE, the model achieved an average error rate of 0.652%±7.3%. For short- to medium-term predictions (interval 1-3, as shown in Figure 7.4), the model tended to underestimate pitted area growth, while for medium- to long-term predictions (interval 4-6), the model tended to overestimate the growth. For PRE, the model achieved an average error rate of -2.56%±4.21%, and the variation range of error rate was lower than that of PAE. For most interval levels, the model tended to slightly underestimate the pitted region growth. Overall, metrics using extracted features from images, including VGG cosine similarity, PAE and PRE, suggest that the model generates high-quality and accurate predictions. On the other hand, the traditional pixel-

wise metric, MS-SSIM, suggests that the prediction accuracy is reasonable for short-term predictions but could notably decrease with extended interval levels.

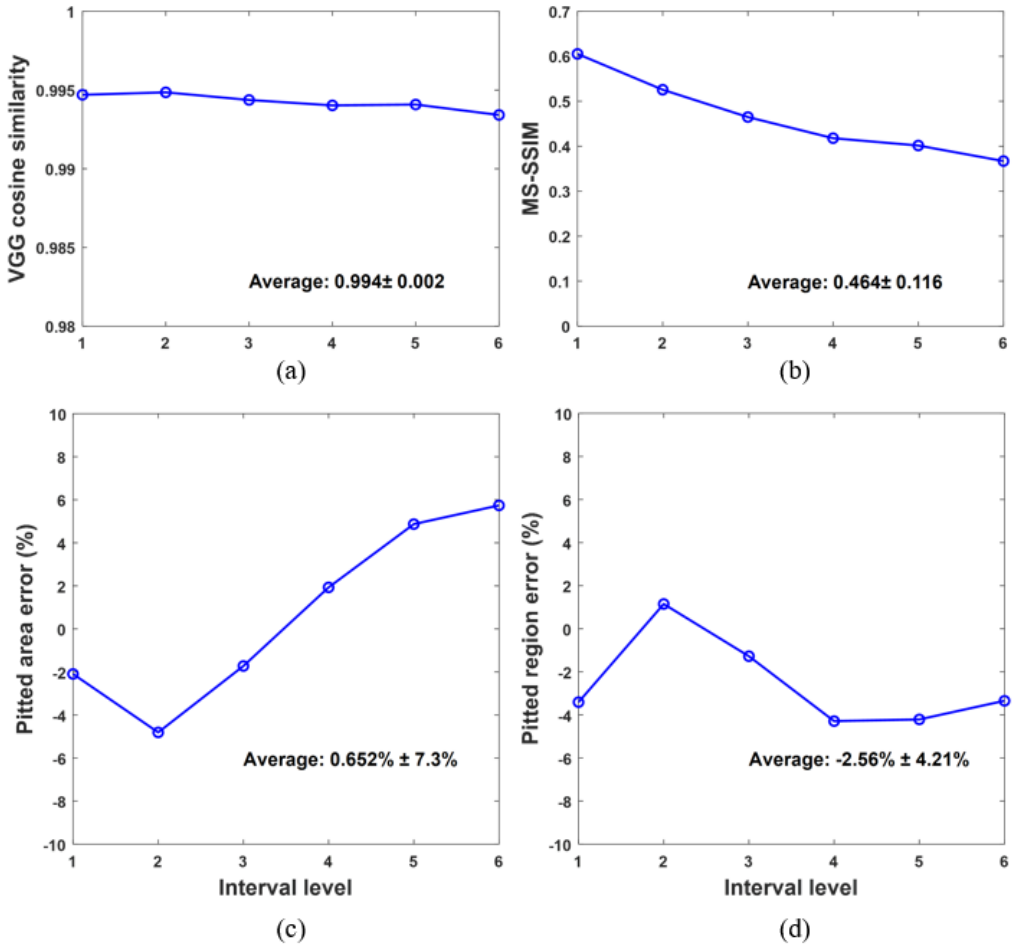


Figure 7.13: Evaluation results of the four metrics with respect to different interval levels. (a) VGG cosine similarity, (b) MS-SSIM, (c) pitted area error, and (d) pitted region error.

7.3.3 Discussion

The ConvLSTM wear prediction model was designed to extract high-level features from image sequences of gear tooth flank and generate realistic and accurate predictions for gear prognosis. Test results have shown that the model has achieved the objective. Qualitatively, the predictions preserve essential tribological features for wear mechanism and severity assessment, and they are well aligned with the ground truth in terms of the prop-

agation of different wear mechanisms on the tooth flank. For image sequences with the shortest interval level (interval 1), the predictions are very realistic and difficult to be distinguished from real images. Quantitatively, the model achieves excellent scores on three metrics regarding both the realism and accuracy of predictions, including VGG cosine similarity (for realism, based on deep features) and PAE and PRE (for accuracy, related to pitting propagation prediction). Compared to other video prediction works [173, 252], the model in this study generates predictions of equivalent or higher scores in terms of VGG cosine similarity. It is also demonstrated that the model prediction accuracy does not decrease when the interval level extends, as reflected by the PRE scores.

Another important aspect of the model is that the predicted image is closer to its ground truth than its previous image in pitted area/region prediction. In other words, the model does not simply copy the last frame (of 5 past frames) and make trivial changes. Instead, normal events related to pitting evolution, such as the generation of new pits and shrinkage of existing pits, can be observed in predictions (see Figure 7.10, Figure 7.11, and Figure 7.12). To quantify this relationship, a pitted area/region-based metric named relative closeness ratio (RCR) was adopted, which measures the relative closeness between the prediction and its ground truth and previous image. Figure 7.14 shows the definition of RCR and scatter plots of RCR for pitted area and pitted region. It can be observed from the figure that a majority (72.2%) of predictions are closer to the ground truth than their previous images.

In terms of MS-SSIM, the model only achieved reasonable scores (around 0.6) for short interval levels, and the scores decreased to 0.4 for long interval levels. However, the scores could be impacted by the inherent defects of the gear tooth images. MS-SSIM is a pixel-wise index for image similarity, which means that two identical images could obtain a low score if one image is slightly shifted, as shown in Figure 7.15. In this study, manual adjustment of the position of the gear tooth mould and image stitching were required to obtain a flat gear tooth image for the curved tooth flank. As a result of the positioning error and stitching error, the gear tooth images could not be fully aligned with each other, and such inherent defects limit the MS-SSIM score of a prediction regardless of its

qualitative realism and accuracy.

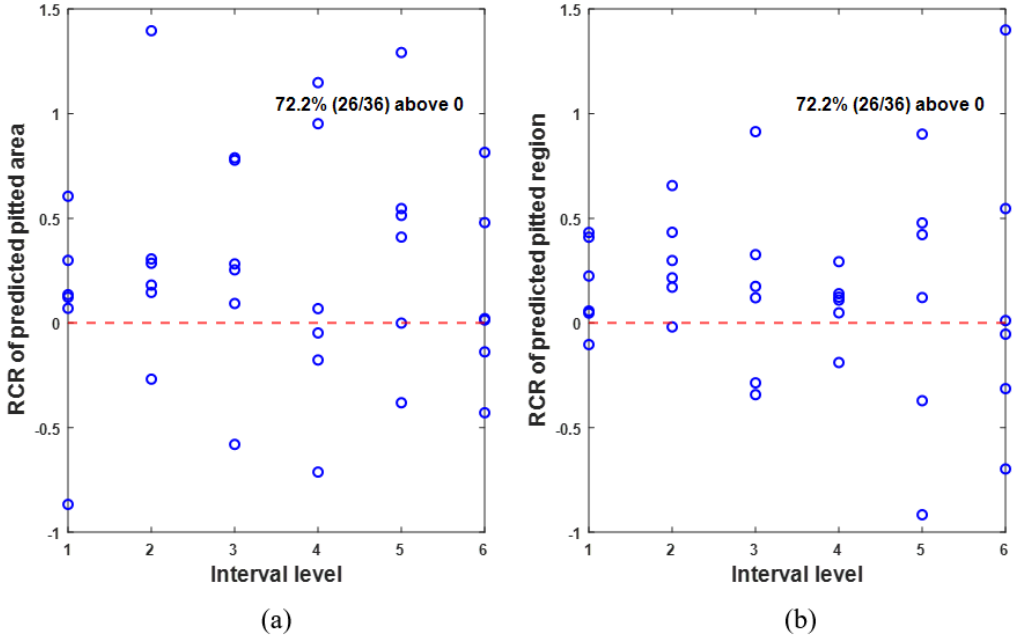


Figure 7.14: The relative closeness between prediction and its ground truth and previous image in (a) pitted area and (b) pitted region. RCR above 0 means that the prediction is closer to its ground truth than its previous image and vice versa.

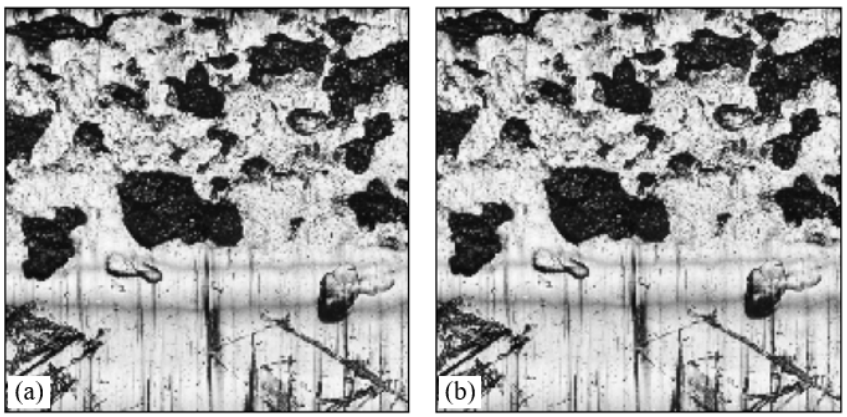


Figure 7.15: An example of MS-SSIM's sensitivity to image position. (a) The original image (200x200) and (b) its shifted version (to upper left by 2x2). The MS-SSIM between two images is 0.6092.

The model in this study is the first one for gear wear prediction that generates realistic and accurate predictions utilising image sequences of gear tooth flank. The model employs mixed inputs (image sequences and time information), and it allows great flexibility of gear tooth inspection and imaging in practice. Moreover, the model can be generally applied to different wear mechanisms and mechanical components since it only relies on image data and time information to make predictions.

7.4 Summary

A novel deep learning model for gear wear prediction was proposed in this chapter. The model takes an image sequence (5 history images) of the same gear tooth flank and the corresponding time information as the inputs and generates a future image of the tooth flank as the output. A dataset of gear tooth image sequences was built using the moulding method from a long-running gear test, and the training and test datasets came from different gear teeth. Considering varying gear inspection times in practice, image sequences with six different time intervals were included in the dataset. The model is a GAN of which the generator is a ConvLSTM network to capture the spatiotemporal dynamics in an image sequence. To improve the realism of predictions, image gradient difference losses and adversarial losses were combined with mean absolute error. Both qualitative and quantitative evaluations were used to ensure a fair measure of the predictions. The effect of time interval on prediction quality was also studied. Test results have shown that the model can generate realistic and accurate predictions for different interval levels.

Findings are summarised as follow.

- The qualitative evaluation of the test results showed that the model preserves essential tribological features in predictions for wear mechanism and severity assessment, and it also predicts surface changes aligned with propagation trends of different wear mechanisms.
- The quantitative evaluation showed that the model performed well on three metrics,

including VGG cosine similarity (0.994±0.002), PAE (0.652%±7.3%) and PRE (-2.56%±4.21%).

- As the time interval of the image sequence increased, the realism of the predictions reduced, but the effect on prediction accuracy was not obvious.
- The model achieved reasonable scores of MS-SSIM for short interval levels (~ 0.6), but the score decreased to around 0.4 for long interval levels, which might be affected by positioning error during imaging.

The model can be extended to other wear mechanisms and mechanical components since it does not require expert knowledge of specific wear problems. The main limitation of this work is that stochasticity was not incorporated in the deep learning model, causing the difficulty in precisely predicting the locations of pits generation, and future work to improve the model will be presented in Chapter 9.

Chapter 8

Discussion

This research developed a set of advanced techniques for gear wear monitoring and prediction. Detailed discussions of these developments can be found in the earlier chapters, in particular, sections 4.5, 5.4, 6.3.4 and 7.3.3. This chapter, from an integrated perspective, discusses (i) potential applications of the developed moulding-imaging techniques, (ii) further insights into WDA for gear wear monitoring, (iii) AI-assisted gear wear assessment and prediction, and (iv) strengths and limitations of the developed techniques.

8.1 Potential applications of the developed moulding-imaging techniques

The combination of surface replication and image analysis, or the moulding-imaging techniques, were developed and applied in this thesis to obtain direct wear information on worn surfaces. Compared to other existing techniques, the novelty of the techniques compared to other existing techniques is the ability to obtain and utilise a variety of direct gear wear information (e.g., macro- and micro-scale surface appearance, tooth profile, surface roughness, pitted area, see Chapter 4). Assisted with the impressions of worn surfaces and the extracted wear information, early warnings for wear-related gear faults could be

provided with high accuracy and reliability. Therefore, the techniques could be potentially applied in industrial maintenance and serve as an efficient tool to reduce machine downtime and maintenance costs. Table 8.1 summarises suggested procedures to implement gear wear monitoring and prediction based on the moulding-imaging techniques for industrial gearboxes.

Table 8.1: Suggested procedures to implement gear wear monitoring and prediction based on the moulding-imaging techniques.

No.	Operation procedure
1	Stop the gearbox
2	Remove the gearbox's inspection cover and rotate the input shaft to move the marked teeth (marking should be done during the first inspection) to an accessible position (above the lubricant level)
3	Clean the teeth using an air tube and remove any remaining oil on them using absorbent materials
4	Apply the replication compounds in the valley between two teeth and ensure the entire tooth flanks are covered
5	Take off the cured moulds after suggested curing time (20 mins at 25°C for the material used in this thesis)
6	Capture images of collected moulds using an optical microscope
7	Pre-process the acquired mould images (e.g., stitching, cropping, stacking) to prepare valid inputs for the two-level wear assessment system (see Chapter 6) and the wear prediction model (see Chapter 7)
8	Input the pre-processed mould images into the two-level wear assessment system for automated wear mechanism and severity assessment
9	Input prepared mould image sequences into the wear prediction model for gear wear prediction. Note that the predicted gear images can also be put into the wear assessment system for automated wear assessment
10	Submit gear wear assessment and prediction results for maintenance decision-making (beyond the scope of this thesis)

The procedures as shown in Table 8.1 can be implemented during regular inspections to minimise interruptions to the normal operation of a gearbox. During gearbox stoppage, maintenance engineers could remove the inspection cover of a gearbox (see Figure 8.1 [253,254]) and then implement the moulds making procedures. For some gearboxes within complex structures that cannot directly accessed, it could be difficult to acquire gear moulds using the suggested procedures. Firstly, as mentioned in section 4.2.2, the target tooth flank(s) must be cleaned to be oil-free to ensure a proper curing of the replication

compounds. Usually, it can be done by using an air tube to blow off most of the oil on the tooth flank(s) and then using absorbent materials to remove the remaining oil. It is worth noting that the absorbent materials should leave no contaminants on gear teeth or in the lubricating oil. Then, replication compounds can be conveniently applied to the tooth flank(s) using a dispenser with an extended nozzle (see Figure 3.4). After the curing process, which takes around 20 minutes for the replication compounds used in this research, gear moulds can be taken off easily, and no contaminants will be left on gear teeth. It is estimated that the entire process, from removing the inspect cover to collecting cured gear moulds, could be finished in less than one hour. Therefore, it is feasible to implement the techniques to regularly collect gear moulds to record gear wear evolution. Compared to other visual inspection methods such as using an endoscope or a videoscope, of which the imaging process must be finished within the inspection session, gear moulds can be stored for an extended amount of time and examined by various microscopy repetitively.

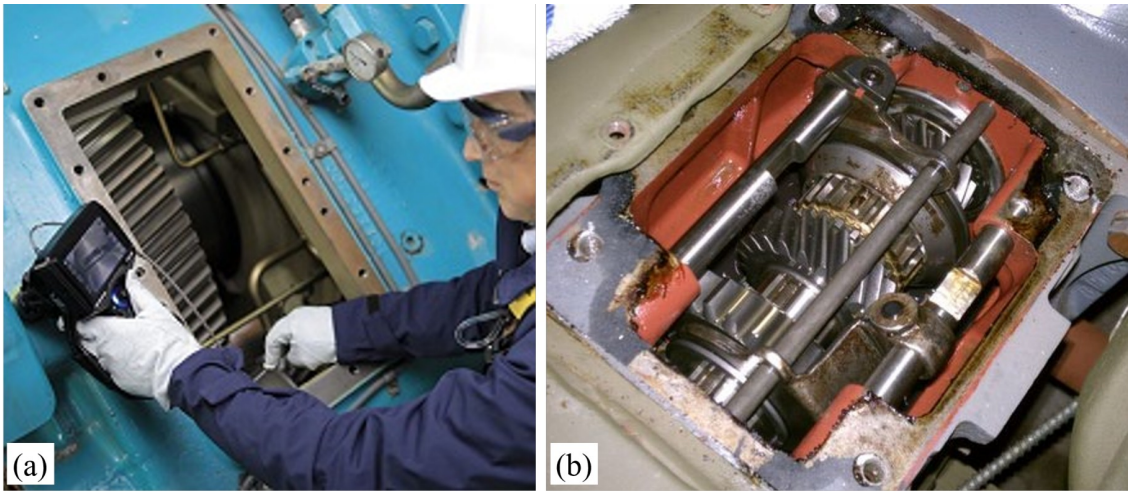


Figure 8.1: Examples of gearboxes with the inspection cover removed. (a) A wind turbine gearbox [253]. (b) An automotive gearbox [254].

The techniques could also be applied to conduct fundamental studies on gear wear to better understand a complex gear wear process such as combination of micropitting and tribocorrosion in offshore drive trains [255, 256] and in a controlled (i.e., lab-based) environment. As reviewed in section 4.2, in laboratory gear tests, detailed surface imaging is

usually only conducted before and after the test, and the evolution of gear wear is usually examined by simple visual inspections. By regularly applying the developed techniques during laboratory gear tests, gear moulds and mould images that record the wear evolution in great detail could be obtained. Compared to industrial machines, laboratory test rigs are usually allowed to stop more frequently, making the implementation of moulds making more feasible. Mould images as well as other quantitative data can be used to study a range of topics including the evolution of a gear wear process, wear performance of new gear materials, anti-wear properties of lubricants and additives.

In terms of gear wear prediction, the insights into gear wear evolution obtained in this research, particularly in Chapter 4, could be used as experimental evidence for the development of physical models of gear wear. For instance, 3D mould images of macropits could be used to calculate the inclination angle of a crack face with respect to the tooth surface, which is an important initial parameter used in numerical simulation of pitting propagation [37, 157]. Moreover, for gear pitting models that predict the distribution of fatigue pitting [40,152], wear maps such as Figure 4.13, Figure 4.14, and Figure 6.21 could be used to quantify the prediction accuracy.

The moulds making process requires a gearbox to be stopped, and therefore, the techniques may not be suitable to provide real-time wear monitoring when it is difficult to access gear teeth. However, regular inspections are commonly implemented to maintain industrial machines, and such opportunities can be utilised for moulds making. For example, wind turbines are typically maintained at least twice per year [257, 258], and regular inspections are conducted more frequently [259]. For machines that do not need to operate continuously, such as helicopters and mining excavators, there could be more windows to implement the techniques to collect direct wear evidence for further analysis or trending, which is vital for RUL prediction.

The direct gear wear information obtained via the combination of surface replication and image analysis has not only been used in gear wear monitoring and prediction per se, but also provided and could provide more important supports for the development of other CM techniques. For techniques such as vibration and AE analysis, direct gear

wear information is essential to establish the correlations between proposed signal features and real surface conditions of gears. The developed techniques in this research have been used in several studies on vibration and AE analysis. In [50], two AE-based gear wear indicators (cyclic kurtosis and cyclic scale parameter) were proposed based on the cyclostationarity of AE. Direct gear wear information, including the precise region affected by macropitting and the growth of total pitted area, served as crucial evidence to quantify the two indicators' sensitivity to different gear wear mechanisms (cyclic scale parameter being more sensitive to macropitting while cyclic kurtosis being more sensitive to abrasive wear). Similarly, in [51], gear moulds images showing detailed evolution of different gear wear mechanisms and the growth of pitted area were employed as important validations in developing vibration indicators for wear mechanism identification. The techniques are generally suitable for experimental studies on other gear wear monitoring and prediction techniques due to the ability to accurately quantify gear wear throughout an entire wear process. Combining the developed techniques in this research and other CM methods (e.g., WDA, vibration analysis) that provide high-frequency or real-time monitoring could potentially bring a synergistic improvement to maintenance reliability. This is because the techniques could provide accurate visualisations of current gear wear and predictions of wear evolution, and the information can be used by other CM methods to make informed adjustments to their monitoring strategies. In turn, real-time CM data could be crucial to optimise the inspection schedule for moulds making.

8.2 Further insights into WDA for wear monitoring

The developed techniques presented in Chapter 4 and discussed in section 8.1 were used to investigate the relationship between worn gear teeth and wear particles in the same wear process. The relationship has a fundamental effect on the implementation of OA/WDA, and the techniques could help deepen the understanding of this relationship. From the investigation in Chapter 5, a major finding is that there was a significant delay (0.85 million cycles) between the variation of wear debris concentration and that of pitted area/region

(see Figure 5.8) and the delay was also found regarding the variation of four numerical features (e.g., aspect ratio) of wear debris and macropits. Reports of such delays were not seen in the literature, but studies on the life cycle of wear debris provide potential explanations. As reviewed in section 2.2.3.2, wear particles could be entrapped between gear contacts after their detachment [10, 25] and experimental tests [10, 12, 103] showed the particle concentration between gear contacts is often much higher than that in the bulk oil. Those findings suggest that the delay might be caused by prolonged entrapment of wear particles between gear contacts. This hypothesis is also supported by many dents that were observed on gear teeth in the lubricated test, especially on driven gear teeth, as shown in Figure 5.9.

Nevertheless, the finding of the delay provides potentially valuable evidence for the use of OA/WDA to monitor gear wear. That is, when the real-time reading of wear debris concentration suggests the occurrence of abnormal gear wear, the actual wear severity could be worse than the suggested level. The extent of the delay could be affected by factors such as gear materials and lubrication systems. For example, in the lubricated test, the oil circulated within the gearbox (no external loop), and the gear material was relatively soft, making it easy for newly generated wear debris to accumulate near the meshing region and be entrapped. Entrapped particles, before their release into bulk oil, may not be detectable by existing methods but there could be a reasonable portion of newly generated particles that are not entrapped between contacts. Among those unaffected particles, the abnormal particles (e.g., fatigue particles, severe sliding particles) could be important indications of the occurrence of abnormal wear without delay. For example, in the lubricated test, the growth rate of pitted area reached its first peak at around 0.6 million cycles (see Figure 5.8), while that of particle concentration remained very low. However, by examining particle images obtained at 0.58 million cycles, of which one was shown in Figure 5.7 (b2), a few large fatigue particles were found, indicating the occurrence of fatigue pitting without a noticeable delay. Therefore, the ability to capture and identify abnormal wear debris might be necessary to provide early warnings of abnormal gear wear. As mentioned in section 2.2.2, particle identification usually requires not only the size of each particle to be captured but also its shape and morphology. Therefore, further

development of online wear debris imaging is necessary to overcome this phenomenon.

8.3 AI-assisted gear wear monitoring and prediction

Artificial intelligence (AI) technologies were used in this research to process gear mould images to automate gear wear monitoring and prediction. Specifically, deep CNNs were employed to classify mould images and wear debris images into different wear mechanisms and severity levels, and a GAN with a ConvLSTM generator was employed to generate future gear tooth images using history images. The use of deep learning significantly simplified the process of image-based wear assessment and prediction, which often involved technically demanding and time-consuming procedures such as image segmentation, feature extraction, and statistical modelling (in prediction), as reviewed in Chapter 2. The developed techniques were able to automatically generate and visualise the results of wear assessment and prediction through wear maps (see Figures 6.21 and 6.22) and predicted images (see Figures 7.10, 7.11, and 7.12) in several seconds (using a normal CPU) or under one second (using a normal GPU). Ideally, the techniques could enable maintenance engineers to carry out the entire process of gear wear monitoring and prediction, from moulds making, to mould imaging, to data input into deep learning models, and to results interpretation by themselves, bringing positive impacts including reduced reliance on trained analysts, lower costs, and improved efficiency.

However, other important aspects need to be reflected for a proper evaluation of the AI-assisted system. The first of them might be whether the deep learning models matched or outperformed human analysts or other conventional AI approaches in terms of the accuracy of wear assessment and prediction. As the development of gear wear monitoring and prediction is still in its early stage, human performance on such tasks has not been reported in the literature, and similar works are rather limited. In terms of wear assessment, the three wear severity CNNs developed in Chapter 6 achieved significantly higher classification accuracies ($\sim$ 95%) than the conventional classification approach using a HOG descriptor did in [114] ($\sim$ 80%) for wear severity assessment. The wear mechanism

CNN obtained an even higher classification accuracy of 98% plus. For wear debris images, the CNNs in this research achieved a comparable performance ($\sim$ 90%) to the CNNs developed in [129, 239] ($\sim$ 94%) for wear debris classification. In terms of wear prediction, due to the scarcity of studies on image-based wear prediction, performance data for comparison were not found. The wear prediction model developed in Chapter 7 achieved high scores (0.994 ± 0.002) on VGG cosine similarity that measures the realism and accuracy of predictions and low errors in pitted area ($0.652\% \pm 7.3\%$) and pitted region ($-2.56\% \pm 4.21\%$) prediction.

Then, the technical efforts spent on building the AI-assisted systems, from model selection to data preparation and to model training, need to be considered despite their desirable features of automated process and high accuracy. In this research, the selection of CNNs for image classification was inspired by their numerous successful applications related to image analysis (see section 2.3.3), while the selection of ConvLSTM for image prediction was inspired by its applications in video prediction. Normal 2D CNNs cannot model the time dimension of time series of gear tooth images, and therefore ConvLSTM, i.e., the combination of CNNs and LSTM networks, was employed. It is worth noting that ConvLSTM is only one of the several popular models for video prediction [243], which is a relatively new research topic, and a comprehensive comparison between them would be beyond the scope of this research.

Regarding data preparation, deep learning models (e.g., deep CNNs) generally have stronger learning capabilities than conventional machine learning models (e.g., SVMs) [63], but they usually require a much higher number of training samples to minimise overfitting [225]. Therefore, the preparation of training, validation and test data, especially the manual process of data labelling, could be one of the most time-consuming procedures. In this research, data (images) labelling for the deep CNNs in Chapter 6 took around two months of full-time manual work, and the author experienced the effect of mental fatigue on the efficiency and correctness of the labelling process. There are many open-source datasets for common applications of deep learning (e.g., classifying common objects such as trees and cars), but for its specific engineering applications such as in this research, the work-

load of dataset building should be well considered. A potential solution to this issue is to build large open-source datasets of gear wear images (either gear tooth images or wear debris images) that cover different wear mechanisms, severity, gear materials, etc., through research-industry collaborations. The open-source datasets could facilitate cross-checks of data labels, model benchmarking, and ultimately industrial applications of AI-assisted wear monitoring and prediction.

Lastly, training deep learning models with different architecture and hyper-parameters has mainly been a trial-and-error process since there have not been widely accepted theories to determine the parameters [176, 225]. For industrial maintenance, the combined costs of time, computing resources, and deep learning experts could become prohibitive to train such models from scratch and train many models for slightly different applications. To solve this issue, transfer learning [260] could be applied to utilise pre-trained models on large datasets, and it could effectively reduce training time and the need for a large number of new training samples.

8.4 Strengths and limitations of the developed techniques

This section discusses the strengths and weaknesses of the developed techniques in this research in terms of several common aspects to evaluate CM techniques as suggested in [9]. Firstly, the strengths of the developed techniques can be identified as follow.

- **The techniques can provide direct and comprehensive gear wear information.** As mentioned in Chapter 4, the combination of surface replication and image analysis can be used to extract a wide variety of direct gear wear information, such as gear tooth appearance and its evolution, macro- & micro-scale surface features, 3D tooth profiles, and pitted area. In comparison, the methods based on vibration, AE, and IRT in practice can only be used to detect severe gear wear (also called wear fault [261]) and other gear faults [5, 54, 55, 59, 261] via indirect wear information (e.g., vibration signals), despite their capabilities in detecting mild to moderate gear

wear (in terms of wear depth) and differentiating wear mechanisms being actively studied [50, 51]. OA and WDA are more direct methods, and they are widely used to determine gear wear rate, wear mechanism, and wear source. However, the actual gear tooth condition, which can be directly accessed by the developed techniques, cannot be reconstructed using discrete wear debris.

- **The techniques can monitor gear wear with high accuracy and high reliability.** As mentioned in Chapter 4, the selected replication material has desirable physical properties such as high resolution, low shrinkage and low thermal expansion, and its replication quality was tested regarding resolution, accuracy, and profile reproduction (see Appendix A). Therefore, replicas of worn gear teeth with a micro-level precision can be obtained through proper moulds making. Common microscopes would be sufficient to capture high-resolution images of the gear moulds, and the images can be analysed by the automated wear assessment system developed in Chapter 6 or trained analysts to ensure accurate and reliable gear wear monitoring. In comparison, the reliability of other CM techniques could be easily impacted by various factors (see sections 2.2.2 and 2.2.3), resulting in reduced accuracy and reliability for gear wear monitoring.

- **The techniques could be a cost-effective tool** for root cause analysis and gear maintenance. Maintenance cost is a major expenditure that many industries attempt to moderate such as wind farms [3], and therefore introducing new and costly maintenance techniques to such industries could be challenging. The developed techniques in this research, however, could be low-cost yet effective in comparison to the current practice of detailed wear analyses that require gearbox disassembly and even tooth cut. The potential costs of applying the techniques in industrial maintenance could be broken down into hardware cost (purchasing replication compounds, imaging facilities, etc.,), personnel cost (for moulds making, imaging, etc.,), and opportunity cost (for interrupting production). Firstly, in terms of the hardware cost, the cost of replication compounds used in this research was around AU$ 80 per 100 ml, which could be very likely lower for bulk purchases, and the cost of a common optical

microscope is around a few thousand Australian dollars. Secondly, automated gear wear assessment and prediction based on the deep learning models could save a major expense for trained analysts, and existing maintenance engineers with essential training can carry out the main tasks such as moulds making and imaging. Lastly, as discussed in section 8.1, regular inspections that are commonly implemented in industries could be well utilised to carry out moulds making which could be finished quickly, and therefore the interruptions to normal production could be minimised.

Nevertheless, the developed techniques have important limitations that need to be considered. Future work to address the issues can be found in the next chapter.

- **The techniques are intrusive and cannot provide real-time wear monitoring.** As mentioned in section 8.1, the moulds making process requires the gearbox to be stopped and its inspection cover to be removed, and hence it has obvious disadvantages compared to non-intrusive techniques such as vibration analysis and IRT. The lack of real-time monitoring capability means that the gear tooth condition could not be inspected between two moulds collections, although the condition can be predicted using history mould images. Therefore, ideally, the techniques should be used with real-time monitoring techniques for enhanced robustness.

- **The study of the relationship between worn gear surfaces and wear debris was based on only 2D images.** 3D images of wear debris had a limited sample size (less than 100 compared to over 10,000 for 2D images of wear debris), and therefore they were not used in this research. Comparing 3D features of worn surfaces and wear debris, such as volume and surface roughness, could provide additional insights into their relationship in gear wear evolution.

- **The automated monitoring and prediction models still need more training data to further improve their reliability.** Successful deep learning systems (e.g., autopilot systems, face recognition systems) all need to be trained on large scale training datasets to ensure that they can generalise well to new data, which

is directly related to their reliability. Similarly, the deep learning models developed in this research also need to be trained on a variety of gear moulds images covering different gear sizes, materials, operating conditions, etc. In this research, the images came from only two gear wear tests although both lubricated and dry tests were conducted. Therefore, the moulds making procedure needs to be applied to more gear tests to obtain more diverse mould images.

Chapter 9

Conclusions and Future Work

This research has developed advanced techniques to collect direct wear information for gear wear assessment, monitoring and prediction. Specifically, Chapter 4 laid the foundation by employing the combination of surface replication and image analysis to obtain gear mould images that contain direct and comprehensive gear wear information. Then, Chapter 5 linked the proposed method to WDA, a popular method for gear wear monitoring, by investigating the morphological relationship between worn gear teeth and wear debris in the same wear process. Then, gear mould images and wear debris images were used to train deep learning models (CNNs) to identify common gear wear mechanisms and severity levels, thus automating the process of gear wear assessment. Lastly, time series of gear mould images, or mould image sequences, were used to train a deep learning model (GAN) to learn gear wear dynamics and predict future appearances of gear teeth using history images.

Overall, the techniques provide a novel tool that allows direct and detailed examinations of worn gear teeth and enables automated wear assessment and prediction based on direct wear information (gear mould images) using deep learning approaches. Compared to other CM techniques for gear wear monitoring and prediction, the techniques have distinctive features of direct monitoring, high accuracy and reliability, and minimised reliance on trained analysts, and they are suitable to be used with real-time monitoring techniques

for enhanced robustness. Specifically, key findings related to the four objectives of this research are summarised in section 9.1, and future work are recommended in section 9.2.

9.1 Summary of key findings

In relation to Objective 1 of this research, i.e., **to develop a method to obtain direct and comprehensive gear wear information without gearbox disassembly**, the following key findings were obtained.

- Surface replication using the selected material attained (i) high replication resolution (capturing surface features spaced at ~ 1 µm), (ii) high-accuracy surface topography replication (accurately replicating surfaces with roughness values ranging from 0.4 µm to 12.5 µm), and (iii) high-accuracy profile replication (less than 1 µm RMS error when replicated profiles were aligned with actual profiles).
- From mould images, a wide variety of direct gear wear information were extracted via image analysis techniques, including (i) the overall appearance of a gear tooth and its evolution, (ii) macro- & micro-scale surface features, (iii) 3D tooth profile, (iv) surface roughness, and (v) wear quantification (e.g., wear depth, pitted area).

In relation to Objective 2, i.e., **to investigate the relationship between the features of worn gear surfaces and those of wear debris**, the following key findings were obtained.

- Qualitative comparisons between gear mould images and wear debris images obtained in the lubricated test showed that the variation of wear debris concentration/quantity followed a similar trend as the propagation of macropitting (the primary wear mechanism) on gear teeth.

- Quantitative comparisons showed that the evolution of wear debris concentration had a similar trend to that of the pitted area derivative in the test but with a notable delay of ~ 0.85 million cycles.
- The same delay was also observed in comparing the evolution of four non-dimensional features (aspect ratio, extent, solidity, and circularity) of wear debris and macropits.

In relation to Objective 3, i.e., **to develop an automated system for gear wear mechanism and severity assessment**, the following key findings were obtained.

- Deep CNNs were employed to automate gear wear assessment using mould images and wear debris images, and the wear assessment CNNs achieved high classification accuracies in tests. The wear mechanism CNN using mould images achieved a notably high accuracy of 98.22% in classifying images of normal condition, abrasive wear, macropitting, and scuffing, and the three wear severity CNNs scored 93.43%, 94.44%, and 97.61% in classifying images of mild, moderate, and severe wear, for abrasive wear, macropitting, and scuffing, respectively. The particle classification CNN using wear debris images achieved a classification accuracy of 88.97% in classifying images of rubbing, fatigue, laminar, and cutting particles, and the wear severity CNN scored 90.11% in classifying images of mild, moderate, and severe wear.
- Using gear mould images, a two-level inference system was built by connecting the wear mechanism CNN (the first level) and three severity CNNs (the second level). The system could automatically convert full-tooth mould images to wear mechanism maps and wear severity maps that display the distribution of wear mechanisms and their severity levels on gear teeth.

In relation to Objective 4, i.e., **to develop a gear wear prediction model using direct wear information**, the following key findings were obtained.

- A GAN with a ConvLSTM generator was developed to generate a future image of a gear tooth using five history images of the same tooth and their time information.

Image sequences with six different time intervals were used in model training to adapt to different gearbox inspection intervals in practice.

- Qualitatively, the predicted images in tests were reasonably realistic and accurate because they contain essential tribological features for wear assessment and plausible surface changes aligned with the trends of wear propagation.
- The predicted images also obtained high scores on quantitative metrics, including VGG cosine similarity (0.994±0.002) that measures the realism and accuracy of predictions and PAE (0.652%±7.3%) and PRE (-2.56%±4.21%) that measure the accuracy of pitting prediction.

9.2 Future work

Based on the limitations of this research as discussed in section 8.4, future work on the developed techniques are recommended as follow.

- The procedures of moulds making and imaging could be streamlined to minimise the interruption to machine operation and to further improve their efficiency. Although the moulds making process could utilise regular inspections and be quickly finished, as discussed in section 8.1, its interruption to machine operation could be costly and needs to be minimised as much as possible. Further developments, such as designing tools to simplify tooth cleaning and applying replication compounds, and standardising the mould imaging process, could be considered.
- Further studies of comparing 3D features of worn surfaces and those of wear debris in a gear wear process can be conducted, to deepen the understanding of the relationship between worn surfaces and wear debris. Several technical challenges could be encountered, such as the acquisition of full 3D shapes of wear particles and the heavy workload of acquiring sufficient 3D images of worn surfaces and wear particles. A potential solution to the former issue could be utilising microfluidic devices with

multiple imaging sensors to capture multiple 2D images of wear particles, and they can be used to reconstruct full 3D shapes of wear particles.

- Utilising the moulding-imaging techniques, further fundamental studies on the influence of gear materials (e.g., ceramics and polymers) and operational and environmental conditions on their wear performance can be conducted. The image-based evidence is critical for understanding their wear modes and wear processes. Also, crucial data will be available to develop AI-based wear assessment and prediction. Collaborations and an open platform could be the key to building large-scale gear wear datasets, which benefit not only the improvement of the deep learning models but also the broader field of gear wear research.

- Stochasticity could be introduced to the wear prediction model for more precise predictions of wear propagation. The process of gear wear propagation is difficult to predict in details, such as the generation location of macropits and their sizes, and therefore, the deterministic GAN model developed in Chapter 7 has limited capability to precisely predict such events. To solve this issue, the model needs to be further improved to incorporate stochasticity. When the prediction accuracy is well improved, a recursive prediction strategy in which the predicted image along with previous frames is used as the input of the generator will be attempted to make multi-step-ahead predictions.

- The integration of the techniques with other CM techniques such as vibration and WDA in monitoring and predicting gear wear should be further studied to reveal if the distinctive strengths of these techniques could be combined to provide a robust maintenance program that could effectively reduce maintenance costs and downtime. Possible approaches are digital twins [262] with multiple sensors and other data fusion techniques [263, 264].

References

[1] Global Industry Analysts Inc., "Industrial gearbox - global market trajectory & analytics," Global Industry Analysts Inc., Tech. Rep., 6 2021. [Online]. Available: https://www.strategyr.com/market-report-industrial-gearbox-forecasts-global-industry-analysts-inc.asp

[2] J. Ribrant, "Reliability performance and maintenance-a survey of failures in wind power systems," Ph.D. dissertation, KTH Royal Institute of Technology, 2006. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.561.2279&rep=rep1&type=pdf

[3] C. Dao, B. Kazemtabrizi, and C. Crabtree, "Wind turbine reliability data review and impacts on levelised cost of energy," *Wind Energy*, vol. 22, no. 12, pp. 1848–1871, 12 2019. [Online]. Available: https://publons.com/publon/10.1002/we.2404

[4] F. D. Harris and E. F. Kasper, "U.S. civil rotorcraft accidents, 1963 through 1997," in *American Helicopter Society Forum 55*. Montreal: NASA, jan 1998. [Online]. Available: https://ntrs.nasa.gov/citations/20020065533

[5] R. B. Randall, *Vibration-based condition monitoring: industrial, automotive and aerospace applications*, 1st ed. New Jersey: John Wiley & Sons, 2011.

[6] A. Kumar, C. P. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Latest developments in gear defect diagnosis and prognosis: A review," *Measurement*, vol. 158, 7 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224120302736

[7] P. J. Blau, *Tribosystem Analysis: A Practical Approach to the Diagnosis of Wear Problems*, 1st ed. Boca Raton: CRC Press, 2016.

[8] S. Feng, B. Fan, J. Mao, and Y. Xie, "Prediction on wear of a spur gearbox by on-line wear debris concentration monitoring," *Wear*, vol. 336-337, pp. 1–8, 8 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164815002057

[9] A. Davies, *Handbook of condition monitoring: techniques and methodology*, 1st ed. New York: Springer, 1998.

[10] R. S. Dwyer-Joyce, "The life cycle of a debris particle," *Tribology and Interface Engineering Series*, vol. 48, pp. 681–690, 1 2005.

[11] G. K. Nikas, "Particle entrainment in elastohydrodynamic point contacts and related risks of oil starvation and surface indentation," *Journal of Tribology*, vol. 124, no. 3, pp. 461–467, 7 2002. [Online]. Available: http://asmedigitalcollection.asme.org/tribology/article-pdf/124/3/461/5819677/461_1.pdf

[12] V. Strubel, N. Fillot, F. Ville, J. Cavoret, P. Vergne, A. Mondelin, and Y. Maheo, "Particle entrapment in hybrid lubricated point contacts," *Tribology Transactions*, vol. 59, no. 4, pp. 768–779, 7 2016. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10402004.2015.1106631

[13] H. Wu, R. Li, N. M. Kwok, Y. Peng, T. Wu, and Z. Peng, "Restoration of low-informative image for robust debris shape measurement in on-line wear debris monitoring," *Mechanical Systems and Signal Processing*, vol. 114, pp. 539–555, 1 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327018302863?via%3Dihub

[14] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 314–334, 1 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327013002860

[15] F. Zhao, Z. Tian, and Y. Zeng, "Overview on gear health prognostics," in *Probabilistic Prognostics and Health Management of Energy Systems*. New York: Springer, 2017, pp. 49–65. [Online]. Available: http://link.springer.com/10.1007/978-3-319-55852-3_4

[16] C. Hu, B. D. Youn, P. Wang, and J. Taek Yoon, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," *Reliability Engineering and System Safety*, vol. 103, pp. 120–135, 7 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832012000427

[17] X. Xu, Y. Lei, and Z. Li, "An incorrect data detection method for big data cleaning of machinery condition monitoring," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2326–2336, 3 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8667006

[18] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 10 2006. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0888327005001512?via%3Dihub

[19] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 135–149, 8 2012. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0360835212000484

[20] D. W. Dudley, "Gear wear," in *Wear Control Handbook*, M. B. Peterson and W. O. Winer, Eds. New York: ASME, 1980, pp. 755–830.

[21] AGMA, "ANSI/AGMA 1010-F14 - Appearance of gear teeth - terminology of wear and failure," 2014.

[22] J. T. Burwell, "Survey of possible wear mechanisms," *Wear*, vol. 1, no. 2, pp. 119–141, 10 1957. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0043164857900054

[23] J. F. Archard, "Wear theory and mechanisms," in *Wear control handbook*, M. B. Peterson and W. O. Winer, Eds. New York: ASME, New York, 1980, ch. II, pp. 35–80.

[24] K. Kato and K. Adachi, "Wear mechanisms," in *Modern Tribology Handbook*, B. Bhushan, Ed. Boca Raton: CRC Press, 2000, ch. 7.

[25] B. J. Roylance, "Wear debris and associated wear phenomena-fundamental research and practice," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 214, no. 1, pp. 79–105, 1 2000. [Online]. Available: http://journals.sagepub.com/doi/10.1243/1350650001543025

[26] Z. Peng and N. Kessissoglou, "An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis," *Wear*, vol. 255, no. 7-12, pp. 1221–1232, 8 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004316480300098X

[27] G. W. Stachowiak and A. W. Batchelor, *Engineering Tribology*, 4th ed. New York: Elsevier, 2014.

[28] H. S. Cheng, "Gears," in *Modern Tribology Handbook*, B. Bhushan, Ed. Boca Raton: CRC Press, 2000, ch. 29.

[29] S. P. Radzevich, *Dudley's handbook of practical gear design and manufacture*, 2nd ed. Boca Raton: CRC Press, 2012.

[30] B. Bhushan, *Modern Tribology Handbook*, B. Bhushan, Ed. Boca Raton: CRC Press, 2000, vol. 1.

[31] M. M. Khruschov, "Principles of abrasive wear," *Wear*, vol. 28, no. 1, pp. 69–88, 4 1974. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0043164874901021

[32] C. N. Rowe, "Lubricated wear," in *Wear Control Handbook*, M. B. Peterson and W. O. Winer, Eds. New York: ASME, 1980, ch. 2, pp. 143–160.

[33] F. Sadeghi, B. Jalalahmadi, T. S. Slack, N. Raje, and N. K. Arakere, "A review of rolling contact fatigue," *Journal of Tribology*, vol. 131, no. 4, p. 041403, 10 2009. [Online]. Available: http://tribology.asmedigitalcollection.asme.org/article.aspx?articleid=1468339

[34] K. J. Miller, "Materials science perspective of metal fatigue resistance," *Materials Science and Technology*, vol. 9, no. 6, pp. 453–462, 1993. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1179/mst.1993.9.6.453

[35] Y. Murakami and M. Endo, "Effects of defects, inclusions and inhomogeneities on fatigue strength," *International Journal of Fatigue*, vol. 16, no. 3, pp. 163–182, 4 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0142112394900019

[36] L. M. Keer and M. D. Bryant, "A pitting model for rolling contact fatigue," *Journal of lubrication technology*, vol. 105, no. 2, pp. 198–205, 1983.

[37] M. T. Hanson and L. M. Keer, "An analytical life prediction model for the crack propagation occurring in contact fatigue failure," *Tribology Transactions*, vol. 35, no. 3, pp. 451–461, 1992.

[38] G. Fajdiga, S. Glodež, and J. Kramar, "Pitting formation due to surface and subsurface initiated fatigue crack growth in contacting mechanical elements," *Wear*, vol. 262, no. 9-10, pp. 1217–1224, 4 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164806004546

[39] K. L. Johnson, *Contact mechanics*. Cambridge: Cambridge University Press, 1985.

[40] S. Li and A. Kahraman, "A micro-pitting model for spur gear contacts," *International Journal of Fatigue*, vol. 59, pp. 224–233, 2 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142112313002302

[41] S. Haizuka, C. Naruse, and J. Tamenaga, "Study on tooth surface strength of spur gears — part i: experimental procedure and determination of tooth surface

strength," *Tribology Transactions*, vol. 42, no. 1, pp. 76–83, 1999. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10402009908982202

[42] H. P. Evans, R. W. Snidle, K. J. Sharif, B. A. Shaw, and J. Zhang, "Analysis of micro-elastohydrodynamic lubrication and prediction of surface fatigue damage in micropitting tests on helical gears," *Journal of Tribology*, vol. 135, no. 1, 11 2013. [Online]. Available: http://asmedigitalcollection.asme.org/tribology/article-pdf/135/1/011501/6282407/trib-135-1-011501.pdf

[43] M. Weibring, L. Gondecki, and P. Tenberge, "Simulation of fatigue failure on tooth flanks in consideration of pitting initiation and growth," *Tribology International*, vol. 131, pp. 299–307, 3 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X18305103?via%3Dihub

[44] K. L. Johnson, "The strength of surfaces in rolling contact," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 203, no. 3, pp. 151–163, 5 1989. [Online]. Available: http://journals.sagepub.com/doi/10.1243/PIME_PROC_1989_203_100_02

[45] S. Raadnui, "Spur gear wear analysis as applied for tribological based predictive maintenance diagnostics," *Wear*, vol. 426-427, pp. 1748–1760, 4 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164818317113

[46] D. P. Townsend and J. Shimski, "Evaluation of the EHL film thickness and extreme pressure additives on gear surface fatigue life," NASA, Tech. Rep., 7 1994. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=19940032811

[47] X. Zhu, C. Zhong, and J. Zhe, "Lubricating oil conditioning sensors for online machine health monitoring – A review," *Tribology International*, vol. 109, pp. 473–484, 5 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X17300166

[48] X. S. Si, W. Wang, C. H. Hu, and D. H. Zhou, "Remaining useful life estimation - A review on the statistical data driven approaches," *European Journal of*

Operational Research, vol. 213, no. 1, pp. 1–14, 8 2011. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0377221710007903

[49] J. Sun, L. Wang, J. Li, F. Li, J. Li, and H. Lu, "Online oil debris monitoring of rotating machinery: A detailed review of more than three decades," *Mechanical Systems and Signal Processing*, vol. 149, p. 107341, 2 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327020307275

[50] P. Feng, P. Borghesani, H. Chang, W. A. Smith, R. B. Randall, and Z. Peng, "Monitoring gear surface degradation using cyclostationarity of acoustic emission," *Mechanical Systems and Signal Processing*, vol. 131, pp. 199–221, 9 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327019303644

[51] K. Feng, W. A. Smith, P. Borghesani, R. B. Randall, and Z. Peng, "Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution," *Mechanical Systems and Signal Processing*, vol. 150, p. 107258, 3 2021. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0888327020306440

[52] S. Bagavathiappan, B. B. Lahiri, T. Saravanan, J. Philip, and T. Jayakumar, "Infrared thermography for condition monitoring – A review," *Infrared Physics & Technology*, vol. 60, pp. 35–55, 9 2013. [Online]. Available: https://doi.org/10.1016/j.infrared.2013.03.006

[53] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Molleda, and F. G. Bulnes, "Infrared thermography for temperature measurement and non-destructive testing," *Sensors*, vol. 14, no. 7, pp. 12 305–12 348, 7 2014. [Online]. Available: https://www.mdpi.com/1424-8220/14/7/12305/htmhttps://www.mdpi.com/1424-8220/14/7/12305

[54] E. Resendiz-Ochoa, J. J. Saucedo-Dorantes, J. P. Benitez-Rangel, R. A. Osornio-Rios, and L. A. Morales-Hernandez, "Novel methodology for condition monitoring of gear wear using supervised learning and infrared thermography,"

Applied Sciences, vol. 10, no. 2, p. 506, 1 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/2/506/htmhttps://www.mdpi.com/2076-3417/10/2/506

[55] C. Hu, W. A. Smith, R. B. Randall, and Z. Peng, "Development of a gear vibration indicator and its application in gear wear monitoring," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 319–336, 8 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327016000200

[56] D. Mba and R. B. R. R. B. Rao, "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines; bearings, pumps, gearboxes, engines and rotating structures," *The Shock and Vibration Digest*, vol. 38, no. 1, pp. 3–16, 1 2006. [Online]. Available: http://svd.sagepub.com/cgi/doi/10.1177/0583102405059054https://dspace.lib.cranfield.ac.uk/handle/1826/1029

[57] C. K. Tan, P. Irving, and D. Mba, "A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 208–233, 1 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327005001500

[58] R. B. Sharma, A. Parey, and N. Tandon, "Modelling of acoustic emission generated in involute spur gear pair," *Journal of Sound and Vibration*, vol. 393, pp. 353–373, 4 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022460X1730010X

[59] T. H. Loutas, D. Roulias, E. Pauly, and V. Kostopoulos, "The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1339–1352, 5 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327010004103

[60] R. B. Sharma and A. Parey, "Modelling of acoustic emission generated due to pitting on spur gear," *Engineering Failure Analysis*, vol. 86, pp. 1–20, 4 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350630717305009

[61] S. Dutta, S. K. Pal, S. Mukhopadhyay, and R. Sen, "Application of digital image processing in tool condition monitoring: A review," *CIRP Journal of Manufacturing Science and Technology*, vol. 6, no. 3, pp. 212–232, 1 2013. [Online]. Available: https://doi.org/10.1016/j.cirpj.2013.02.005

[62] W. Wang, F. Guan, S. Ma, and J. Li, "Measurement system of gear parameters based on machine vision," *Measurement and Control*, vol. 48, no. 8, pp. 242–248, 10 2015. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/0020294015595997

[63] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 5 2015. [Online]. Available: http://www.nature.com/articles/nature14539

[64] A. Flodin, "Wear investigation of spur gear teeth," *TriboTest*, vol. 7, no. 1, pp. 45–60, 9 2000. [Online]. Available: http://doi.wiley.com/10.1002/tt.3020070106

[65] H. N. Hansen, R. J. Hocken, and G. Tosello, "Replication of micro and nano surface geometries," *CIRP Annals - Manufacturing Technology*, vol. 60, no. 2, pp. 695–714, 1 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0007850611002101

[66] S. Andersson, "Initial wear of gears," *Tribology International*, vol. 10, no. 4, pp. 206–210, 8 1977. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0301679X77900214

[67] S. Andersson, "Plastic replicas for optical and scanning electron microscopy," *Wear*, vol. 29, no. 2, pp. 271–274, 8 1974. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0043164874900775

[68] A. Flodin, "Wear of spur and helical gears," Ph.D. dissertation, Royal Institute of Technology Sweden, 2000.

[69] V. Moorthy and B. A. Shaw, "Contact fatigue performance of helical gears with surface coatings," *Wear*, vol. 276-277, pp. 130–140, 2 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164811006521

[70] V. Moorthy and B. A. Shaw, "Effect of as-ground surface and the BALINIT® C and Nb-S coatings on contact fatigue damage in gears," *Tribology International*, vol. 51, pp. 61–70, 7 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X12000795

[71] E. C. Ames, "Repair of high value/high demand spiral Bevel gears by superfinishing," in *The American Helicopter Society 67th Annual Forum.* Virginia Beach, VA: AHS International, 2011, pp. 2503–2514.

[72] ASTM D7670-10(2017), "Standard practice for processing in-service fluid samples for particulate contamination analysis using membrane filters," West Conshohocken, PA, 2017. [Online]. Available: www.astm.org

[73] ASTM D7684-11(2020), "Standard guide for microscopic characterization of particles from in-service lubricants," West Conshohocken, PA, 2020.

[74] ASTM D7690-11(2017), "Standard practice for microscopic characterization of particles from in-service lubricants by analytical ferrography," West Conshohocken, PA, 2017.

[75] B. Roylance, "Ferrography—then and now," *Tribology International*, vol. 38, no. 10, pp. 857–862, 10 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X05000897

[76] Spectro Scientific, "MicroLab 30 - On-site Oil Analyzer." [Online]. Available: https://www.spectrosci.com/product/microlab-30/

[77] W. Cao, G. Dong, Y. B. Xie, and Z. Peng, "Prediction of wear trend of engines via on-line wear debris monitoring," *Tribology International*, vol. 120, pp. 510–519, 4 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0301679X1830015X?via%3Dihub

[78] Y. Peng, T. Wu, S. Wang, and Z. Peng, "Wear state identification using dynamic features of wear debris for on-line purpose," *Wear*, vol. 376-377, pp. 1885–1891,

4 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164817300674

[79] X. Zhu, L. Du, and J. Zhe, "A 3×3 wear debris sensor array for real time lubricant oil conditioning monitoring using synchronized sampling," *Mechanical Systems and Signal Processing*, vol. 83, pp. 296–304, 1 2017. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0888327016301960

[80] W. Hong, W. Cai, S. Wang, and M. M. Tomovic, "Mechanical wear debris feature, detection, and diagnosis: A review," *Chinese Journal of Aeronautics*, vol. 31, no. 5, pp. 867–882, 5 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1000936117302637

[81] Gastops, "MetalSCAN." [Online]. Available: https://www.gastops.com/products/metalscan/

[82] Parker, "MWDS - Metal Wear Debris Sensor." [Online]. Available: https://ph.parker.com/au/en/mwds-metal-wear-debris-sensor

[83] PAMAS, "PAMAS OLS4031." [Online]. Available: https://www.pamas.de/particle-counters/products-by-name/pamas-ols4031.html

[84] L. Du and J. Zhe, "A high throughput inductive pulse sensor for online oil debris monitoring," *Tribology International*, vol. 44, no. 2, pp. 175–179, 2 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X10002653

[85] X. Zhu, L. Du, and J. Zhe, "An integrated lubricant oil conditioning sensor using signal multiplexing," *Journal of Micromechanics and Microengineering*, vol. 25, no. 1, p. 015006, 1 2015. [Online]. Available: https://iopscience.iop.org/article/10.1088/0960-1317/25/1/015006/meta

[86] L. Du, J. Zhe, J. Carletta, R. Veillette, and F. Choy, "Real-time monitoring of wear debris in lubrication oil using a microfluidic inductive Coulter counting device," *Microfluidics and Nanofluidics*, vol. 9, no. 6, pp. 1241–1245, 12 2010. [Online]. Available: http://link.springer.com/10.1007/s10404-010-0627-y

[87] J. Zhe, X. Zhu, and C. Zhong, "A high sensitivity wear debris sensor using ferrite cores for online oil condition monitoring," *Measurement Science and Technology*, vol. 28, no. 7, p. 075102, 7 2017. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6501/aa6adb

[88] S. Morris, R. J. K. Wood, T. J. Harvey, and H. E. G. Powrie, "Use of electrostatic charge monitoring for early detection of adhesive wear in oil lubricated contacts," *Journal of Tribology*, vol. 124, no. 2, pp. 288–296, 4 2002. [Online]. Available: https://doi.org/10.1115/1.1398293

[89] T. J. Harvey, S. Morris, L. Wang, R. J. K. Wood, and H. E. G. Powrie, "Real-time monitoring of wear debris using electrostatic sensing techniques," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 221, no. 1, pp. 27–40, 2 2007. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1243/13506501JET196

[90] Z. Wen, X. Yin, and Z. Jiang, "Applications of electrostatic sensor for wear debris detecting in the lubricating oil," *Journal of The Institution of Engineers (India): Series C*, vol. 94, no. 3, pp. 281–286, 10 2013. [Online]. Available: https://link.springer.com/article/10.1007/s40032-013-0072-2

[91] H. Mao, H. Zuo, and H. Wang, "Electrostatic sensor application for on-line monitoring of wind turbine gearboxes," *Sensors*, vol. 18, no. 10, p. 3574, 10 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/10/3574/htm

[92] L. Du and J. Zhe, "An integrated ultrasonic–inductive pulse sensor for wear debris detection," *Smart Materials and Structures*, vol. 22, no. 2, p. 025003, 12 2012. [Online]. Available: https://iopscience.iop.org/article/10.1088/0964-1726/22/2/025003

[93] ARGO-HYTOS, "OPCount - Portable particle counter." [Online]. Available: https://www.argo-hytos.com/products/sensors-measurement/portable-particle-counters/opcount.html

[94] T. Wu, Y. Peng, H. Wu, X. Zhang, and J. Wang, "Full-life dynamic identification of wear state based on on-line wear debris image features," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 404–414, 1 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327013004433

[95] T. H. Wu, J. H. Mao, J. T. Wang, J. Y. Wu, and Y. B. Xie, "A new on-line visual ferrograph," *Tribology Transactions*, vol. 52, no. 5, pp. 623–631, 9 2009. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10402000902825762

[96] Y. Peng, T. Wu, S. Wang, and Z. Peng, "Oxidation wear monitoring based on the color extraction of on-line wear debris," *Wear*, vol. 332-333, pp. 1151–1157, 5 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004316481400427X

[97] T. Wu, H. Wu, Y. Du, N. Kwok, and Z. Peng, "Imaged wear debris separation for on-line monitoring using gray level and integrated morphological features," *Wear*, vol. 316, no. 1-2, pp. 19–29, 8 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0043164814001380

[98] B. Li, Y. Xi, S. Feng, J. Mao, and Y.-B. Xie, "A direct reflection OLVF debris detector based on dark-field imaging," *Measurement Science and Technology*, vol. 29, no. 6, p. 065104, 4 2018. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6501/aab9fc

[99] H. Wu, T. Wu, Y. Peng, and Z. Peng, "Watershed-based morphological separation of wear debris chains for on-line ferrograph analysis," *Tribology Letters*, vol. 53, no. 2, pp. 411–420, 2014. [Online]. Available: https://link.springer.com/article/10.1007/s11249-013-0280-1

[100] Y. Peng, T. Wu, S. Wang, Y. Du, N. Kwok, and Z. Peng, "A microfluidic device for three-dimensional wear debris imaging in online condition monitoring," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 231, no. 8, pp. 965–974, 8 2017. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1350650116684707

[101] T. Wu, H. Wu, Y. Du, and Z. Peng, "Progress and trend of sensor technology for on-line oil monitoring," *Science China Technological Sciences*, vol. 56, no. 12, pp. 2914–2926, 12 2013. [Online]. Available: http://link.springer.com/10.1007/s11431-013-5400-5

[102] D. P. Anderson, "Wear particle atlas (revised)," Foxboro Analytical, Tech. Rep., 1982.

[103] R. S. Dwyer-Joyce and J. Heymer, "The entrainment of solid particles into rolling elastohydrodynamic contacts," *Tribology Series*, vol. 31, pp. 135–140, 1 1996. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167892208707776

[104] G. K. Nikas, "A state-of-the-art review on the effects of particulate contamination and related topics in machine-element contacts," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 224, no. 5, pp. 453–479, 3 2010. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1243/13506501JET752

[105] A. A. Reda, R. Bowen, and V. C. Westcott, "Characteristics of particles generated at the interface between sliding steel surfaces," *Wear*, vol. 34, no. 3, pp. 261–273, 10 1975. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0043164875900952?via%3Dihub

[106] A. Umeda, J. Sugimura, and Y. Yamamoto, "Characterization of wear particles and their relations with sliding conditions," *Wear*, vol. 216, no. 2, pp. 220–228, 4 1998.

[107] C. Q. Yuan, Z. Peng, X. C. Zhou, and X. P. Yan, "The surface roughness evolutions of wear particles and wear components under lubricated rolling wear condition," *Wear*, vol. 259, no. 1-6, pp. 512–518, 7 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0043164804004223

[108] C. Q. Yuan, Z. Peng, X. P. Yan, and X. C. Zhou, "Surface roughness evolutions in sliding wear process," *Wear*, vol. 265, no. 3-4, pp. 341–348, 7

2008. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S004316480700693X

[109] Z. Peng and T. B. Kirk, "Automatic wear-particle classification using neural networks," *Tribology Letters*, vol. 5, no. 4, p. 249–257, 1998. [Online]. Available: http://link.springer.com/10.1023/A:1019126732337

[110] G. W. Stachowiak and P. Podsiadlo, "Characterization and classification of wear particles and surfaces," *Wear*, vol. 249, no. 3-4, pp. 194–200, 5 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164801005622

[111] S. Raadnui, "Wear particle analysis—utilization of quantitative computer image analysis: A review," *Tribology International*, vol. 38, no. 10, pp. 871–878, 10 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X05000915

[112] G. P. Stachowiak, G. W. Stachowiak, and P. Podsiadlo, "Automated classification of wear particles based on their surface texture and shape features," *Tribology International*, vol. 41, no. 1, pp. 34–43, 1 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X07000849

[113] H. Wu, N. M. Kwok, S. Liu, T. Wu, and Z. Peng, "A prototype of on-line extraction and three-dimensional characterisation of wear particle features from video sequence," *Wear*, vol. 368-369, pp. 314–325, 12 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164816303854

[114] C. Gonzalez-Arias, C. C. Viafara, J. J. Coronado, and F. Martinez, "Automatic classification of severe and mild wear in worn surface images using histograms of oriented gradients as descriptor," *Wear*, vol. 426-427, pp. 1702–1711, 4 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164818315023

[115] C. Locateli, R. Martins, and J. Seabra, "Evolution of tooth flank roughness during gear micropitting tests," *Industrial Lubrication and Tribology*, vol. 63, no. 1, pp. 211–212, 2 2011. [Online]. Available: https://www.emeraldinsight.com/doi/10.1108/00368791111101821

[116] J. Wei, A. Zhang, and P. Gao, "A study of spur gear pitting under EHL conditions: Theoretical analysis and experiments," *Tribology International*, vol. 94, pp. 146–154, 2 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X15003795

[117] B. J. Roylance and S. Raadnui, "The morphological attributes identifying wear mechanisms of wear particles their role in," *Wear*, vol. 175, no. 1-2, pp. 115–121, 6 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0043164894901740

[118] Y. Tian, J. Wang, Z. Peng, and X. Jiang, "A new approach to numerical characterisation of wear particle surfaces in three-dimensions for wear study," *Wear*, vol. 282-283, pp. 59–68, 4 2012. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164812000336

[119] M. Guagliano, E. Riva, and M. Guidetti, "Contact fatigue failure analysis of shot-peened gears," *Engineering Failure Analysis*, vol. 9, no. 2, pp. 147–158, 4 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350630701000024

[120] D. J. Stokes, *Principles and Practice of Variable Pressure/Environmental Scanning Electron Microscopy (VP-ESEM)*. Chichester: John Wiley & Sons, 11 2008.

[121] H. Chang, P. Borghesani, W. A. Smith, and Z. Peng, "Application of surface replication combined with image analysis to investigate wear evolution on gear teeth - a case study," *Wear*, vol. 430-431, pp. 355–368, 7 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164819304442

[122] Z. Peng and T. B. Kirk, "Computer image analysis of wear particles in three-dimensions for machine condition monitoring," *Wear*, vol. 223, no. 1-2, pp. 157–166, 12 1998. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0043164898002804

[123] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, 1st ed. Upper Saddle River: Pearson/Prentice Hall, 2003.

[124] J. Wang, L. Zhang, F. Lu, and X. Wang, "The segmentation of wear particles in ferrograph images based on an improved ant colony algorithm," *Wear*, vol. 311, no. 1-2, pp. 123–129, 3 2014. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164814000131

[125] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. [Online]. Available: https://ieeexplore.ieee.org/document/4310076

[126] Z. Peng and T. Kirk, "Wear particle classification in a fuzzy grey system," *Wear*, vol. 225-229, pp. 1238–1247, 4 1999. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164898004001

[127] M. Kumar, P. Shankar Mukherjee, and N. Mohan Misra, "Advancement and current status of wear debris analysis for machine condition monitoring: a review," *Industrial Lubrication and Tribology*, vol. 65, no. 1, pp. 3–11, 2 2013. [Online]. Available: http://www.emeraldinsight.com/doi/10.1108/00368791311292756

[128] S. Wang, T. H. Wu, T. Shao, and Z. X. Peng, "Integrated model of BP neural network and CNN algorithm for automatic wear debris classification," *Wear*, vol. 426-427, pp. 1761–1770, 4 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164818317083

[129] P. Peng and J. Wang, "FECNN: A promising model for wear particle recognition," *Wear*, vol. 432-433, 8 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164819304107

[130] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 12 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7803544

[131] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in Neural Information Processing Systems*,

vol. 26, 2013. [Online]. Available: https://proceedings.neurips.cc/paper/2013/file/f7cade80b7cc92b991cf4d2806d6bd78-Paper.pdf

[132] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ

[133] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2016, pp. 770–778. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[134] H. Wu, N. M. Kwok, S. Liu, R. Li, T. Wu, and Z. Peng, "Restoration of defocused ferrograph images using a large kernel convolutional neural network," *Wear*, vol. 426-427, pp. 1740–1747, 4 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164818317125

[135] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Mathematical Problems in Engineering*, vol. 2015, pp. 1–17, 5 2015. [Online]. Available: http://www.hindawi.com/journals/mpe/2015/793161/

[136] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery," *Advances in Mechanical Engineering*, vol. 8, no. 8, pp. 1–21, 8 2016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1687814016664660

[137] P. Kundu, A. K. Darpe, and M. S. Kulkarni, "A review on diagnostic and prognostic approaches for gears," *Structural Health Monitoring*, pp. 1–41, 12 2020. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/1475921720972926

[138] O. Fink, Q. Wang, M. Svensén, P. Dersin, W. J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health

management applications," *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103678, 6 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197620301184

[139] J. F. Archard, "Contact and rubbing of flat surfaces," *Journal of Applied Physics*, vol. 24, no. 8, pp. 981–988, 8 1953. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.1721448

[140] D. Park and A. Kahraman, "A surface wear model for hypoid gear pairs," *Wear*, vol. 267, no. 9-10, pp. 1595–1604, 9 2009. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164809004190

[141] M. Watson, C. Byington, D. Edwards, and S. Amin, "Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems," *Tribology Transactions*, vol. 48, no. 2, pp. 208–217, 2005.

[142] S. Wu and H. S. Cheng, "A sliding wear model for partial-EHL contacts," *Journal of Tribology*, vol. 113, no. 1, pp. 134–141, 1 1991. [Online]. Available: https://asmedigitalcollection.asme.org/tribology/article-abstract/113/1/134/437666/A-Sliding-Wear-Model-for-Partial-EHL-Contacts

[143] S. Wu and H. S. Cheng, "Sliding wear calculation in spur gears," *Journal of tribology*, vol. 115, no. July, pp. 493–500, 7 1993. [Online]. Available: http://tribology.asmedigitalcollection.asme.org/article.aspx?articleid=1460336

[144] A. Flodin and S. Andersson, "Simulation of mild wear in spur gears," *Wear*, vol. 207, no. 1-2, pp. 16–23, 6 1997. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164896074674

[145] A. Flodin and S. Andersson, "Simulation of mild wear in helical gears," *Wear*, vol. 241, no. 2, pp. 123–128, 7 2000. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164800003847

[146] P. Bajpai, A. Kahraman, and N. E. Anderson, "A surface wear prediction methodology for parallel-axis gear pairs," *Journal of tri-*

bology, vol. 126, no. 3, pp. 597–605, 2004. [Online]. Available: https://asmedigitalcollection.asme.org/tribology/article-abstract/126/3/597/464309/A-Surface-Wear-Prediction-Methodology-for-Parallel

[147] T. L. Krantz and A. Kahraman, "An experimental investigation of the influence of the lubricant viscosity and additives on gear wear," *Tribology Transactions*, vol. 47, no. 1, pp. 138–148, 1 2004. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/05698190490278949

[148] A. Kahraman, P. Bajpai, and N. E. Anderson, "Influence of tooth profile deviations on helical gear wear," *Journal of Mechanical Design*, vol. 127, no. 4, p. 656, 7 2005. [Online]. Available: https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/127/4/656/729169/Influence-of-Tooth-Profile-Deviations-on-Helical

[149] J. A. Brandão, J. H. Seabra, and J. Castro, "Surface initiated tooth flank damage. Part I: Numerical model," *Wear*, vol. 268, no. 1, pp. 1–12, 1 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164809004311

[150] J. A. Brandão, J. H. Seabra, and J. Castro, "Surface initiated tooth flank damage. Part II: Prediction of micropitting initiation and mass loss," *Wear*, vol. 268, no. 1, pp. 13–22, 1 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164809004748

[151] S. Li, A. Kahraman, and M. Klein, "A fatigue model for spur gear contacts operating under mixed elastohydrodynamic lubrication conditions," *Journal of Mechanical Design*, vol. 134, no. 4, p. 041007, 4 2012. [Online]. Available: http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?articleid=1450793

[152] G. E. Morales-Espejel, P. Rycerz, and A. Kadiric, "Prediction of micropitting damage in gear teeth contacts considering the concurrent effects of surface fatigue and mild wear," *Wear*, vol. 398-399, pp. 99–115, 3 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164817309535

[153] M. F. AL-Mayali, S. Hutt, K. J. Sharif, A. Clarke, and H. P. Evans, "Experimental and numerical study of micropitting initiation in real rough surfaces in a micro-elastohydrodynamic lubrication regime," *Tribology Letters*, vol. 66, no. 4, pp. 1–14, 11 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11249-018-1110-2

[154] S. Glodež, H. Winter, and H. P. Stüwe, "A fracture mechanics model for the wear of gear flanks by pitting," *Wear*, vol. 208, no. 1-2, pp. 177–183, 7 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164897000082

[155] J. W. Blake and H. S. Cheng, "A surface pitting life model for spur gears: part i - life prediction," *Journal of Tribology*, vol. 113, no. 4, p. 719, 10 1991. [Online]. Available: http://tribology.asmedigitalcollection.asme.org/article.aspx?articleid=1459989

[156] K. Aslantaş and S. Taşgetiren, "A study of spur gear pitting formation and life prediction," *Wear*, vol. 257, no. 11, pp. 1167–1175, 12 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164804001802

[157] J. Flašker, G. Fajdiga, S. Glodež, and T. K. Hellen, "Numerical simulation of surface pitting due to contact loading," *International Journal of Fatigue*, vol. 23, no. 7, pp. 599–605, 8 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142112301000202

[158] G. Fajdiga and M. Sraml, "Fatigue crack initiation and propagation under cyclic contact loading," *Engineering Fracture Mechanics*, vol. 76, no. 9, pp. 1320–1335, 6 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013794409000484

[159] D. Wang, K. L. Tsui, and Q. Miao, "Prognostics and health management: a review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 11 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8115325

[160] D. He, E. Bechhoefer, P. Dempsey, and J. Ma, "An integrated approach for gear health prognostics," in *AHS International 68th Annual Forum*. Fort Worth,TX: NASA, 2012. [Online]. Available: https://ntrs.nasa.gov/citations/20120012591

[161] S. S. H. Zaidi, S. Aviyente, M. Salman, K. K. Shin, and E. G. Strangas, "Prognosis of gear failures in dc starter motors using hidden Markov models," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1695–1706, 5 2011. [Online]. Available: https://ieeexplore.ieee.org/document/5484485

[162] J. Kocijan and V. Tanko, "Prognosis of gear health using Gaussian process model," in *EUROCON 2011 - International Conference on Computer as a Tool - Joint with Conftele 2011*. Lisbon, Portugal: IEEE, 4 2011. [Online]. Available: https://ieeexplore.ieee.org/document/5929179

[163] J. Zhu, J. M. Yoon, D. He, and E. Bechhoefer, "Online particle-contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines," *Wind Energy*, vol. 18, no. 6, pp. 1131–1149, 6 2015. [Online]. Available: http://doi.wiley.com/10.1002/we.1746

[164] W. Hong, S. Wang, M. M. Tomovic, H. Liu, J. Shi, and X. Wang, "A novel indicator for mechanical failure and life prediction based on debris monitoring," *IEEE Transactions on Reliability*, vol. 66, no. 1, pp. 161–169, 3 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7792701

[165] A.-L. Boulesteix and M. Schmid, "Machine learning versus statistical modeling," *Biometrical Journal*, vol. 56, no. 4, pp. 588–593, 7 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/bimj.201300226

[166] N. K. Chauhan and K. Singh, "A review on conventional machine learning vs deep learning," in *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, 2018, pp. 347–352. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8675097

[167] W. Q. Wang, M. F. Golnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing*, vol. 18, no. 4, pp. 813–831, 7 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327003000797

[168] P. Kundu, A. K. Darpe, and M. S. Kulkarni, "An ensemble decision tree methodology for remaining useful life prediction of spur gears under natural pitting progression," *Structural Health Monitoring*, vol. 19, no. 3, pp. 854–872, 8 2019. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/1475921719865718

[169] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Systems with Applications*, vol. 83, pp. 187–205, 10 2017. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0957417417302750

[170] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic flow prediction with big data: A deep learning approachpproach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 4 2015. [Online]. Available: https://ieeexplore.ieee.org/document/6894591

[171] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo, and H. Kong Observatory, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*. Neural information processing systems foundation, 6 2015, pp. 802–810. [Online]. Available: http://arxiv.org/abs/1506.04214

[172] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 11 2015. [Online]. Available: http://arxiv.org/abs/1511.05440

[173] A. X. Lee, R. Zhang, F. Ebert, P. Abbeel, C. Finn, and S. Levine, "Stochastic adversarial video prediction," *arXiv preprint*, 4 2018. [Online]. Available: http://arxiv.org/abs/1804.01523

[174] S. Xiang, Y. Qin, C. Zhu, Y. Wang, and H. Chen, "Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction," *Engineering Applications of Artificial Intelligence*, vol. 91, p.

103587, 5 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0952197620300634

[175] H. Yan, Y. Qin, S. Xiang, Y. Wang, and H. Chen, "Long-term gear life prediction based on ordered neurons LSTM neural networks," *Measurement*, vol. 165, p. 108205, 12 2020.

[176] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016. [Online]. Available: http://www.deeplearningbook.org

[177] C.-C. Lo, C.-H. Lee, and W.-C. Huang, "Prognosis of bearing and gear wears using convolutional neural network with hybrid loss function," *Sensors*, vol. 20, no. 12, p. 3539, 6 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/12/3539

[178] D. Nélias, M. L. Dumont, F. Champiot, A. Vincent, D. Girodin, R. Fougeres, and L. Flamand, "Role of inclusions, surface roughness and operating conditions on rolling contact fatigue," *Journal of tribology*, vol. 121, no. 2, pp. 240–251, 1999. [Online]. Available: https://asmedigitalcollection.asme.org/tribology/article-abstract/121/2/240/437915/Role-of-Inclusions-Surface-Roughness-and-Operating

[179] N. Soda and T. Yamamoto, "Effect of tangential traction and roughness on crack initiation/propagation during rolling contact," *ASLE Transactions*, vol. 25, no. 2, pp. 198–206, 1982. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/05698198208983081

[180] P. H. Dawson, "Effect of metallic contact on the pitting of lubricated rolling surfaces," *Journal of Mechanical Engineering Science*, vol. 4, no. 1, pp. 16–21, 1962. [Online]. Available: http://journals.sagepub.com/doi/10.1243/JMES_JOUR_1962_004_005_02

[181] Microset, "Microset 101 Thixotropic technical info." [Online]. Available: http://www.microset.co.uk/technical-info/microset-101-thixotropic/

[182] J. S. Stecki and M. Anderson, "Machine condition monitoring using filtergram and ferrographic techniques," *The Research Bulletin of the Centre of Machine Condition Monitoring*, 1991.

[183] Olympus, "Olympus BX53 microscope," 2021. [Online]. Available: https://www.olympus-lifescience.com/en/landing/bx53f2/

[184] Olympus, "Olympus DP72." [Online]. Available: https://www.olympus-ims.com/en/microscope/dp72/

[185] Keyence, "Keyence VK-X200 3D laser scanning microscope." [Online]. Available: https://www.keyence.com/keyence-tv/VK-X200_3D_Laser_Scanning_Microscope.jsp

[186] F. K. Choy, V. Polyshchuk, J. J. Zakrajsek, R. F. Handschuh, and D. P. Townsend, "Analysis of the effects of surface pitting and wear on the vibration of a gear transmission system," *Tribology International*, vol. 29, no. 1, pp. 77–83, 2 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0301679X95000375

[187] P. J. Dempsey, "Gear damage detection using oil debris analysis," in *14th International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management*, 9 2001. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20020009087

[188] T. Toutountzakis, C. K. Tan, and D. Mba, "Application of acoustic emission to seeded gear fault detection," *NDT and E International*, vol. 38, no. 1, pp. 27–36, 1 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S096386950400091X

[189] K. Kato, "Micro-mechanisms of wear - wear modes," *Wear*, vol. 153, no. 1, pp. 277–295, 3 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/004316489290274C

[190] S. Li and A. Kahraman, "Micro-pitting fatigue lives of lubricated point contacts: Experiments and model validation," *International Journal of Fatigue*, vol. 48, pp. 9–18, 3 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142112312003507

[191] R. B. Randall, "A new method of modeling gear faults," *Journal of Mechanical Design*, vol. 104, no. 2, p. 259, 4 1982. [Online]. Available: http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?articleid=1451667

[192] P. Borghesani, W. A. Smith, X. Zhang, P. Feng, J. Antoni, and Z. Peng, "A new statistical model for acoustic emission signals generated from sliding contact in machine elements," *Tribology International*, vol. 127, pp. 412–419, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0301679X18303244

[193] Y. Ding and N. F. Rieger, "Spalling formation mechanism for gears," *Wear*, vol. 254, no. 12, pp. 1307–1317, 11 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164803001261

[194] G. Hoffmann, F. G. Hanejko, and R. H. Slattery, "Crack initiation and propagation in RCF – A new approach to understanding pitting failure of highly loaded gears," in *SAE Technical Papers*. Detroit: SAE, 4 2006. [Online]. Available: https://www.sae.org/content/2006-01-0383/

[195] S. Li, M. T. Devlin, J. Milner, R. Iyer, T.-C. Jao, M. R. Hoeprich, and T. M. Cameron, "Investigation of Pitting Mechanism in the FZG Pitting Test," in *SAE Technical Paper Series*, vol. 1, no. 724. Pittsburgh: SAE, 10 2003. [Online]. Available: https://www.sae.org/content/2003-01-3233/

[196] M. J. Owen, "Low surface energy inorganic polymers," *Comments on Inorganic Chemistry*, vol. 7, no. 4, pp. 195–213, 7 1988. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/02603598808072308

[197] S. Way, "Pitting due to rolling contact," *ASME J. of Applied Mecanics*, vol. 2, pp. A49–A58, 1935. [Online]. Avail-

able: https://asmedigitalcollection.asme.org/appliedmechanics/article-abstract/2/2/A49/1099689/Pitting-Due-to-Rolling-Contact

[198] A. Nakajima, K. Ichimaru, F. Hirano, and M. Nishimura, "Effects of combination of rolling direction and sliding direction on pitting of rollers," *JSLE Int.*, vol. 4, p. 94, 1983.

[199] A. V. Olver, "The mechanism of rolling contact fatigue: An update," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 219, no. 5, pp. 313–330, 2005. [Online]. Available: https://journals.sagepub.com/doi/10.1243/135065005X9808

[200] J. A. Williams, "Wear and wear particles - Some fundamentals," *Tribology International*, vol. 38, no. 10, pp. 863–870, 10 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X05000903

[201] J. E. Merwin and K. L. Johnson, "An analysis of plastic deformation in rolling contact," *Proceedings of the Institution of Mechanical Engineers.*, vol. 203, no. 1, pp. 1–14, 6 1963. [Online]. Available: http://journals.sagepub.com/doi/10.1243/PIME_PROC_1963_177_052_02

[202] S. Ebersbach, Z. Peng, and N. J. Kessissoglou, "The investigation of the condition and faults of a spur gearbox using vibration and wear debris analysis techniques," *Wear*, vol. 260, no. 1-2, pp. 16–24, 1 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0043164804004521

[203] H. Heirani and K. Farhangdoost, "Predicting depth and path of subsurface crack propagation at gear tooth flank under cyclic contact loading," *Journal of Solid Mechanics*, vol. 9, no. 3, pp. 587–598, 2017. [Online]. Available: http://jsm.iau-arak.ac.ir/article_533188.html

[204] W. E. Littmann and R. L. Widner, "Propagation of contact fatigue from surface and subsurface origins," *Journal of Basic Engineering*, vol. 88, no. 3, p. 624, 9 1966. [Online]. Available: http://fluidsengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1432814

[205] P. J. Fernandes and C. McDuling, "Surface contact fatigue failures in gears," *Engineering Failure Analysis*, vol. 4, no. 2, pp. 99–107, 6 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S135063079700006X

[206] P. J. Blau, "An editor's perspective on the quality and content of wear research manuscripts," *Wear*, vol. 426-427, pp. 1384–1390, 4 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004316481831665X?via%3Dihub

[207] C. Yuan, X. Yan, and Z. Peng, "Prediction of surface features of wear components based on surface characteristics of wear debris," *Wear*, vol. 263, no. 7, pp. 1513–1517, 9 2007. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S004316480700333X

[208] W. Liu, "A concise filtergram wear particle atlas and some case studies," in *Handbook of Materials Failure Analysis with Case Studies from the Chemicals, Concrete and Power Industries*. New York: Elsevier, 1 2016, ch. 13, pp. 311–353. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081001165000132

[209] G. E. Totten, A. W. Batchelor, H. Liang, C. Y. H. Lim, T. W. Scharf, E. Van Der Heide, A. Nolan, S. Lampman, W. Haws, E. Kubel, H. Lampman, L. Ryan, J. H. Leyda, E. Marquard, and B. Sanders, *ASM handbook - friction, lubrication, and wear technology*. ASM International, 2017, vol. 18. [Online]. Available: www.asminternational.org

[210] M. Widmark, "Effect of material, heat treatment, grinding and shot peening on contact fatigue life of carburised steels," *International Journal of Fatigue*, vol. 21, no. 4, pp. 309–327, 4 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142112398000772

[211] Y. Peng, J. Cai, T. Wu, G. Cao, N. Kwok, S. Zhou, and Z. Peng, "Online wear characterisation of rolling element bearing using wear particle morphological features," *Wear*, vol. 430-431, pp. 369–375, 7 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164818315552

[212] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 7 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/wics.101

[213] J. Wang and X. Wang, "A wear particle identification method by combining principal component analysis and grey relational analysis," *Wear*, vol. 304, no. 1-2, pp. 96–102, 7 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164813002639

[214] R. S. Dwyer-Joyce, R. S. Sayles, and E. Ioannides, "An investigation into the mechanisms of closed three-body abrasive wear," *Wear*, vol. 175, no. 1-2, pp. 133–142, 6 1994. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0043164894901767

[215] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: A review," *Measurement*, vol. 48, pp. 292–305, 2 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224113005551

[216] Z. Peng, "An integrated intelligence system for wear debris analysis," *Wear*, vol. 252, no. 9-10, pp. 730–743, 5 2002. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164802000315

[217] S. Ebersbach, "Artificial intelligent system for integrated wear debris and vibration analysis in machine condition monitoring," Ph.D. dissertation, University of Queensland, 6 2007. [Online]. Available: https://researchonline.jcu.edu.au/2109/

[218] M. A. Younes, A. M. Khalil, and M. N. Damir, "Automatic measurement of spur gear dimensions using laser light, part 2: measurement of flank profile," *Optical Engineering*, vol. 44, no. 10, p. 103603, 2005. [Online]. Available: https://doi.org/10.1117/1.2114987

[219] W. Wang, P. L. Wong, J. B. Luo, and Z. Zhang, "A new optical technique for roughness measurement on moving surface," *Tribology International*, vol. 31, no. 5, pp. 281–287, 5 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X98000346

[220] J. Zhang, M. Korsten, and R. Paul, "A vision system for online wear detection," in *Proceedings XVII IMEKO World Congress*. Dubrovnik: Croatian Metrology Society, 2003, pp. 1960–1964. [Online]. Available: https://research.utwente.nl/en/publications/a-vision-system-for-online-wear-detection

[221] B. Lee, S. Yu, and H. Juan, "The model of surface roughness inspection by vision system in turning," *Mechatronics*, vol. 14, no. 1, pp. 129–141, 2 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741580200096X

[222] S. Soleimani, J. Sukumaran, A. Kumcu, P. De Baets, and W. Philips, "Quantifying abrasion and micro-pits in polymer wear using image processing techniques," *Wear*, vol. 319, no. 1-2, pp. 123–137, 11 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0043164814002440

[223] I. Bodini, C. Petrogalli, M. Faccoli, M. Lancini, S. Pasinetti, G. Sansoni, F. Docchio, and A. Mazzù, "Evaluation of wear in rolling contact tests by means of 2D image analysis," *Wear*, vol. 400-401, pp. 156–168, 4 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004316481730933X

[224] G. Stachowiak and P. Podsiadlo, "Towards the development of an automated wear particle classification system," *Tribology International*, vol. 39, no. 12, pp. 1615–1623, 12 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X06000302

[225] M. A. Nielsen, *Neural networks and deep learning*. San Francisco: Determination press, 2015. [Online]. Available: http://neuralnetworksanddeeplearning.com/

[226] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11263-015-0816-y

[227] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE*

Transactions on Industrial Electronics, vol. 63, no. 5, pp. 3137–3147, 5 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7386639

[228] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2 2017. [Online]. Available: http://www.mdpi.com/1424-8220/17/2/425

[229] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Computers in Industry*, vol. 105, pp. 182–190, 2 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0166361518305190

[230] A. Prosvirin, J. Kim, and J.-M. Kim, "Bearing Fault Diagnosis Based on Convolutional Neural Networks with Kurtogram Representation of Acoustic Emission Signals," in *Lecture Notes in Electrical Engineering*. Springer, 2018, vol. 474, pp. 21–26. [Online]. Available: http://link.springer.com/10.1007/978-981-10-7605-3_4

[231] M. J. Hasan, M. M. Islam, and J.-M. Kim, "Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions," *Measurement*, vol. 138, pp. 620–631, 5 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0263224119301939

[232] D. Godfrey, "Diagnosis of wear mechanisms," in *Wear control handbook*, M. B. Peterson and W. O. Winer, Eds. New York: ASME, 1980, pp. 283–311.

[233] A. Dyson, "Scuffing - a review," *Tribology International*, vol. 8, no. 2, pp. 77–87, 4 1975. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0301679X75900560

[234] J. Castro and J. Seabra, "Global and local analysis of gear scuffing tests using a mixed film lubrication model," *Tribology International*, vol. 41, no. 4, pp. 244–255, 4 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X07001144

[235] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Online]. Available: http://ieeexplore.ieee.org/document/726791/

[236] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *The 3rd International Conference for Learning Representations*, San Diego, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[237] H. Düzcükoğlu and H. İmrek, "A new method for preventing premature pitting formation on spur gears," *Engineering Fracture Mechanics*, vol. 75, no. 15, pp. 4431–4438, 10 2008. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0013794408001458

[238] S. Odi-Owei, B. Roylance, and L. Xie, "An experimental study of initial scuffing and recovery in sliding wear using a four-ball machine," *Wear*, vol. 117, no. 3, pp. 267–287, 7 1987. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/0043164887901499

[239] P. Peng and J. Wang, "Wear particle classification considering particle overlapping," *Wear*, vol. 422-423, pp. 119–127, 3 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164818311785

[240] D. Wang, Q. Miao, Q. Zhou, and G. Zhou, "An intelligent prognostic system for gear performance degradation assessment and remaining useful life estimation," *Journal of Vibration and Acoustics*, vol. 137, no. 2, p. 021004, 4 2015. [Online]. Available: http://vibrationacoustics.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4028833

[241] J. J. Coy, D. P. Townsend, and E. V. Zaretsky, "Dynamic capacity and surface fatigue life for spur and helical gears," *Journal of Lubrication Technology*, vol. 98, no. 2, p. 267, 4 1976. [Online]. Available: http://tribology.asmedigitalcollection.asme.org/article.aspx?articleid=1463284

[242] F. Zhao, Z. Tian, and Y. Zeng, "Uncertainty Quantification in Gear Remaining Useful Life Prediction Through an Integrated Prognostics Method," *IEEE*

Transactions on Reliability, vol. 62, no. 1, pp. 146–159, 3 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6423860/

[243] S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J. A. Castro-Vargas, S. Orts-Escolano, J. Garcia-Rodriguez, and A. Argyros, "A review on deep learning techniques for video prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 4 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9294028

[244] C. Finn, I. Goodfellow, S. Levine, I. G. Openai, S. Levine, and G. Brain, "Unsupervised learning for physical interaction through video prediction," *Advances in Neural Information Processing Systems*, pp. 64–72, 5 2016. [Online]. Available: http://arxiv.org/abs/1605.07157

[245] Y. Wang, L. Jiang, M.-H. Yang, L.-J. Li, M. Long, and L. Fei-Fei, "Eidetic 3D LSTM: A model for video prediction and beyond," in *ICLR 2019*, 2019. [Online]. Available: https://openreview.net/forum?id=B1lKS2AqtX

[246] I. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," in *NIPS*. arXiv, 12 2016. [Online]. Available: http://www.iangoodfellow.com/slides/2016-12-04-NIPS.pdf

[247] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 5 2015. [Online]. Available: http://arxiv.org/abs/1505.00853

[248] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 1 2018, pp. 586–595. [Online]. Available: http://arxiv.org/abs/1801.03924

[249] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *The 37th Asilomar Conference on Signals,*

Systems & Computers. Pacific Grove: IEEE, 2003. [Online]. Available: https://ieeexplore.ieee.org/document/1292216

[250] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*. ICLR, 9 2015. [Online]. Available: http://www.robots.ox.ac.uk/

[251] H. Chang, P. Borghesani, and Z. Peng, "Investigation on the relationship between macropits and wear particles in a gear fatigue process," *Wear*, vol. 484-485, p. 203724, 11 2021. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0043164821001137

[252] R. Villegas, A. Pathak, H. Kannan, D. Erhan, Q. V. Le, and H. Lee, "High fidelity video prediction with large stochastic recurrent neural networks," in *33rd Conference on Neural Information Processing Systems*, 11 2019. [Online]. Available: http://arxiv.org/abs/1911.01655

[253] Olympus, "Inspections of wind turbine gearboxes." [Online]. Available: https://www.olympus-ims.com/en/applications/rvi-wind-turbine/

[254] G503 Military Vehicle Message Forums, "Automotive gearbox inspection." [Online]. Available: https://forums.g503.com/viewtopic.php?t=276187

[255] S. Sheng, "Monitoring of wind turbine gearbox condition through oil and wear debris analysis: A full-scale testing perspective," *Tribology Transactions*, vol. 59, no. 1, pp. 149–162, 1 2016. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/10402004.2015.1055621

[256] Z. Liu and L. Zhang, "A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings," *Measurement*, vol. 149, 1 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0263224119308681

[257] TWI, "Wind turbine inspection." [Online]. Available: https://www.twi-global.com/technical-knowledge/faqs/how-long-do-wind-turbines-last

[258] D. Chan and J. Mo, "Life cycle reliability and maintenance analyses of wind turbines," *Energy Procedia*, vol. 110, pp. 328–333, 3 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610217301789

[259] S. Muthukkumaran, "Maintenance schedule checklist for wind generators," GE Energy, Tech. Rep., 7 2017. [Online]. Available: https://www.gepowerconversion.com/sites/gepc/files/MaintenanceScheduleChecklistforWindGenerators.pdf

[260] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. [Online]. Available: https://ieeexplore.ieee.org/document/5288526

[261] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review," *Mechanical Systems and Signal Processing*, vol. 126, pp. 662–685, 7 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0888327019301426

[262] I. Errandonea, S. Beltrán, and S. Arrizabalaga, "Digital Twin for maintenance: A literature review," *Computers in Industry*, vol. 123, p. 103316, 12 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0166361520305509

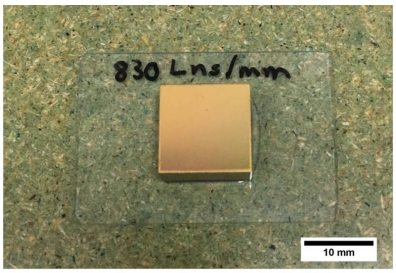
[263] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang, and Z. Zhao, "An overview of data fusion techniques for Internet of Things enabled physical activity recognition and measure," *Information Fusion*, vol. 55, pp. 269–280, 3 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519302258

[264] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Information Fusion*, vol. 57, pp. 115–129, 5 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519303902

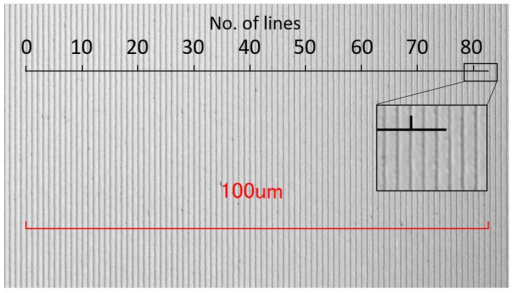
Appendix - Evaluation of Replication Quality

The surface replication method combined with image analysis was applied to reference samples to evaluate the replication quality in resolution, accuracy and profile retention. The micro-scale resolution of the material was tested using a reference surface with 830 line pairs per mm, as shown in Figure A.1. The results of the test (Figure A.1 (b) and (c)) show the line pairs on the reference surface and its mould, respectively, using LSCM with a 50x magnification objective lens. The fact that the mould could capture the surface features spaced at $\sim$ 1 µm suggests a micrometric order for the resolution of the selected moulding material.

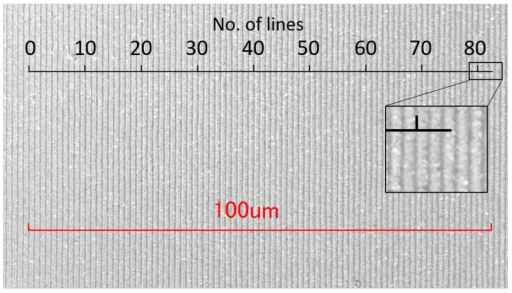
The accuracy and profile retention were assessed by moulding a set of specimens with flat and curved surfaces of known surface roughness. Six surface samples were used: three flat surfaces with ground finishing and three curved surfaces cut off a sample obtained by turning. The flat surface samples had R_a values of 0.4 µm, 0.8 µm, and 1.6 µm, while the curved surfaces had R_a values of 3.2 µm, 6.3 µm, and 12.5 µm (Figure A.2 (a)). The curved surfaces were characterised by manufacturing marks resulting in a characteristic profile wavelength in the order of a few hundred microns.



(a)

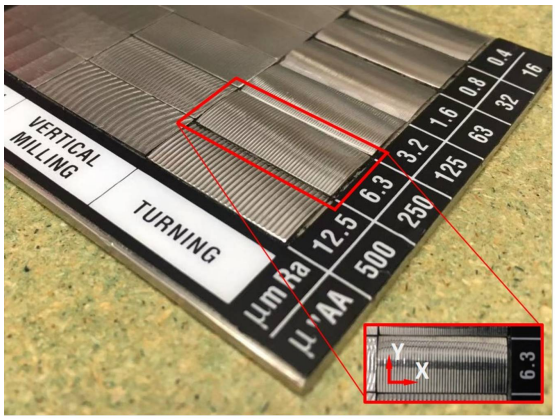


(b)

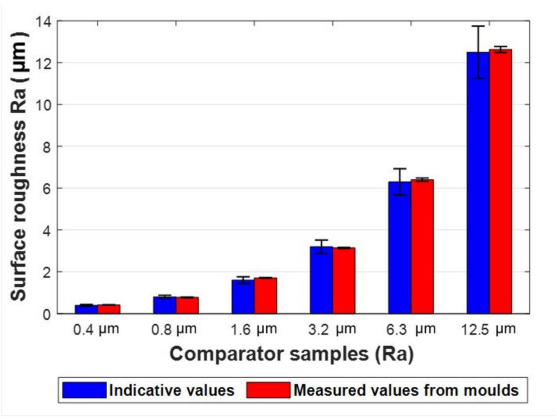


(c)

Figure A.1: (a) The resolution reference sample with a flat surface, (b) a LSCM image of 83 line pairs on the top surface of the reference sample, and (c) a LSCM image of the line pairs on the top of the mould.



(a)



(b)

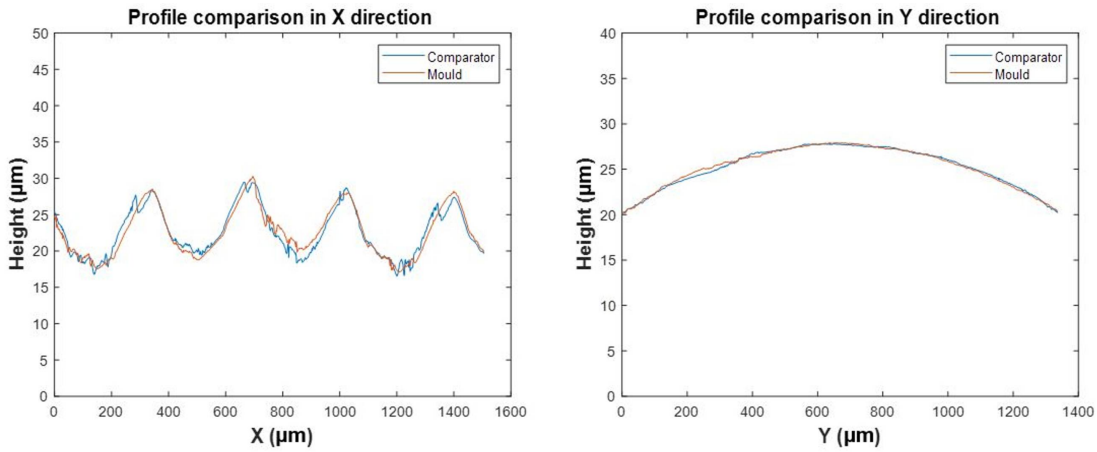
Figure A.2: (a) A surface comparator and (b) the surface roughness of its moulds.

The accuracy was measured in terms of surface roughness, and the surface roughness of the moulds obtained on these samples were measured using LSCM with a 20x objective lens at 5 random locations on each sample. Figure A.2 (b) shows the comparison between

the nominal roughness values (in blue, with tolerance indicated by black whiskers) and the measured quantities (mean in red with whiskers indicating ±1 standard deviation). It can be clearly seen that the measured surface roughness of all six moulds is within the tolerance of nominal values and its standard deviation is lower than 3% of the mean.

Profile retention of the moulding material was confirmed by imaging the curved surface with $R_a = 6.3$ µm and the corresponding mould. The profile of the mould and the surface were scanned using LSCM with a 20x objective lens across two rectangular strips oriented perpendicularly to each other. Each strip had a length of $\sim$ 1,400 µm and a width of $\sim$ 700 µm. The profiles of Figure A.3 (a) and (b) were then obtained by averaging the scan along the strips' width to remove local random features. The averaged profiles were characterised by the manufacturing-induced waviness with a wavelength of $\sim$ 300 µm in the X direction (Figure A.3 (a)) and an arc with a radius of $\sim$ 30 mm in the Y direction (Figure A.3 (b)), respectively.

In the X direction, the RMS error, which is the RMS of the height difference between the actual and mould-based estimation of the profile, is 0.99 µm, and the maximum error is 2.54 µm. In the Y direction, the RMS error is 0.22 µm and the maximum error is 0.64 µm.



(a) (b)

Figure A.3: Comparisons of profiles between the comparator and its mould in (a) X direction, and (b) Y direction.