APPROACHES FOR DIAGNOSIS AND PROGNOSIS OF ASSET CONDITION: APPLICATION TO RAILWAY SWITCH SYSTEMS

by

QIANYU CHEN

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Department of Electronic,

Electrical and Systems Engineering

School of Engineering

The University of Birmingham

August 2021

ABSTRACT

This thesis presents a novel fault diagnosis and prognosis methodology which is applied to railway switches. To improve on existing fault diagnosis, energy-based thresholding wavelets (EBTW) are introduced. EBTW are used to decompose sensor measurement signals, and then to reconstruct them within a lower dimensional feature vector. The extracted features replace the original signals and are fed into a neural network classifier for fault diagnosis. Compared to existing wavelet-based feature extraction methods, the new EBTW method has the advantage of an intrinsic energy conservation property during the wavelet transform process. The EBTW method localises and redistributes the signal energy to realise an efficient feature extraction and dimension reduction.

The presented diagnosis approach is validated using real-world switch data collected from the Guangzhou Metro in China. The results show that the proposed diagnosis approach can achieve 100% accuracy in identifying a railway switch overdriving fault with various severities, improving upon existing methods of conventional discrete wavelet transform (C-DWT) and soft-thresholding discrete wavelet transform (ST-DWT) by 8.33% and 16.67%, respectively.

The presented prognosis approach is constructed based on traditional data-driven prognosis modelling. The concept of a remaining maintenance-free operating period (RMFOP) is introduced, which transforms the usefulness of sensor measurement data that is readily available from operations prior to failure. Useful features are then extracted from the original measurement data, and modelled using linear and exponential regression curve fitting models. By extracting key features, the original measurement data can be transformed into degradation

signals that directly reflect the variations in each movement of a switch machine. The features are then fed into regression models to derive the probability distribution of switch residual life. To update the probability distribution from one operation to the next, Bayesian theory is incorporated into the models.

The proposed RMFOP-based approach is validated using real-world electrical current sensor measurement data that were collected between January 2018 and February 2019 from multiple operational railway switches across Great Britain. The results show that the linear model and the exponential model can both provide residual life distributions with a satisfactory prediction accuracy. The exponential model demonstrates better predictions, the accuracy of which exceeds 95% when 90% life percentage has elapsed. By applying the RMFOP-based prognosis approach to operational data, the railway switch health condition that is affected by incipient overdriving failure is predicted.

ACKNOWLEDGEMENTS

The author thanks the following people and parties for their contributions to the research: Professor Clive Roberts, for his experienced guidance and assistance as the first supervisor; Dr. Gemma Nicholson, for her detailed advice and support as the second supervisor; Mr. Louis Saadé, for his help with data collection; The School of Engineering, for covering the course fees and maintenance fees during the author's research course; Network Rail, for facilitating the data collection from point machines over multiple trainlines.

Finally, I would like to give my heartfelt thanks to my parents and boyfriend for their experienced advice, love, and encouragement throughout the research course.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1-D CNN	One-dimensional CNN
AI	Artificial intelligence
ANFIS	Adaptive neural-fuzzy inference system
ANN	Artificial neural network
BN	Bayesian network
CCNN	Cascade correlation neural network
cdf	Cumulative distribution function
C-DWT	Conventional discrete wavelet transform
CH	Calinski-Harabasz
CNN	Convolutional neural network
CVI	Cluster validity index
DB	Davies-Bouldin
DBN	Dynamic Bayesian network
Deep NN	Deep neural network
DGA	Dissolved gas analysis
DRR	Dimension reduction ratio
DWT	Discrete wavelet transform
EBTW	Energy-based thresholding wavelets
FFNN	Feedforward neural network
GAN	Generative adversarial network

GB	Great Britain
HMM	Hidden Markov model
kNN	k-nearest neighbour
MTTR	Mean time to repair
OOBN	Object-oriented Bayesian network
ORR	Office of rail and road
PCA	Principal component analysis
pdf	Probability density function
PHM	Prognostics and health management
QTA	Qualitative trend analysis
RMFOP	Remaining maintenance free operating period
RUL	Remaining useful life
SBPD	State-based prognostics with state duration information
Shallow NN	Shallow neural network
SSBP	Simple state-based prognostic model
ST-DWT	Soft-thresholding discrete wavelet transform
SVM	Support vector machine
POE	Point operating equipment

CHAPTER 1

INTRODUCTION

The railway transportation system has been under huge pressure to facilitate increased passenger volumes with higher reliability, lower costs and higher speed. In order to enhance the overall system performance, effective management of inspection, maintenance and renewal of the relevant railway assets has great significance. According to an annual assessment report from Office of Rail and Road (ORR), the network-wide reliability across Great Britain reduced in 2019-20 [1]. Typically, the number of failures in electrically powered assets, such as switches and level crossings, has increased with more negative impacts on normal train operations. For example, in 2018-19, train delays caused by asset failures accounted for over half of the total delay minutes in Great Britain [2]. A railway switch system, which directs the coming train onto the correct tracks at railway junctions, is a key railway asset. Current inspection and maintenance strategies could be improved by investigating more timely and

effective condition monitoring methods. In the following sections, the precise research problem is defined.

1.1 Research Problem Definition

1.1.1 Application Background: Railway Switch System

A railway switch, or (set of) points is a mechanism that can divert trains from one track to another by driving a moving rail at a railway junction. The switch is composed of a pair of linked tapering rails, known as points (switch rails), situated between the diverging outer rails (stock rails), as illustrated in Fig. 1.1 [3]. The points can be moved laterally in either of two directions (i.e., normal-to-reverse or reverse-to-normal) within a few seconds, to direct the arriving trains towards the correct track. Historically, it would require a human operator to move the points from one position to another, and there still remain some hand-controlled switches. Most switches are now driven remotely by actuators, called point machines. Railway switches are categorized into electro-mechanical, electro-hydraulic and electro-pneumatic types based on their distinct power operations. Among the three, the electro-mechanical and electro-hydraulic powered point machines are applied more commonly.

Fig. 1.1 Railway switch in an operational railway environment

At present, visual inspections of switches are carried out periodically by experienced personnel. Some mechanical measurement tools are normally carried to identify faults in railway switches. Reactive maintenance is then employed based on the inspection results. However, it cannot be denied that the possibility of human error always exists during normal inspections. Faults and failures may also happen between inspections. Thus, a balance needs to be found on inspection frequency. Failure to plan and carry out inspection frequently enough has the potential to cause severe consequences. Too frequent inspections are costly and give increased exposure to risk to the staff carrying them out. For example, the Grayrigg derailment was a fatal railway accident that happened on 23 February 2007 [4]. The scheduled inspection on 18 February 2007 did not take place and switch faults had gone undetected. From this point of view, the construction of effective condition monitoring

techniques that are applicable to railway switches is an essential part of the whole procedure, which can dramatically affect the switch maintenance quality and the overall railway capability to perform at high capacity.

1.1.2 Research Topic: Condition Monitoring and Fault Analysis

Condition monitoring techniques are commonly used on rotating equipment such as gearboxes and bearings, and other equipment including pumps, electric motors, compressors and presses [5]. Accelerometers are usually installed to measure the bearing casing vibrations [6]. Eddy-current sensors are also applied to measure radial displacement of the shaft [7]. Interpreting and analysing these sensor signals requires specialized experience and knowledge [8]. In other words, the developed condition monitoring algorithms are specifically tailored to the applications [9]. The research methods can significantly vary by machine type, from the selection of measurement sensors, the understanding of signal features, to the construction of analysis models [10].

A railway switch is an industrial reciprocating machine, and there is still a lack of comprehensive condition monitoring methodologies for this type of machine. Therefore, this research is focused on investigating the potential of condition monitoring techniques as a feasible approach to addressing the important

problem of railway switch fault/failure analysis. When performing advanced condition monitoring techniques on railway switches, the system state can be evaluated more precisely. Maintenance activities can be scheduled more efficiently with an enhanced system reliability, and the delay caused by unexpected maintenance intervention will be mitigated and track availability can be improved.

An asset management framework based on the ISO55000 series is used by Network Rail to continuously improve efficiency and reliability of operational railway assets [11]. The stages in the asset management framework are mainly composed of establishing proper objectives and standards, planning delivery strategies, executing planned work, and improving current performance, as demonstrated in Table 1-1. Condition monitoring techniques normally appear at the final stage to support improvement of the planned work by means of measuring and analysing system performance.

Table 1-1: Asset Management Framework Adjusted From [11]

	Framework Stage
Plan	1-Organizational Objectives and Strategy
	2-Rules and Standards
	3-Annual Delivery Plan
Do	4-Detailed Delivery Schedules
	5-Execution of Work
Review	6-Performance Analysis and Improvement

Condition monitoring is defined as the process of measuring specific machine parameters that can reflect the machine health condition, in order to identify the signs of any abnormal changes that could indicate an impending fault/failure. It is a major step in scheduling predictive maintenance, which is the type of maintenance based on knowledge of what fault or failure might happen and therefore what maintenance intervention should be planned to terminate or slow down the development of such faults or failures. The lifetime of a monitored machine is also generally extended as the result of proper maintenance, preventing development into severe functional failures. In other words, effective condition monitoring brings decreased maintenance costs, decreased downtime, extended asset lifetime and enhanced reliability.

Historically, data collection for condition monitoring has been carried out at discrete intervals and undertaken manually by trained individuals, who are responsible for taking a periodic condition check while a machine is still operating without obvious faults. The collected data can only give a snapshot of the machine status at the time of collection. More recently, with the rapid development of measurement sensors and wireless transmission, condition monitoring has employed the digital measurement of machine parameters through wirelessly transmitted machine-mounted sensors. A remote condition monitoring system can receive and display the sensor signals continuously. Advanced condition monitoring systems for some applications can even automatically pre-process and analyse multiple types of received sensor data, in order to accurately detect, diagnose and give a prognosis of machine fault/failure performance. In terms of machinery condition monitoring, fault and failure indicate two distinct machine health conditions. A failure implies the incapability of a machine to implement what is supposed to, while a fault means the potential for failure at a point when it may be possible to prevent the failure with an intervention.

Fig. 1.2 shows a data processing flowchart of the type needed to implement condition monitoring effectively. It can be seen that machine condition

assessment is divided into five distinct, layered processing blocks [12]. The **Data Collection block** combines the outputs from a sensor/transducer and represents them digitally as a physical quantity (i.e., temperature, humidity, time period, current supply and voltage supply). The **Fault Detection block** searches for abnormalities whenever new data are obtained. An alert or alarm may be raised when faults are identified. The **Fault Diagnosis block** determines the nature of the fault and rates the current health of the machine considering all state information. The **Fault Prognosis block** predicts future failures and remaining useful life (RUL) based on current and past health condition, as well as on projected usage frequencies and maintenance records. The **Advisory Generation block** provides recommended actions to extend the equipment useful life. Since the information flow goes sequentially from data collection to advisory generation, each processing block requires reliable outputs from the earlier processing blocks. In addition, extra information is obtained from/sent to external systems. Information displays are also essential in that the data from distinct processing blocks are converted to a format that represents clearly the information for making corrective-action decisions, such as the fault/failure type, an estimation of the severity, a prediction of the condition and, eventually, suggested operations.

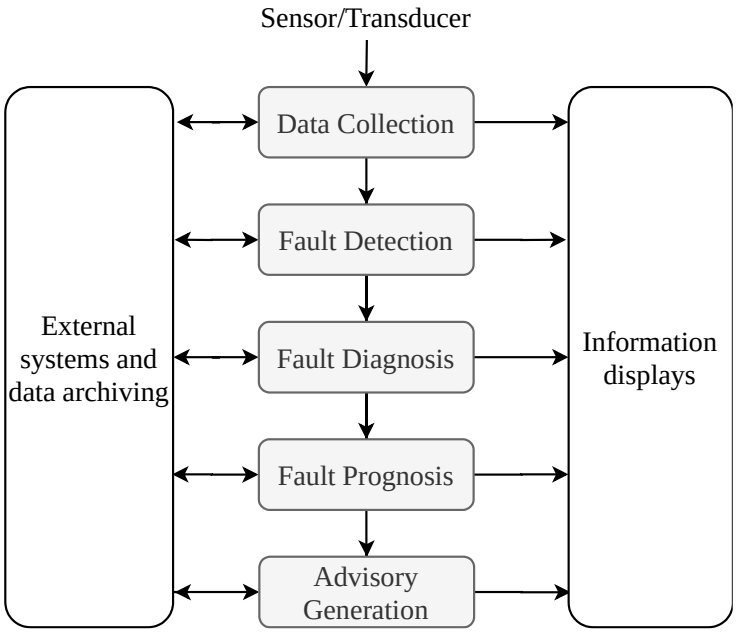

Fig. 1.2 Data processing flowchart

Among the five layered processing blocks, the first four blocks (i.e., data collection, fault detection, fault diagnosis and fault prognosis) are technology focused, requiring signal processing, meaningful descriptor derivations and effective modelling constructions from the raw measurements. Among these four, data collection and fault detection are excluded from this research scope because they have been well developed in both research and practice, the techniques of which are relatively mature compared with fault diagnosis and prognosis [13]. Therefore, the focus of this research is to enhance the present fault diagnosis and prognosis techniques with respect to the diagnosis reliability and quality, as well as the predictions of health condition and potential faults/failures for a real-world railway switch application.

Fault diagnosis evaluates the machine current health condition and determines the actual fault type by assessing the monitored sensor information [14]. Diagnosis approaches can be classified into three categories. These are: 1) statistical approaches, 2) artificial intelligence (AI) approaches, and 3) physical approaches. Statistical approaches aim to recognize, differentiate and put measurement data sets into categories (i.e., corresponding to fault types) based on their similarities [15]. The data within the same category are more similar (e.g., smaller distance or denser data space) to each other, than to those in other categories. Artificial intelligence approaches have become more popular in recent years. They are a type of computational model used to mimic the human brain structure [16]. Typically, multiple processing units are connected to each other in a complex layered configuration, in order to express the non-linear relationships between the multi-input and multi-output systems. Another type of machine diagnosis method is model-based physical approaches, which use specifically built mathematical models of the monitored device and vary significantly for different machine types [17]. Model-based approaches generate and evaluate residuals with respect to their physical structures. A residual is an indicative index of machine faults, which shows an anomalous deviation from the ideal behaviour.

Compared with diagnosis, the capability in prognosis is relatively immature. Fault prognosis aims to estimate the future changes in machine health condition over time and derive the RUL based on the observations from current and past histories, such as the machine usage frequencies and maintenance records [18]. Prognosis models can be divided into three categories. These are: 1) data-driven models, 2) physics-based models, and 3) knowledge-based models. Data-driven models use run-to-failure data to establish an artificial intelligence or statistical model, which can accommodate the degradation process and estimate a system's residual life [19]. Usually, the degradation patterns are extracted from original data with calculated confidence intervals. Physics-based models incorporate mathematical expressions that are used to describe the machine physical degradation behaviours from first principles [20]. Knowledge-based models give the residual life predictions by matching a newly observed situation with failure events that were defined previously in a databank [21]. A databank is a series of rules that experts use to describe the encountered experience of failures.

Fault diagnosis is concerned with detection and identification when the faults have occurred. Prognosis, however, focuses on predicting faults before they

happen. The reason for investigating both diagnosis and prognosis together in the area of condition monitoring is that each cannot completely replace the other in real life applications. Instead, diagnosis and prognosis can work as complementary tools for each other, in order to improve condition monitoring effectiveness and provide better maintenance support since there will always exist some sudden faults that cannot be well predicted. Under these circumstances, fault diagnosis can be implemented to give comprehensive decision support. The diagnosis information can also be regarded as useful feedback for system redesign.

The methods described in both diagnosis and prognosis have their own advantages and disadvantages in practice. There is no agreement on a universally recognized best approach owing to the variations in accessible data amount, applicable experiment scenarios and model parameter assumptions. This research considers the benefits and limitations of different method categories, followed by investigating solutions to improve fault diagnosis and fault prognosis capability for a real-world railway switch application.

The scope of this research is to investigate and improve fault diagnosis and fault prognosis methodologies that are applicable to the condition monitoring of a

real-world railway switch system. As mentioned earlier in Section 1.1.2, current condition monitoring diagnosis literature is categorized into statistical approaches, AI approaches, and other approaches. Among these three types of methods, AI diagnosis methods have been selected for further deep investigation owing to their properties of modelling real-world complex systems without detailed requirements of physical understanding of the system behaviour, which can be difficult to determine in defects for assets such as switches that work in a range of scenarios and environmental conditions. Subsequently, among the three prognosis models described in Section 1.1.2 (i.e., data-driven models, physics-based models and knowledge-based models), data-driven prognosis models are selected for exploration for this application.

1.2 Research Aims & Objectives

The aim of this research is to investigate and improve current fault diagnosis and failure prognosis techniques in terms of the performance (i.e., reliability and quality of diagnosis, and prediction of failures) and the applicability to a real-world system, that is, a railway switch system.

The objectives of this research are as follows:

1. To categorize the established literature on machine fault diagnosis and fault

prognosis, in order to identify the potential type of methods that may be applicable to the condition monitoring of a real-world railway switch system;

2. To investigate the effectiveness and sensitivity of the installed sensors to machine faults by comparing the sensor data collected from various measurement devices in a real-world railway switch system;

3. To propose a fault diagnosis methodology and validate its effectiveness in terms of the diagnosis performance, such as the computation complexity and decision accuracy, using sensor data collected from real-world railway switches for comparison with some established diagnosis techniques;

4. To propose a fault prognosis technique that is applicable to railway switches in practice using real-world rather than idealized lab data, and prove its capability compared to existing fault prognosis methods;

5. To determine the applicability of and recommend the generalization of the proposed methodologies beyond the railway switch case.

1.3 Case Studies

The proposed diagnosis and prognosis methodologies are validated using two separately collected sets of data from railway switches. Dataset 1 includes condition monitoring data from different fault types and can be used for diagnosis in AI modelling. Dataset 2 consists of continuous degradation data and will be utilized for prognosis in data-driven modelling. The applicability of the proposed methods is explored by assessing their mathematical principles, approximation assumptions and applicable scenarios. The analysis outcomes contribute to making more intelligent and precise fault detection, identification and prediction for the application of a railway switch system. Compared with the traditional reactive 'go-and-fix' schemes, condition-based maintenance can be scheduled and conducted in a proactive way.

1.4 List of Publications

A list of publications related to this research is listed below:

Journal papers:

1. Chen, Q., Nicholson, G., Ye, J., Zhao, Y. and Roberts, C., 2020. Estimating Residual Life Distributions of Complex Operational Systems Using a Remaining

Maintenance Free Operating Period (RMFOP)-Based Methodology. *Sensors*, 20(19), p.5504; **DOI:** 10.3390/s20195504

2. Chen, Q., Nicholson, G., Roberts, C., Ye, J. and Zhao, Y., 2020. Improved Fault Diagnosis of Railway Switch System Using Energy-Based Thresholding Wavelets (EBTW) and Neural Networks. *IEEE Transactions on Instrumentation and Measurement*, 70, pp.1-12; **DOI:** 10.1109/TIM.2020.3029365

Conference proceedings:

3. Chen, Q., Nicholson, G., Ye, J. and Roberts, C., 2020, May. Fault Diagnosis Using Discrete Wavelet Transform (DWT) and Artificial Neural Network (ANN) for A Railway Switch. In *2020 Prognostics and Health Management Conference (PHM-Besançon)* (pp. 67-71). IEEE; **DOI:** 10.1109/PHM-Besancon49106.2020.00018

1.5 Structure of the Thesis

The remainder of this thesis is organized as follows:

Chapter 2 introduces the state-of-the-art diagnosis and prognosis literature based on their categorizations. The challenges and limitations are also introduced by comparing the approaches within each category.

Chapter 3 demonstrates the operational setup of the railway switch system. The selected measurement sensors for data collection, as well as the dataset descriptions are explained in this chapter.

Chapter 4 describes in detail the newly proposed methodologies for improving the current AI diagnosis modelling and data-driven prognosis modelling.

Chapter 5 presents the performance results obtained from every methodology discussed in Chapter 4 by manipulating the data sets described in Chapter 3.

Chapter 6 provides a conclusion based on the research presented in this thesis and discusses possible directions for related future work.

CHAPTER 2

LITERATURE REVIEW

The scope of this research is to explore and enhance fault diagnosis and prognosis methods that are applicable to condition monitoring of real-world railway switches. Currently, switch inspections are conducted at intervals based on different track categories defined by the traffic speed and equivalent tonnage using the line. For example, visual inspections of switches and crossings are conducted every week and the maximum permitted interval between two adjacent inspections is 8 days, except for non-strengthened switches and crossings in some specific track categories. Routine periodic maintenance is then carried out, such as repairing fasteners in sections and greasing the fasteners and joint bolts, as well as replacing defective rails, welding and polishing rails and repairing joint damage, etc. However, there always exists the possibility of human error during both the inspection and maintenance process. Extreme weather and intense traffic flow might also increase the risk of accidents.

There are more than twenty failure modes that have been identified for a railway switch system [22]. The most frequently failed components are the motor, slide chair and drive rod. The failures related to the motor are mainly due to its inner components, such as worn brushes. The slide-chair-related issues mainly come from contaminated, dried or defective working conditions. Problems associated with the drive rod happen most frequently among the three, occupying 40% of failure reports across Great Britain [22]. The common failures are mis-adjustment and defective physical structures. Except for the aforementioned motor, slide chair and drive rod in Britain, other sub-components including heaters, cables and relays may also create switch failures. This research is focused on the problems associated with the switch drive rod only.

An enormous amount of fault diagnosis and prognosis literature has already been published, but with very limited discussion on the application scenarios of railway switches. In other words, most published research methods are not directly applicable to the railway switch related problems considered as part of this research. Therefore, this chapter aims to present a comprehensive literature review of current fault diagnosis and fault prognosis methods that are developed for various engineering applications. These modelling methods are categorized and the advantages and disadvantages of the methods in each category are

discussed and compared, in order to find the most appropriate method category that can potentially solve the proposed research problem of railway switch health monitoring.

2.1 Fault Diagnosis

Fault diagnosis aims to detect machine abnormalities and to identify the fault type based on analysis of the machine's health monitoring data. Condition-based monitoring with diagnosis results helps schedule relevant maintenance activities once an abnormality has been identified, the application of which to railway switches can prevent the progression of faults to failures during normal train operations.

Based on this review of relevant research articles, there is almost no consensus in the diagnosis field as to which model type can potentially show the best performance for a specified application, here that is railway switch systems, in terms of diagnosis quality and accuracy. Generally speaking, diagnosis models are categorized into three types, which are:

1. Statistical approaches,
2. Artificial intelligence (AI) approaches, and

3. Physical approaches.

A categorization chart is shown in Fig. 2.1, consisting of a hierarchy of diagnosis models considering the model complexity, cost, accuracy, and the range of applicability. Among the three types of approaches, statistical approaches are considered to have the lowest relative implementation costs, with the largest range of applicability in engineering. However, the diagnosis accuracy of these approaches is inferior to that of physical approaches. AI approaches are situated in the middle of aforementioned two approach types with moderate model complexity and mathematics requirement. The different methods listed in each category will be discussed in detail in the following subsections (i.e., Sections 2.1.1-2.1.3).

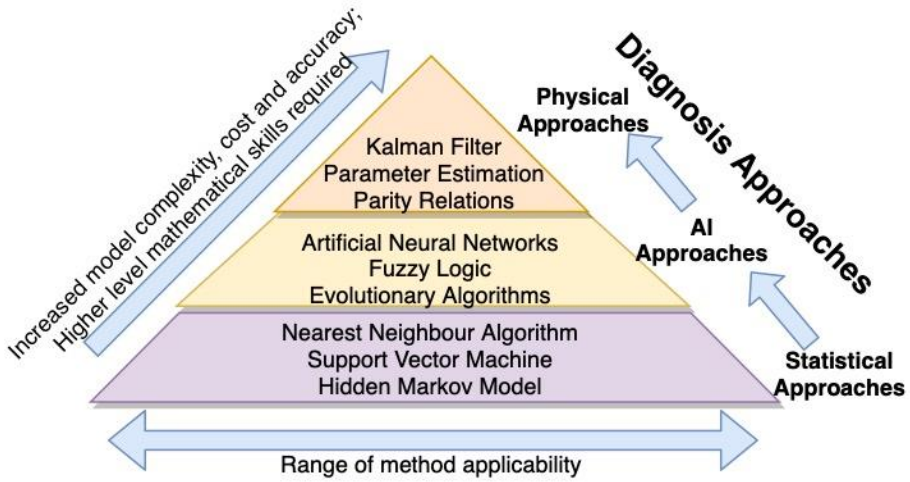

Fig. 2.1 Diagnosis models hierarchy

2.1.1 Statistical Approaches

Statistical approaches are a common method for fault diagnosis to determine whether a specific fault exists or not under known conditions.

A Bayesian Network (BN) is one of the earliest applications of statistics principles to diagnosis analysis, which classifies machine health monitoring data (fault-free and faulty) by applying probabilistic graphical modelling that can deal with different uncertainty problems considering probabilistic information presentation and inference [23]. In recent research, some advances beyond traditional BNs to achieve an improved performance have been made. A Dynamic Bayesian Network (DBN) is demonstrated to have an increased capability in fault diagnosis and prognosis by incorporating information from multiple sources, with its successful applications to hydraulic actuator system [24] and offshore drilling [25]. An Object-Oriented Bayesian Network (OOBN) also demonstrates its superiority over traditional static BNs by introducing hierarchical representations [26] [27]. Although Bayesian techniques can precisely model multivariate and noisy processes, the measurement noise and the whole process are assumed to follow Gaussian distributions, which is difficult to achieve for a real-world machine with a variety of correlated parameters and unknown distributions.

The nearest neighbour algorithm is another conventional tool in classifying machine faults, which assigns an observation to the health condition most common among its nearest neighbours [28]. In this algorithm, the number of nearest neighbours and distance measures can be tuned for optimized results. For example, a fault diagnosis algorithm was developed for photovoltaic systems based on nearest neighbours algorithms, which achieved a high validation accuracy in classifying four common faults including line-line fault and bypass diode fault [28]. Also, a nearest neighbour based Dissolved Gas Analysis (DGA) was investigated for insulating oil faults in power transformers [29]. It should be noted that the nearest neighbour algorithms may not achieve a reliable outcome when applied to a complex industrial system exposed to noisy and variable working environments.

A method called Support Vector Machine (SVM) is normally applied to solve fault classification problems by searching for suitable decision boundaries that allow the data to be split into different sets, with each set belonging to a distinct fault category [30]. A decision boundary is a curve splitting the historical data into various areas corresponding to different known fault states. A suitable boundary curve helps to map unknown data into the correct data space. As a result of its

superior performance in the case of small data size, SVM has been widely employed to machinery fault diagnosis in rolling element bearings [31], induction motors [32], turbines [33], and power systems [34]. When considering its applicability to railway switch machines proposed in this research, it may not be the best choice if dealing with large noisy data sets.

The Hidden Markov Model (HMM) can also be used for fault classification. It is a stochastic method for modelling signals evolving through a number of discrete states. These states are assumed to be hidden and responsible for generating observations. Fault diagnosis is achieved by characterizing the hidden states from observations [35]. Its combination with Principal Component Analysis (PCA) has been applied for fault diagnosis of wind energy converter systems [36]. Another combination with wavelet packet decomposition was used for fault diagnosis of rolling bearings [37]. The combination of HMMs with other techniques, such as the aforementioned PCA and wavelet packets, can provide relatively accurate results while compensating for the weakness of intensive computation if using HMMs alone.

2.1.2 Artificial Intelligence (AI) Approaches

With the rapid development of smart sensory techniques over the past two decades, huge amounts of data can be automatically collected in a short period of time. Compared with many traditional statistical approaches, AI approaches have recently gained more attention in solving machine diagnosis problems, and have demonstrated an enhanced capability in improving data quality, reducing data redundancy and boosting analysis efficiency.

Artificial Neural Networks (ANNs) are the most common AI technique, which contain a complex configuration with multiple layers of data processing units enabling modelling of the non-linear relationship between multiple inputs and multiple outputs [38]. A Feed-Forward Neural Network (FFNN) with a backpropagation algorithm is the most common type of ANN used for machine fault diagnosis, with successful application to railway switches [39], voltage source inverters [40] and bearings [41]. A Cascade Correlation Neural Network (CCNN) requires no initialization of network construction; therefore, it is more suitable for on-line diagnosis compared with the traditional FFNN. An application of CCNN to bearing fault diagnosis shows that minimal network construction is required to achieve a satisfactory accuracy [42]. Other neural networks that are applicable to machine fault diagnosis include recurrent neural networks [43],

generative adversarial networks [44], and convolutional neural networks [45]. In terms of a concrete fault diagnosis problem, ANN variants show significant superiority in modelling complex, non-linear systems without the requirement of physical understanding of system behaviours. In other words, their strong applicability facilitates use in real-world complex problems without limitations on the input data type and application scenarios. In addition, user-friendly computer software is available for modelling, which reduces the difficulty of implementation.

Unlike neural networks that automatically learn knowledge from training observed data with known input and output pairs, fuzzy logic can make full use of domain expert knowledge to solve machine diagnosis problems more robustly. In other words, the uncertainties in domain expert knowledge can be measured qualitatively using fuzzy logic methods. An example of utilizing fuzzy logic for machine diagnosis was shown for dissolved gas analysis of various transformer incipient faults [46]. A more recent application of fuzzy logic is the combination with other techniques. An adaptive neural-fuzzy inference system (ANFIS) is a popular method incorporating fuzzy logic with neural networks, which has been employed for fault diagnosis in planetary gears [47], power transformers [48] and induction motors [49]. When modelling noisy and imprecise inputs, fuzzy

logic often gains an advantage in transferring imprecise inputs to precise rules using domain expert knowledge. As such, other algorithms combined with fuzzy logic could give more reliable outcomes.

In addition, evolutionary algorithms such as genetic algorithms, which mimic the biological evolution progression of a population to achieve probabilistic searches of some combinations, have also shown some merits in specific applications [50, 51]. However, the key objective function within the genetic algorithm is not easy to design, especially for some application areas that have not been studied in depth, such as the railway switches proposed in this research.

2.1.3 Physical Approaches

In addition to the aforementioned statistical approaches in Section 2.1.1, and the AI approaches in Section 2.1.2, another type of solution for machine fault diagnosis is physical approaches. These approaches are built upon physics-based mathematical models of the specified machine faults. Specifically, the residual, which is an indicative signal of the presence of machine faults, is first generated from machine observations by utilizing a Kalman filter, parameter estimation, parity relations, or other residual generation approaches. These residuals are then evaluated to arrive at fault identification. For example, the fault diagnosis

problem was considered for a linear drive system that is subject to system noise [52]. To solve this problem, a Kalman-filter-based residual generator was designed, and a fault-tolerant control method was used to accommodate the specific failure types. Also, a combination of an adaptive Kalman filter with a joint-state parameter estimation was considered for actuator fault diagnosis in analysing lateral dynamics of a remotely piloted aircraft [53]. Some published literature applied the residuals principle to detect and diagnose railway switch faults, demonstrating the applicability of physics-based fault diagnosis approaches to real-world railway switches [54-56]. One advantage of physics-based approaches is that they require less performance data than AI approaches, which may be difficult to collect. Performance data in physics-based modelling is used for model validation, whereas it is central to the effectiveness of the whole approach in AI methods.

2.1.4 Challenges in Diagnosis Modelling

The added value of this fault diagnosis literature review is summarized in Table 2-1, which discusses the advantages and disadvantages of the diagnosis methods in each category, with an emphasis on the possibility and existing challenges in being applied to real-world railway switch machines.

The comparisons summarized in Table 2-1 indicate that AI approaches are most likely to be successfully applied to railway switches because of their capability to model practical systems with complex structures and working principles by learning automatically from a large amount of monitoring data. Therefore, in order to solve the proposed condition monitoring problem for real-world railway switches, AI diagnosis approaches have been selected to be investigated further. In the next chapters (i.e., Chapters 3, 4 and 5), data collection, method exploration, and result analysis will be successively carried out around the topic of AI-based fault diagnosis for the railway switch fault diagnosis application.

Table 2-1: Summary of fault diagnosis literature review

Methodologies	Advantages	Disadvantages	Applicability for real-world railway switches
Statistical Approaches:			
Bayesian Techniques	Can model multivariate, dynamic and noisy process	Measurement noise and whole process must be Gaussian distributed	Statistical approaches use specific models containing parameters of specific distributions, in order to interpret the condition monitoring information into an expression. However, it is hard to establish an explicit statistical diagnosis model for real-world railway switches considering the various parameters it contains, and their distributions and correlations.
Nearest Neighbour Algorithm	Simple and versatile technique to use and explain	Results are affected by condition variations and system noise	
Support Vector Machine	Works well on small amount of data	Not suitable for noisy, large-size data	
Hidden Markov Model	Can model temporal and spatial data	Large computation required, proportional to number of hidden units	
AI Approaches:			
Artificial Neural Networks	Can model practical systems without expert knowledge	Need a large amount of data; large model complexity and computation time	AI approaches can make full use of condition monitoring data to automatically learn and update model structure. The advanced sensor techniques that have been applied to railway switches provide a large amount of effective data for model training and validation.
Fuzzy Logic	Can model noisy, imprecise inputs	Require domain expert knowledge to set rules	
Evolutionary Algorithms	Easily modified and adaptable to various problems	Objective function is hard to design; computationally expensive	
Physical Approaches:			
Physical Models	Can provide a sound understanding of the system performance	A reference model needs to be built, requiring more prior knowledge than AI approaches	In the case of a real-world railway switch system, physics-based approaches use the residuals principle to diagnose switch faults showing some successful applications [54-56].

2.2 Fault Prognosis

In addition to fault diagnosis, fault prognosis is another major concern in condition monitoring and condition-based maintenance. Fault prognosis involves forecasting the future health condition of the current machine of interest by exploring the continuous changes in machine health descriptors from the present until a defined fault threshold condition is reached. The outcome from prognosis is the time left before a fault occurs, which is usually called Remaining Useful Life (RUL). As such, condition-based maintenance with RUL estimations could contribute to more effective decision making on operational and maintenance activities, resulting in improved Mean Time To Repair (MTTR) and the associated costs, for example, in its application to railway switch systems.

Similar to fault diagnosis, the approaches to fault prognosis fall into three main categories. There is little agreement on which category provides the most effective prognosis models overall. The categories are:

1. Knowledge-based approaches,
2. Data-driven approaches, and
3. Physics-based approaches.

Fig. 2.2 shows a hierarchy of the three modelling types with respect to the range of applicability, model complexity, cost and accuracy. Among the three types of prognosis approaches, knowledge-based approaches are the most affordable with the widest range of applicability. At the same time, the prediction accuracy of knowledge-based approaches is the lowest, followed by data-driven approaches. Physics-based approaches are the most accurate of the aforementioned two approach types; they have the highest level of model complexity. In the following three subsections (i.e., Sections 2.2.1, 2.2.2 and 2.2.3), the common methods belonging to each category are discussed in detail, with the purpose of identifying the most appropriate model type that can be potentially applicable to the specified field of railway switch health monitoring.

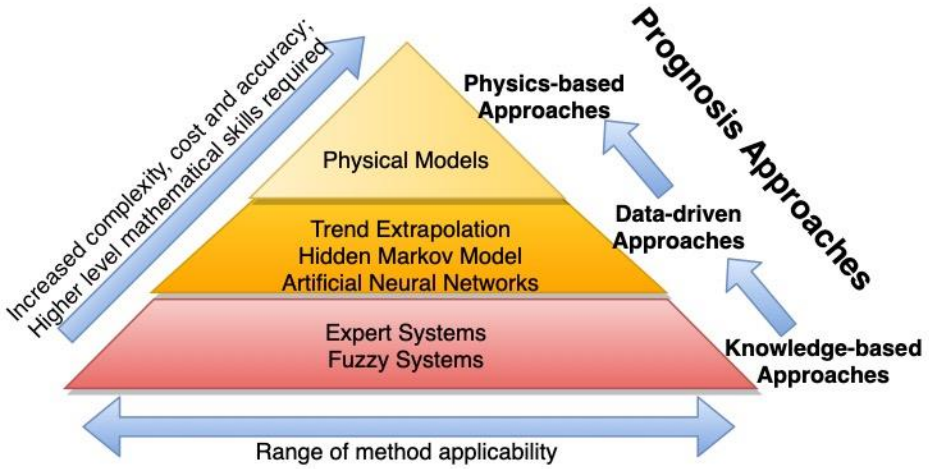

Fig. 2.2 Prognosis models hierarchy

2.2.1 Knowledge-Based Approaches

Knowledge-based approaches derive the machine's residual life by comparing the similarity between an observed condition with a databank of previously defined events. Expert systems and fuzzy systems are the main two methods within this category.

An expert system contains the accumulated professional knowledge in a specific area and formulates it as exact IF-THEN rules in a software program. An effective fault prognosis system was established to predict several typical faults of railway track circuits by developing a complete and accurate expert system [21]. Another expert system was combined with neural networks to predict the residual life distributions of gearboxes [57]. Although expert system methods are easy to understand and implement, the issue of 'combinatorial explosion' happens when the machine behaviour and the corresponding prediction results are described in great detail [57]. That is, the increased numbers of system inputs and the desired outputs cause the number of possible combinations to increase dramatically. One rule is required for each combination, which makes it difficult to check throughout and manually for consistency. It becomes even more difficult to implement on a real-world system, because all formulated rules need

to be updated by incorporating working environment related parameters as new inputs.

Fuzzy systems, serving as another type of knowledge-based approaches, can overcome the aforementioned combination problem in expert systems by allowing empirically derived, but intentionally inexact IF-THEN statements to address these problems. As such, one fuzzy rule can substitute a great number of traditional rules. Successful applications of fuzzy system for RUL prediction include a set of boiler tubes [58], plant equipment [59], and wind turbine pitch [60]. Similar to expert systems, a sufficient amount of expert knowledge of the underlying fault process is compulsory for fuzzy systems, which is not easy to achieve for real-world complex systems.

Note that some approaches introduced in the fault diagnosis sections (i.e., Sections 2.1.1 - 2.1.3) can also be used for prognosis, such as the aforementioned fuzzy systems, because fault prognosis is related to and depends highly on fault diagnosis. In addition to several common approaches that are applicable for both diagnosis and prognosis, the following subsections (i.e., Section 2.2.2-2.2.3) will also introduce some methods that are only suitable for prognosis.

2.2.2 Data-Driven Approaches

Data-driven approaches determine the machine's RUL based on a mathematical expression that has been obtained from observation data instead of from physical knowledge of the fault progression.

Trend extrapolation serves as the simplest approach in this category, which defines a monotonic, single-parameter fault progression trend. The variable parameter that is directly related to collected data is plotted against time, and the residual life is found when the parameter curve reaches a pre-established alarm threshold. Note that measurement noise and variations in working conditions could significantly affect prediction results in trend extrapolation. An improved exponential trend analysis was proposed for rolling element bearings, which utilizes particle filters to reduce the random errors of the stochastic process, so as to increase the prediction accuracy [61]. In addition to the exponential function, many other distributions can also be applied to model fault progression, including Normal, Lognormal and Weibull functions. A Weibull trend analysis was specifically presented for RUL predictions of bearings exposed to different operating conditions [62]. Normal functions have demonstrated their superiority in modelling gradual and monotonic deteriorations in bearings [63]. As a result of its superior ability to model complex systems and relatively

low requirements for physical knowledge, the trend extrapolation approach shows high potential to be applicable to the proposed railway switch problem.

Hidden Markov models (HMMs) can model multiple practical fault scenarios without the precondition of monotonic behaviour. It has been successfully applied to fault prediction in power transformers [64] and bearings [65]. More recent research makes full use of the characteristics of state-based analysis in HMMs, but improves their overall performance including computation complexity and result consistency. For example, a simple state-based prognostic model (SSBP) was proposed to forecast the residual life of a railway switch system, demonstrating an improved prediction accuracy compared with traditional HMMs [66]. Another technique called state-based prognostics with state duration information (SBPD) that incorporates the time duration spent in each transition state demonstrated its effectiveness for the same application [67]. Based on the literature that has been listed in previous sections (i.e., Sections 2.2.1 - 2.2.2), it can be found that fault prognosis methodologies have been widely developed for a variety of engineering applications, especially for rolling element bearings working in an idealized lab environment. Two of the few published studies that are directly related to railway switch fault prognosis [66, 67], which means that, in addition to the aforementioned trend extrapolation,

HMMs is another method type that is potentially applicable to the proposed problem.

Compared with the several types of artificial neural networks introduced in Section 2.1.2 for fault diagnosis, the types of applicable networks for fault forecasting are quite different. Three types of network structures have been explored for machine prognosis, which are recurrent neural networks, time delay neural networks, and long short-term memory neural networks. The common point of these three methods is that the temporal information that reflects the continuous changes of fault progression is 'stored' by the network in a particular form, in order to deal with time-variant forecasting problems. The term time-variant is used to describe a system property when a time advance or time delay in system input changes not only the system output in time, but also the system performance and other parameters. Some specific applications using the above three types of methods can be individually found for aircraft turbofan engine [68], for cracked aluminum sheet [69], and for a battery system [70]. By training a large number of data, artificial neural networks tend to obtain a good prediction accuracy. The characteristic of modelling complex systems without expert knowledge also makes it possible to be applicable to the proposed research problem.

2.2.3 Physics-Based Approaches

The final type of solution for machinery fault prognosis is physics-based approaches. This type of approach requires a comprehensive understanding of the system performance in response to fault progression, at both microscopic and macroscopic levels. This allows a quantitative description of the system performance based on first principles. For example, crack progression behaviours in different materials are commonly analysed using physical approaches [71] [72].

2.2.4 Challenges in Prognosis Modelling

The added value of this fault prognosis literature review is summarized in Table 2-2, which compares the advantages and disadvantages of prognosis approaches introduced in earlier sections (i.e., Sections 2.2.1-2.2.3). Their applicability for real-world railway switches is also discussed.

According to Table 2-2, the types of knowledge-based approaches and physics-based approaches both require detailed understanding of the fault mechanism for a specific application. In contrast, data-driven approaches are a generic type of model that can predict RUL of a complex system without relevant physical

knowledge. By learning automatically from a large amount of data, data-driven approaches tend to achieve satisfactory prognosis results. In terms of the proposed condition monitoring problem for real-world railway switches, comprehensive knowledge of system operations and external conditions are not easy to obtain. As such, data-driven approaches show the highest potential to be applied to railway switches and therefore will be further explored in this research. The next few chapters (i.e., Chapters 3, 4 and 5) are concerned with the topic of data-driven fault prognosis, with respect to data collection, method exploration, and result analysis processes, respectively.

Table 2-2: Summary of this fault prognosis literature review

Methodologies	Advantages	Disadvantages	Applicability for real-world railway switches
Knowledge-based Approaches:			
Expert Systems	Easy to understand and implement	Considerable rules and expert knowledge needed	A sufficient amount of expert knowledge of the underlying fault process is mandatory, which can be hard to obtain in terms of a real-world complex system.
Fuzzy Systems	Fewer rules required; model noisy, imprecise inputs	Require expert knowledge to set rules	
Data-driven Approaches:			
Trend Extrapolation	Suitable for monotonic, single-parameter trend	Affected by noise and various working conditions	Data-driven approaches show high potential to be applicable to real-world railway switches, because of their ability to model complex, non-linear systems with the lowest requirement for physical knowledge. They tend to achieve a good prognosis accuracy by training a large amount of data.
Hidden Markov Models	Can model noisy, non-monotonic fault scenarios	Large computation effort required	
Artificial Neural Networks	Can model complex scenarios without expert knowledge	A large amount of data is required for model training	
Physics-based Approaches:			
Physical models	Can achieve the highest prediction accuracy when compared with the other two method types	Require detailed physical understanding of the fault mechanism's behaviours	It is usually impractical to apply physical approaches to a real-world complex system due to a lack of comprehensive physical knowledge of system behaviours.

CHAPTER 3

EXPERIMENTAL SETUP & DATA COLLECTION

This chapter first introduces the increasing requirement for railway transportation availability expected to be delivered by improved asset management and condition monitoring strategies, so as to accommodate and serve more efficiently the increasing number of rail passengers in Great Britain. A railway switch system has been chosen as the experimental basis for validation of the condition monitoring algorithms proposed in this research, because of the significant role it plays in traffic flow by shifting the train's moving direction and ensuring safe transfer between tracks. Subsequently, it is presented in terms of its physical structure and the sensor measurements used for monitoring its health status. Finally, at the end of this chapter, the data collection process is described. Two sets of data (named Dataset 1 and Dataset 2) are individually collected for diagnosis and prognosis, respectively. More detailed explanation of

the data type, data volume, and precautions when collecting the data is provided in the following subsections (i.e., Section 3.2.1-3.2.2).

3.1 Railways and Switch System

3.1.1 Railway Transportation

Railway transportation has become increasingly in demand in recent years, because of the ability to efficiently and cost effectively move freight and passengers whilst avoiding the congestion inherent with road traffic. However, as the number of passengers increases year by year, the requirements for railway operation and asset management have also increased. Condition monitoring, serving as a crucial stage in railway asset management, has great significance in enhancing the reliability and availability of the whole system. Common assets in an operational railway include tracks, signals, stations, bridges, tunnels, switches and level crossings. Among these assets, electrically powered assets, such as railway switches, have caused an increased number of incidents during 2019-20 across Great Britain, with more faults influences normal operations [1]. From this point of view, an improved management and condition monitoring over these asset components is urgently needed, in respect of the optimization of inspection, maintenance and renewal strategies, the extension

of long-term usage cycles, and the reduction of management costs over the long-term.

The railway switch is a key asset category that diverts a train from one line to another, as shown in Fig. 3.1 [73]. At the same time, switches are also a priority risk area for railway operations because they create the potential for serious accidents. Due to the important role the switch plays in changing a train's moving direction and ensuring safe transfer between tracks, an efficient management in respect of inspection, maintenance, and renewal can positively affect railway operations whilst reducing the risks of serious incidents [74].

Fig. 3.1 An operational railway switch [75]

Preventive and predictive maintenance techniques for railway switches have the potential to reduce asset faults and failures, leading to a reduction in costs and

disturbances caused by unplanned maintenance interventions. Therefore, the deployment of condition-based maintenance on railway switch systems shows great importance in enhancing the dependability of an overall railway network.

3.1.2 Switch Instrumentation and Measurement

Trains can move between tracks with the help of a railway switch that moves the rails to the needed position before they pass. Since being patented in 1832, the mechanical structure of railway switches has remained virtually unchanged. However, different types of point operating equipment (POE) have been used, and the relevant operation and control strategies have been developed throughout the decades [76].

A schematic of a railway switch system is shown in Fig. 3.2, and a corresponding practical system diagram is shown Fig. 3.3. It can be seen that a switch is a complicated piece of electromechanical equipment consisting of stock rails, switch rails, stretcher bar, slide chairs, drive and detection rods. The operation of a railway switch can be divided into four phases: inrush, unlock, move and lock. Switch rails are locked in position after each movement; the first two phases of a switch operation are to increase the electric power supply to the switch motor and unlock it from its current position, respectively. Subsequently,

the drive rod that is connected to an actuator will move the switch rails to the correct position and lock them.

Generally speaking, the key features of POE are actuation, locking and detection. Actuation delivers the mechanical force to move the switch rail from one place to another. Locking of the switch rail ensures that it maintains in its final position after actuating successfully. Detection provides confirmation to the signalling system that the position of the switch is safe for train passages. In practice, there are many types of POE. The main categories are external or in-bearer, and single or multiple machine [76]. The type of POE applied in this research is an external single switch, which describes an external actuator box mounted on a trackside frame that is fixed to the bearers, and a single machine sufficient for switch actuation.

The railway switch is a piece of reciprocal equipment that can be moved in either of two directions (normal-to-reverse and reverse-to-normal). The actuator is also called a point machine (shown in Fig. 3.2) and can be categorized into electric (DC or AC), hydraulic and pneumatic based on its power supply type. The overall operation duration is only a few seconds and the data to be used in condition-based monitoring are collected during this period. The size of data sets

depends on the time duration of each switch movement, as well as the frequency of the movements.

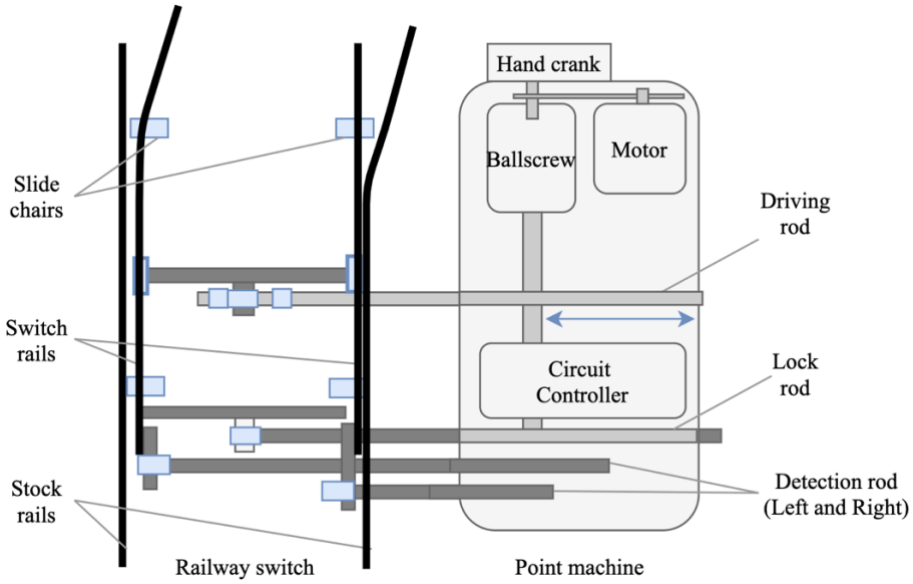

Fig. 3.2 Schematic of a railway switch system adjusted from [77]

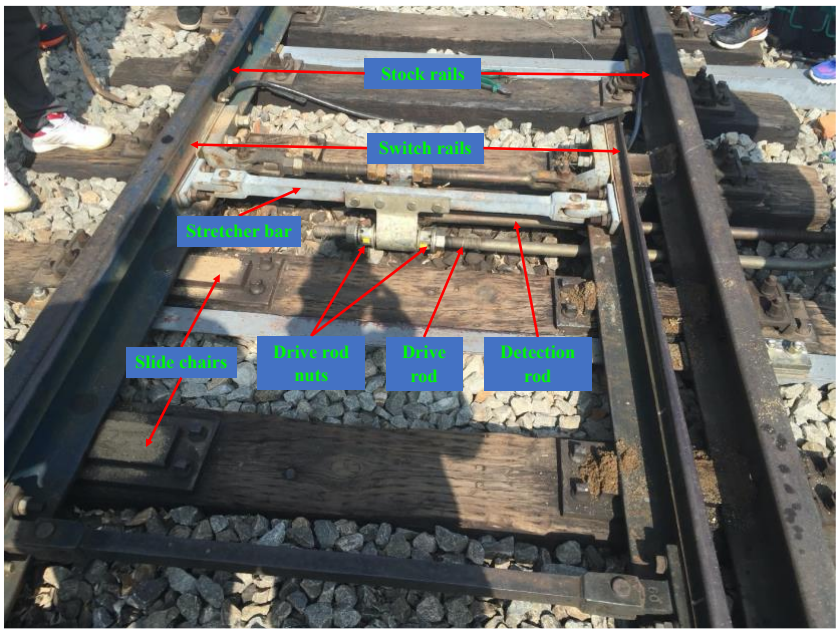

Fig. 3.3 A practical switch system

Before starting the data collection process for the purposes of validating the algorithms proposed in this work, an analysis of the suitability of various sensors

under different failure modes has been conducted. It was found that a large proportion of service affecting failures are caused by the 'Out of Adjustment' fault type. Failures attributed to the mechanical drive rods ranked within the top three between 2015 and 17 [76]. Most failure modes cause a more difficult drive rod movement, resulting in an increased required force, change in motor voltage and current to supply the required power, and a reduced moving speed.

While considering the approximate cost, reliability, robustness, and fault sensitivity, as listed in Table 3-1, the force sensor, current sensor and voltage sensor were chosen to be applied in this study. The proximity sensor that detects the distance of movement of the switch rail is insufficiently robust for practical applications.

Table 3-1: Comparison of sensors for railway switch application [78]

Sensor Type	Approx. Cost (£)	Reliability	Robustness	Sensitivity to faults
Force	350	Reliable	High	High
Current	25	Reliable	High	Moderate
Voltage	45	Reliable	Moderate	Moderate
Proximity	15	Reliable for short distance	Low	Low

Fig. 3.4 shows an experimental setup for data collection from different sensors in a DC point machine. The point machine shown in Fig. 3.4 is of ZD6 type, which is powered by a DC motor. It can be seen that there are many inner components within a point machine, such as the motor, gearbox, circuit controller, and the drive and detection rods that are connected to the switch rails shown in Fig. 3.3. A load pin is installed as the force sensor applied to monitor the load cells within the drive rod during the switch movement. The force sensor provides a range of -50,000 N to +50,000 N analogue output measuring the force in compression and tension. A DC current sensor installed in the circuit controller is used to measure the changes in motor current. The current sensor outputs an analogue current value in the range of -10A to +10A. Similarly, a DC voltage sensor in the circuit

controller is applied to measure the changes in motor voltage during its operation. The voltage sensor is capable of measuring in the range of -250V to +250V.

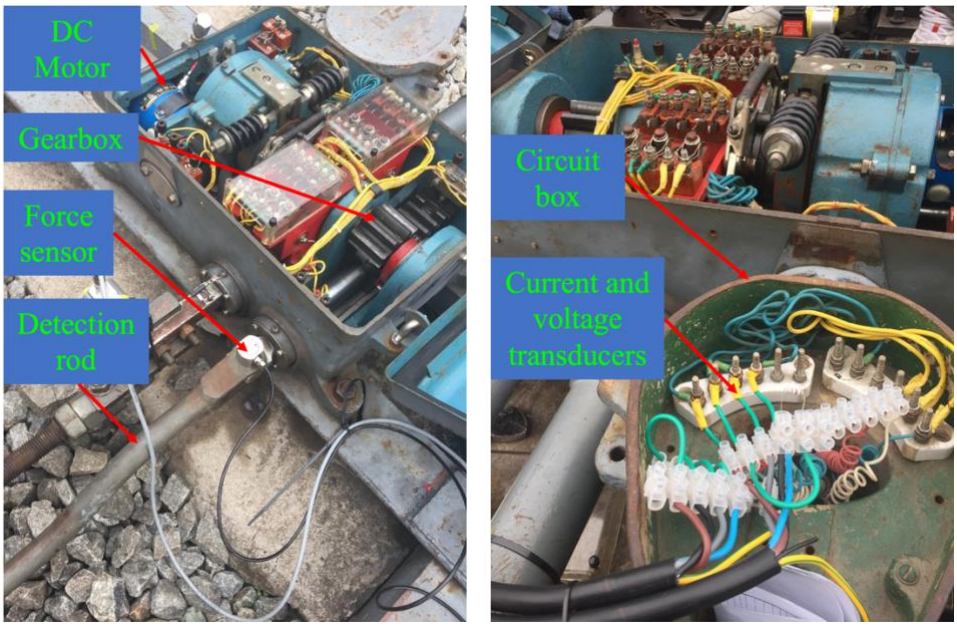

Fig. 3.4 Experimental setup for data collection from a DC point machine

3.2 Data Collection

In order to validate the proposed condition monitoring algorithms, a group of in-service switch systems were selected from which to collect sensor data. An embedded data acquisition system was designed, and all the sensor data were collected and transmitted to a remote PC via a LABVIEW program. Note that the switch machines instrumented were from an operational system, rather than an idealized laboratory instrument. Specifically, in terms of the data collection process, more considerations are present in the case of practical railway

switches. For example, the data collection process must not affect normal railway operations, and the safety of staff carrying them out must be ensured. In contrast, an instrument working in the laboratory environment is free from those limitations and the data collection is rather straightforward to implement. More specific explanation will be given for the scenarios for both diagnosis and prognosis in the following subsections (i.e., Section 3.2.1-3.2.2).

3.2.1 Dataset 1 for Diagnosis

In this research, a drive rod overdriving failure mode with varying severities is considered, which corresponds to 9% of switch failures in Great Britain [79]. Overdriving is a state in which the force between the switch rail and the stock rail is beyond the ideal range. Considering that the development of a natural overdriving failure in a switch system may take years, overdriving conditions with incremental severities were simulated to obtain statistically sufficient data sets.

As shown in Fig. 3.5, the switch rails are moved by the lateral driving force of the driving rod. The driving force actuates the switch rail through the locking nut and the arm fixed on the switch rails. When manually fastening the locking nut on the right-hand side of the arm, it generates a lateral offset to the left via the screw thread. As such, the original distance between the right locking nut and

the arm is shortened. When the driving rod is actuated to move laterally from right to left, the right locking nut touches the arm sooner due to the shortened travel distance. The switch rails linked to the arm are then actuated and transit to the locking position earlier, while the point machine is still producing the driving force. The redundant actuation power will cause a greater force between the switch rail and the stock rail, and overdriving occurs.

Specifically, one hexagonal screw nut has six faces. By fastening the hexagonal screw nut mounted on the drive rod one side (one face) at a time, one level of overdriving severity can be simulated.

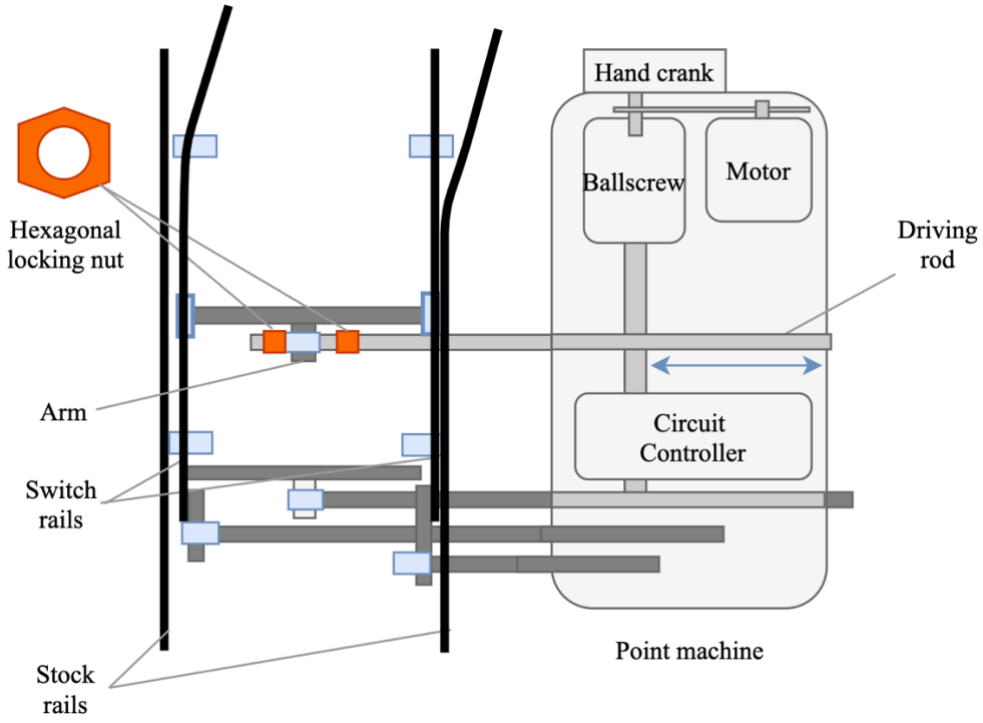

Fig. 3.5 Principles of the simulation of overdriving fault

A total of eight overdriving severities, which are ‘fault free’, ‘overdriving 3 faces’, ‘overdriving 6 faces’, ‘overdriving 7 faces’, ‘overdriving 8 faces’, ’overdriving 9 faces’, ‘overdriving 10 faces’, ‘overdriving 11 faces’ can then be sequentially simulated and obtained. Specifically, 20 sets of data, with each composed of force, electrical current and voltage signals, were collected for each overdriving severity. As such, a total of 160 data sets were obtained during the switch normal-to-reverse movements.

3.2.2 Dataset 2 for Prognosis

Unlike Dataset 1 that was collected separately from discrete faulty conditions, Dataset 2 was collected during a continuous period. By monitoring and analysing the machine condition over a continuous operational period, a prediction of remaining useful life can potentially be achieved. After comparing the cost and fault sensitivity, as shown earlier in Table 3-1, the electrical current sensors were chosen to collect prognosis-related condition monitoring in-service data due to the characteristics of low-cost, high reliability, high robustness, and high sensitivity to faults. From January 2018 to February 2019, 50 groups of current data were collected from 50 electro-hydraulic switch machines, with each data group recording a specific switch’s every movement. Note that the railway switches are a type of practical operational system that must not be permitted

to run to a failure state, wherever possible. In other words, the collected 50 groups of data contain the progression of faults rather than failures. As introduced in Section 1.1.2, the difference between fault and failure is that the machine with a fault can still achieve what is supposed to, but in a degraded mode, while the machine with a failure means the incapability to complete what is supposed to.

A summary of applied point machine type, sensor type, applicable scenario and data description is demonstrated in Table 3-2 for dataset 1 and dataset 2. In the next chapter, the methodology of fault diagnosis and prognosis will be proposed, followed by results analysis of implementing the methodology towards the two datasets.

Table 3-2: A summary of dataset 1 and dataset 2

	POE Category	Actuation Type	Sensors Used	Applicable Scenario	Data Description
Dataset 1	Style-63 point machines from a Chinese railway	Electro-mechanical	Force, current, voltage	Overdriving fault diagnosis	160 sets collected separately from 8 discrete fault severities
Dataset 2	Clamplock point machines from a British railway	Electro-hydraulic	Current	Overdriving fault prognosis	50 groups of degradation data collected over a continuous operational period

CHAPTER 4

METHODOLOGY

In order to begin addressing the research challenges identified in Section 1.1, the condition monitoring methodologies developed and presented in this chapter are focused on the application of data science signal processing techniques to the sensor information from previously conducted experiments. Therefore, this chapter is organised around two topics: 1) AI-based diagnosis modelling, and 2) data-driven prognosis modelling. These were selected based on the outcomes of Chapter 2.

4.1 AI Diagnosis Modelling

4.1.1 Challenges in State of the Art

Fault diagnosis is concerned with the detection of machine abnormalities and identification of the failure mode, by analysing a machine's health monitoring data. As discussed in Chapter 2, AI-based diagnosis methods are found to be potentially applicable to railway switches owing to their ability to model practical

and complex problems. The following research will be developed around the topic of AI-based diagnosis models.

Within the area of artificial intelligence, machine fault diagnosis can be treated as a classification problem [80]. Specifically, a diagnosis model is established by feeding pre-processed sample datasets into a defined network architecture, which is normally composed of layers of data processing units that are capable of mapping the complex non-linear relationship between the network input and output [80]. The sample datasets are called training datasets, each of which is composed of a pair of input data and the corresponding expected output data points [14]. After training using training datasets, an inferred function for mapping the newly encountered data (i.e., test data) is generated and the fault category for these test samples can be determined [14].

Feature extraction plays an extremely important role in solving problems of classification, the quality of which can dramatically affect the overall diagnosis performance [81]. Feature extraction begins with an initial measurement dataset and generates derived values (features) expected to be non-redundant and informative. It is regarded as a dimension reduction process because the

initial measurement datasets are shrunk to a more manageable size, facilitating the subsequent model training and testing procedures [81].

When performing analysis of a dataset, extracting the features from the time domain is the simplest and most intuitive, since the analysis is directly based on the time waveform itself. Common time-domain features include mean, variance, root mean square, kurtosis and skewness. Qualitative trend analysis (QTA) is one common technique to deal with time-domain datasets, the waveforms of which are partitioned and assigned distinct shapes. Specifically, a group of datasets collected in transmission lines were transformed using QTA into a sequence of episodes representing qualitative and quantitative properties, before feeding into Bayesian classifiers [82]. Also, a variety of time-domain features can be extracted and combined for a comprehensive assessment of the machine faults, as seen in [83].

Compared with the aforementioned time-domain features, time-frequency features have gained more attention in recent years because they can more directly demonstrate the changes in signal frequency components with time. A comprehensive review of current time-frequency methods is presented in [84], which mainly covers pseudo-Wigner-Ville distributions, short-time Fourier

transforms and wavelet transforms. Among them, wavelet transforms show the highest capability owing to their characteristics of concise mathematical expressions and robustness, with successful applications to fault diagnosis of a variety of engineering equipment such as rolling element bearings [85] and gears [86].

A conventional discrete wavelet transform (C-DWT) was used for fault diagnosis of railway switch systems, which derives the scaling coefficients at an appropriate decomposition level as feature inputs for a diagnosis model [87]. However, the disadvantages of poor noise sensitivity and limited applicability to sensors make it quite difficult to implement in practical situations. In contrast, the soft-thresholding discrete wavelet transform (ST-DWT) can effectively reduce noise within sampled datasets by applying the soft-thresholding method to wavelet coefficients at a certain decomposition level [88, 89]. However, the characteristics of poor adaptability may result in an inefficient feature extraction when applied to a specific industrial application, because the computed threshold in ST-DWT is a non-adjustable parameter with lack of a physical meaning.

It can be seen that the two aforementioned traditional wavelet-based methods (i.e., C-DWT and ST-DWT) are not comprehensive measurement approaches. The application of the inherent energy conservation characteristics within a wavelet transform could potentially make the feature extraction process more efficient, as will be seen in the newly proposed energy-based thresholding wavelets (EBTW) methodology in the following Section 4.1.2.

4.1.2 Proposed Diagnosis Methodology: Energy-Based Thresholding Wavelets (EBTW) and Neural Networks

The proposed diagnosis methodology is summarised in Fig. 4.1. First, fault-free and faulty signals are collected from a data acquisition system. Subsequently, the features are extracted from the original signals, using the proposed EBTW method in three steps. These features are then treated as the input to a 2-layer convolutional neural network (CNN) for model training and testing. The eventual diagnosis performance depends on the proportion of the test data that are correctly classified between fault-free and three faulty levels.

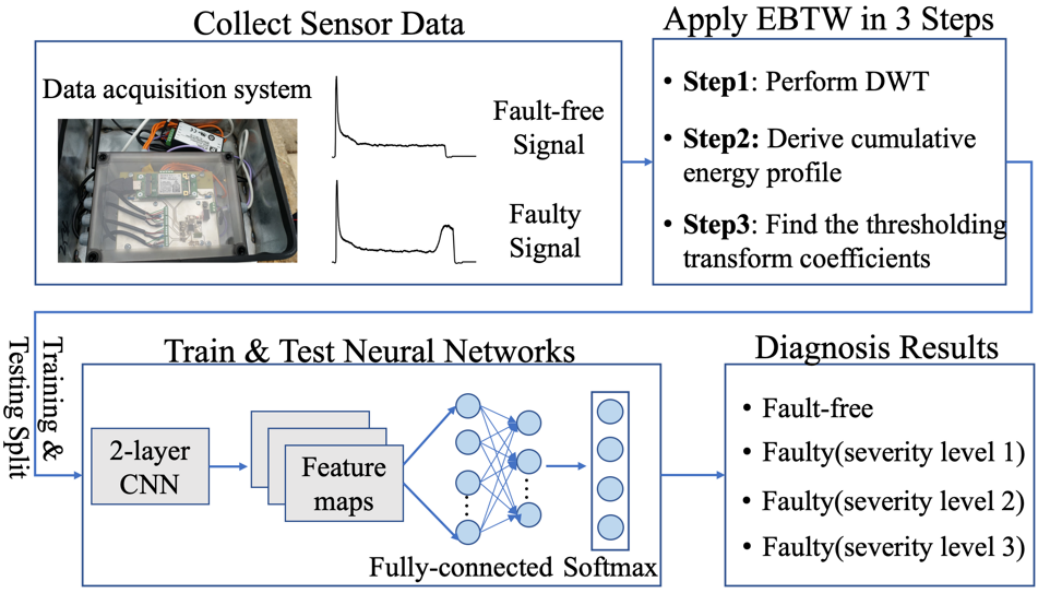

Fig. 4.1 Flowchart of the EBTW diagnosis methodology

4.1.2.1 Feature Extraction

The definition of a wavelet transform, as well as the property of energy conservation is first explained in this subsection. Subsequently, the details of the proposed EBTW methodology are described in three individual steps.

The continuous wavelet transform of a signal $f(t)$ is calculated by implementing a convolution of the signal itself with a scaled and translated wavelet basis function [90]:

$$F(a,b) = \int_{-\infty}^{\infty} f(t)\psi^{*}_{(a,b)}(t)dt \tag{4.1}$$

where $\psi^*_{(a,b)}(t)$ represents the complex conjugate (*) of a wavelet basis function $\psi(t)$ at scale a and translation b. The specific representation of a wavelet basis function is shown as [90]:

$$\psi_{a,b}(t) = a^{-1/2}\psi\left(\frac{t-b}{a}\right) \tag{4.2}$$

The scale parameter a is used to adjust the signal resolution. An increasing value of a will generate a more stretched-out signal with fewer details but lasts the increasing duration in time domain. The translation parameter b is to adjust the location of the wavelet basis function across the time length of the signal. By dilating and translating the wavelet basis function $\psi_{(a,b)}$, a series of daughter wavelets are generated and form a wavelet family. For the purpose of processing efficiently by computers, discretization is required to take discrete samples from the original continuous signals. A discrete wavelet transform (DWT) is introduced for discrete-time signal processing, which states [90]:

$$\psi_{m,n}(t) = 2^{-m/2}\psi\left(\frac{t-n\cdot 2^m}{2^m}\right). \tag{4.3}$$

$a = 2^m$ and $b = n \cdot 2^m$. $\psi_{m,n}(t)$ is a discretised representation of the wavelet basis function specified in Equation (4.2). The original signal $f(t)$ is then reconstructed as [90]:

$$f(t) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\psi_{m,n}(t)d_{m,n} \tag{4.4}$$

$$d_{m,n} = \int_{-\infty}^{\infty} f(t)\,\psi^*_{(m,n)}(t)dt \tag{4.5}$$

where $\{d_{m,n}\}$ are referred to as the wavelet coefficients. Another group of coefficients called scaling coefficients are generated from a wavelet basis function, $\varphi(t)$, that is orthogonal to $\psi(t)$ [90]:

$$s_{m,n} = \int_{-\infty}^{\infty} f(t)\, \varphi^{*}_{(m,n)}(t)dt \tag{4.6}$$

$$\varphi_{m,n}(t) = 2^{-m/2}\varphi\left(\frac{t - n \cdot 2^{m}}{2^{m}}\right) \tag{4.7}$$

where $\{s_{m,n}\}$ are the scaling coefficients at mth order. The wavelet basis function $\varphi(t)$ is also called the scaling function that achieves the approximations of the original signal $f(t)$ in multiple resolutions. Additionally, the wavelet coefficients and scaling coefficients can be calculated iteratively for a higher $(m+1)$th order by decomposing the scaling coefficient at the mth order [90]:

$$d_{m+1,n} = \sum_{j=-\infty}^{\infty} h(j-2n)s_{m,n} \tag{4.8}$$

$$s_{m+1,n} = \sum_{j=-\infty}^{\infty} l(j-2n)s_{m,n} \tag{4.9}$$

where h and l individually represent the high-pass and low-pass filter coefficients that are both obtained from the wavelet basis $\psi(t)$. It can be seen that the wavelet coefficients $\{d_{m,n}\}$ represent the fluctuation or running difference within the higher-frequency components. The scaling coefficients $\{s_{m,n}\}$ represent the trend or running average within the lower-frequency

components of the original signal. The lengths of the wavelet coefficients and scaling coefficients are both half of the original.

In comparison with the two traditional wavelet-based methodologies (i.e., C-DWT and ST-DWT) that were introduced in Section 4.1.1, the newly proposed EBTW method could potentially achieve a more efficient feature extraction by applying the inherent energy conservation characteristics within a wavelet transform.

An m-level wavelet transform of the signal $\boldsymbol{x}$ is represented as:

$$\boldsymbol{x} \longmapsto (\boldsymbol{s}^m|\boldsymbol{d}^m|\boldsymbol{d}^{m-1}|\cdots|\boldsymbol{d}^1). \tag{4.10}$$

$\boldsymbol{x}$ is a N- length signal represented as:

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_N). \tag{4.11}$$

Thus, the energy ε of signal $\boldsymbol{x}$ can be calculated as:

$$\varepsilon_{\boldsymbol{x}} = x_1^2 + x_2^2 + \cdots + x_N^2. \tag{4.12}$$

According to the energy conservation property, the energy among the transform coefficients is equal to the energy within the original signal. This yields:

$$\varepsilon_{(\boldsymbol{s}^m|\boldsymbol{d}^m|\boldsymbol{d}^{m-1}|\cdots|\boldsymbol{d}^1)} = \varepsilon_{\boldsymbol{x}}. \tag{4.13}$$

The newly proposed energy-based thresholding wavelet (EBTW) methodology takes advantage of the energy conservation property during the wavelet

transform process. By localizing and redistributing the energy within the original signal, EBTW could potentially achieve an efficient dimension reduction and feature extraction. A summary of its implementation steps is shown below.

Step 1: Perform a DWT on the collected sensor data with a certain wavelet type. In order to investigate the best performance, comprehensive experimental results from using different wavelet functions, such as the near symmetric 'Coiflet' wavelet, discrete 'Haar' wavelet and continuous 'Daubechies' wavelet, should be used and compared. The 'Haar' wavelet function that serves as the fundamental prototype for all other wavelet families is chosen as a demonstration case to study in this research.

Step 2: Obtain the cumulative energy map by rearranging the transform coefficients. The absolute values of the coefficients obtained from the DWT are sorted in descending order as:

$$e_1 \geq e_2 \geq \cdots \geq e_M \tag{4.14}$$

where e_1 and e_M signify the greatest and the smallest absolute values of the transform coefficients, respectively. As such, the cumulative energy map can be obtained by following:

$$\left(\frac{e_1^2}{\varepsilon_x}, \frac{e_1^2 + e_2^2}{\varepsilon_x}, \frac{e_1^2 + e_2^2 + e_3^2}{\varepsilon_x}, \cdots, 1\right) \tag{4.15}$$

Equation (4.15) tells that if the energy map quickly reaches its maximum value of unity, most of the signal energy is expected to be distributed on a small portion of coefficients with larger absolute values, which allows the following thresholding technique to become potentially efficient.

Step 3: Determine the energy percentage to be conserved within the signal, labelled as η, and calculate the corresponding wavelet threshold e_{th}. Specifically, the wavelet threshold e_{th} is calculated indirectly by assessing the first value of e_j that can make the energy percentage to be conserved exceed the selected value η:

$$\frac{\sum_{j=1}^{th-1} e_j^2}{\varepsilon_{\boldsymbol{x}}} \leq \eta \tag{4.16}$$

where:

$$\frac{e_1^2 + e_2^2 + e_3^2 + \cdots + e_{th-1}^2 + e_{th}^2}{\varepsilon_{\boldsymbol{x}}} = \eta \quad (0 \leq \eta \leq 1) \tag{4.17}$$

Subsequently, retain those coefficients that have greater absolute values than threshold e_{th} and assign all others to be zero:

$$e_j = \begin{cases} sgn(e_j)|e_j|, & |e_j| \geq e_{th} \\ 0, & |e_j| < e_{th} \end{cases} \tag{4.18}$$

where $j = 1, 2, \cdots, M$ and $\mathrm{sgn}()$ is the sign function.

As such, the original transform coefficients obtained from a DWT have been rearranged and thresholded based on the analysis of the signal energy distribution. The surviving coefficients are regarded as the extracted feature components to feed into the following fault diagnosis algorithms. Note that the basis of this new thresholding method of extracting wavelet coefficients is to determine the value of the wavelet threshold. The main novelty of the proposed method is that the determination of the wavelet threshold is transformed to the determination of energy percentage to be conserved within the signal. This allows the relevant experts in a specified area to physically understand and adjust the feature components generated from the proposed method. The general principle of determining a proper energy conservation percentage is to find its lowest value that can still retain an acceptable level of waveform resolution. A high energy conservation percentage ensures the waveform will be reconstructed to a high resolution, but requires a large computation time. On the other hand, a low energy percentage brings about an efficient dimension reduction, at the expense of waveform resolution and diagnosis accuracy. Therefore, there always exists a trade-off between retained signal resolution and computation time.

In order to validate the proposed EBTW method and demonstrate its enhanced capability over existing methods, a comparison between EBTW and two mature methods (i.e., C-DWT and ST-DWT) is implemented under various classifiers, and will be shown in the following Section 4.1.2.2.

4.1.2.2 Diagnosis and Classification

The fast development of neural networks and deep learning over the past two decades has seen their increasing application in the area of fault diagnosis. As a type of machine learning, deep learning methods have an extended architecture with several hidden layers of processing units. This allows the network to automatically learn the complex linear or non-linear relationship between input and expected output, so as to correctly recognize the fault patterns. Deep learning can usually realize a rather satisfactory diagnosis accuracy in comparison with some traditional classifiers.

The application of a generative adversarial network (GAN) to solving the fault diagnosis problem within analogue circuits has been explored and validated [91]. Another deep learning algorithm called the convolutional neural network (CNN) was applied to address gearbox fault diagnosis problems by learning and classifying frequency-domain features extracted from vibration signals [92]. A

comprehensive review of AI-based machine health management methodologies discusses the recent trend and potential restrictions in deep learning diagnosis [93]. However, the adaptive ability of these algorithms for a real-world engineering application still needs to be further investigated using field data, instead of idealized lab-collected data.

A convolutional neural network (CNN) that serves as an efficient and powerful deep learning approach for solving fault classification problems is introduced here in detail. A CNN is a type of feed-forward neural network, which consists of four different layer types, named the convolutional layer, pooling layer, fully connected layer, and softmax layer. First of all, the original datasets or the extracted features are treated as the input to the convolutional layer, which calculates and adds the convolution results of each input with various kernels. Each kernel is composed of one bias vector and one weight matrix. The summation results are then transformed using a non-linear activation function. A pooling layer always follows every convolutional layer, the aim of which is to down sample and reduce data dimension. The fully connected layer works the same way as in a regular neural network, connecting all its neurons to every output from the former layer. The softmax layer is usually the final layer used to

calculate the probabilities of every input being classified in every category and decide the predicted category for the given input.

To seek the model type that can achieve the best diagnosis performance, a comparison is conducted for a variety of machine learning algorithms of both modern and conventional approach types. Modern neural networks include the aforementioned CNNs, shallow neural networks and deep neural networks. Conventional machine learning algorithms including k-nearest neighbours (kNNs) and support vector machines (SVMs) are also applied in this research for measuring diagnosis accuracy and efficiency.

The architecture and related parameters of each classifier need to be adjusted for the best outcomes. Specifically, ten neighbours and a cosine distance matrix are adopted for the kNN in this research. A Gaussian kernel function is applied to the SVM. A shallow neural network (shallow NN) with one hidden layer of two neurons, a deep neural network (deep NN) with three hidden layers separately containing one neuron, two neurons and two neurons, as well as a one-dimensional CNN (1-D CNN) with two layers are applied in this research.

4.2 Data-Driven Prognosis Modelling

4.2.1 Challenges in State of the Art

Fault prognosis is used to determine the future health condition of a current machine of interest, by means of analysing the continuous changes within the machine health descriptor from the present to a pre-defined fault threshold. As discussed previously in Chapter 2, data-driven prognosis approaches are considered to have the highest potential for application to railway switch prognosis. The proposed Remaining Maintenance-free Operating Period (RMFOP)-based prognosis methodology is established upon the traditional data-driven models.

In terms of modern asset management regulations in Great Britain's (GB) railway industry, prognostics and health management (PHM) is normally carried out to maintain and manage the reliability of the machines that are affected by faults and failures. With the increasing requirement for a more efficient condition monitoring and asset management, prognosis information that is mainly obtained from staff knowledge and experience cannot achieve the targets of prognosis for high infrastructure safety and dependability. Expert practitioner knowledge is difficult to accumulate and can be imprecise, especially for integrated systems exposed to complicated working conditions.

At the same time, a large number of condition monitoring models have been explored in recent years to predict the asset condition [94-96], but there are still very few practical implementations in industry as a result of data shortages and comprehensive methodologies. The problem of insufficient data is caused by the fact that operational systems are not allowed to run to failure and faults are normally fixed before progressing to serious functional failures. This poses a serious challenge of collecting the complete run-to-failure degradation data that are required by most prognosis methods. In addition to the problem of data collection, those well-developed methodologies usually contain specific mathematical simplifications and assumptions that are very difficult to hold for real industry implementations. The two aforementioned problems would both increase the level of risk for a practical industry to train particular prediction models.

A new prognosis methodology is designed to collect and process degradation data in terms of the operational systems in real world, and presented in the following Section 4.2.2. The statistical properties of an individual complex system among a population of the same machine type, and its residual life

distributions will be explored in the proposed method. Again, the methodology's effectiveness is fully assessed using railway switch applications.

4.2.2 Proposed Prognosis Methodology: Remaining Maintenance-Free Operating Period (RMFOP)-Based Degradation Modelling

A new perspective on the traditional prognosis definition is put forward, with respect to the remaining maintenance free operating period. Next, two regression models, which are the exponential model and the linear model, are described and applied. An explicit summary of the methodology is then presented.

4.2.2.1 Prognosis Definition and Data Preparation

In order to address the challenge of a shortage of proper data, three types of approach are proposed in the published literature. The first approach is to artificially create a group of continuous degradation paths by fitting discrete data collected from levels of fault severities into a continuous model. Many time-series processing techniques, such as autoregressive moving average and long short-term memory network, can subsequently be applied for residual life predictions. However, it becomes quite challenging to validate whether the selected models and parameters follow the natural degradation patterns, since

the continuous data were generated unnaturally without properly validating its similarity to real-world degradation.

In the second approach, a piece of vulnerable equipment is built to substitute original ones for experimental data collection. Although the complete lifecycle data is easier to collect in this means of accelerated failure progression, it is less likely to be applicable for railway switches with complex electromechanical structures.

The third solution is to build mathematical models that can simulate the real-life physical degradation process. With respect to a railway switch system, both its inner components, such as the static rails, movable rails, and the switch motor, as well as the outer environment, such as train loads and extreme weather conditions, should all be considered. Compared with the two aforementioned methods, it generates continuous data waveforms following the natural propagation. Significant expert knowledge about the system behaviour is a pre-requisite for their development.

The problem of data shortages is addressed in this methodology, by considering a different time point at which to achieve a forecast. Unlike in most published

research, which predicts the time from which a machine cannot achieve its expected behaviour as a result of a functional failure, a new time period called the remaining maintenance-free operating period (RMFOP) is proposed here. It defines the residual time gap of a machine that operates normally without maintenance intervention activities. Therefore, the prediction results will be able to show the remaining suggested operating duration. As such, obtaining a sufficient amount of prognosis data is no longer hard to achieve because the data during normal operations is much easier to collect when compared to full run-to-failure data. Note that fault prognosis rather than failure prognosis is usually a more reasonable consideration for real-world systems since most of them must not be permitted to run to failure, wherever possible.

4.2.2.2 Model Selection

As discussed in Section 2.2.1-2.2.4 of the fault prognosis literature review, data-driven approaches are demonstrated to have the highest potential to be applicable to railway switches. Common data-driven methods include trend extrapolation, hidden Markov models, and dynamic neural networks. Among them, trend extrapolation models are introduced and employed in this study for two reasons. First of all, the required regression analysis within the trend

extrapolation is simple. Second, different regression functions can be chosen from to describe the fault progression trend.

In this research, the linear model and the exponential model are applied to describe the monotonically increasing, incipient fault progression. Regression models are concerned with the establishment of parametric progression paths considering the random effects within the collected condition monitoring data. A threshold needs to be set to indicate the occurrence of an asset fault when the monitored data exceeds the threshold. It is assumed that the assets from a population of the same type share identical degradation patterns. However, due to the varying usage frequency and the environment, the behaviour of the population merely provides some references and cannot accurately indicate the health condition of each individual asset. The following two models take advantage of Bayesian theory. The stochastic distributions of model parameters derived from a known group of progression paths will be later updated for each individual monitored asset.

Model 1: Linear model

A linear model can be mathematically represented as [97]:

$$y(t_i) = \theta t_i + c + \varepsilon(t_i) \tag{4.19}$$

where $y(t_i)$ signifies the degradation signal at time t_i. On the right-hand side of Equation (4.19), θ is a Gaussian distributed random coefficient with a prior mean μ_θ and a prior variance σ_θ^2. Equivalently, $\pi(\theta) \sim N(\mu_\theta, \sigma_\theta^2)$. c is a constant indicating the machine initial state. $\varepsilon(t_i)$ represents the noise within the signal, which is assumed to follow independent and identically distributed (i.i.d.) $N(0, \sigma^2)$. Bayesian theory is then applied to update the distribution of θ. This yields [97]:

$$p(\theta|y_1, \cdots, y_k) = \pi(\theta)p(y_1, \cdots, y_k|\theta) \tag{4.20}$$

where y_k is the degradation signal at time $t(k)$, or equivalently, $y_k = y(t_k)$. Equation (4.20) shows that by observing the continuous degradation signals from initial time t_1 until the current time t_k, the posterior probability of variable θ can be obtained. Subsequently, the posterior mean $\tilde{\mu}_\theta$ and posterior variance $\tilde{\sigma}_\theta^2$ can be separately calculated as [97]:

$$\tilde{\mu}_\theta = \frac{sum2 \cdot \sigma_\theta^2 + \mu_\theta \sigma^2}{sum1 \cdot \sigma_\theta^2 + \sigma^2} \tag{4.21}$$

$$\tilde{\sigma}_\theta^2 = \frac{\sigma_\theta^2 \sigma^2}{sum1 \cdot \sigma_\theta^2 + \sigma^2} \tag{4.22}$$

where $sum1 = \sum_{j=1}^{k}(t_j^2)$ and $sum2 = \sum_{j=1}^{k}\{(y_j - c)t_j\}$. Assuming that the continuous observations from the initial time to current time t_k have been obtained, the probability of residual life T_R not exceeding the time t is equal

to the probability of the degradation signal being greater than the threshold D at a future time t [97]:

$$\begin{aligned} & P(T_R \leq t | y_1, \cdots, y_k) \\ & = P(y(t_k + t) \geq D | y_1, \cdots, y_k) \\ & = \frac{\phi\left(\frac{c + \tilde{\mu}_\theta t - D}{\sqrt{\tilde{\sigma}_\theta^2 t^2 + \sigma^2}}\right) - \phi\left(\frac{c - D}{\sigma}\right)}{1 - \phi\left(\frac{c - D}{\sigma}\right)} \end{aligned} \tag{4.23}$$

where $\phi(\cdot)$ represents the cumulative distribution function (cdf) of a normal distribution. Eventually, in order to obtain the probability density function (pdf) of residual life, Equation (4.23) is differentiated with respect to t. A more detailed derivation process has been previously published [97].

Model 2: Exponential model

The exponential degradation model can be mathematically represented as [97]:

$$y(t_i) = c \cdot \exp\left(\theta t_i + \varepsilon(t_i) - \frac{\sigma^2}{2}\right) \tag{4.24}$$

where c is a constant. Variable θ follows a prior distribution $\pi(\theta)$ with a Gaussian mean μ_θ and variance σ_θ^2. $\varepsilon(t_i)$ represents the noise term following i.i.d. $N(0, \sigma^2)$. Usually, a logarithm is performed upon exponential equations to simplify calculations. After taking a logarithm on both sides on Equation (4.24), this generates [97]:

$$x_i = \left(lnc - \frac{\sigma^2}{2} \right) + \theta t_i + \varepsilon(t_i) \tag{4.25}$$

Similarly, Equation (4.20) is also applied here upon the logarithm amplitude x_i to update the posterior distributions of parameter θ. This generates [97]:

$$\tilde{\mu}_\theta = \frac{\sum_{j=1}^{k} \left\{ (y_j - lnc + \frac{\sigma^2}{2}) t_j \right\} \cdot \sigma_\theta^2 + \mu_\theta \sigma^2}{\sum_{j=1}^{k} (t_j^2) \cdot \sigma_\theta^2 + \sigma^2} \tag{4.26}$$

$$\tilde{\sigma}_\theta^2 = \frac{\sigma_\theta^2 \sigma^2}{\sum_{j=1}^{k} (t_j^2) \cdot \sigma_\theta^2 + \sigma^2} \tag{4.27}$$

Note that Gaussian distribution is assumed in both models above because of its capability of modelling a monotonic degradation process, which is suitable for the single incipient fault prognosis problem considered in this research. Other distribution functions including Weibull and lognormal can also be applied for modelling failure progression. The best choice of the distribution function relies on knowing the noise level and operating conditions.

4.2.2.3 Summary of the Methodology

A summary of the proposed RMFOP-based methodology is presented below in 10 individual steps:

Step 1: Collect N groups of condition monitoring data with each group from a specific machine of the same type. Each data group is a time series, describing the changes of a machine's sensor data over a continuous, maintenance-

included operating period. Note that the term “continuous” refers to the temporal relevance between adjacent data points, instead of the uninterrupted time in a narrow sense.

Step 2: Extract a proper type of “characteristic pattern” from the monitored data. Specifically, the data in each group is transformed into one specific characteristic pattern, which can directly illustrate a machine’s degradation progression. Therefore, the characteristic patterns are also called “degradation signals”. Common degradation signals are the minimum value, maximum value, mean, variance, slope, etc. The selected type of degradation signal will be validated in the experiment section: in the case of an undesirable prediction accuracy, another degradation signal will be considered for replacement. In addition, a maintenance threshold for the selected degradation signal must have been carefully selected by industry experts and can be referred to.

Step 3: Visualize the machine’s degradation process by extracting and plotting the degradation signals over the operating period for each data group. As stated in Section 4.2.2.1, the proposed methodology predicts the remaining period a machine can operate without maintenance interventions. The complete

degradation signals between two adjacent maintenance records are considered as one “degradation path” describing the process from fault-free to faulty.

Step 4: Normalize the degradation paths and apply to a degradation model, which is either linear or exponential. The linear model is suggested to be applied first for its simplicity. Specifically, M out of N ($M < N$) normalized degradation paths are treated as the prior knowledge to decide the values of the constant and the prior distribution of the stochastic parameter, as demonstrated in Equation (4.19). Among the M paths, fit each path with the linear equation and obtain the intersection and slope for each path. The average intersection value calculated among these M paths is assigned to constant c in Equation (4.19). The mean and variance of slope values are individually assigned to prior mean μ_θ and prior variance σ_θ^2 of parameter θ. The variance of error term σ^2 is obtained from a sequence of initial observations of the degradation path with the greatest fluctuation level. Normalization is a common pre-processing technique to re-scale data to be centred around zero and fluctuated within a unit.

Step 5: The prior distributions derived from the M paths are used as the population-wide properties known for the set of monitored machines. Among the residual $(N - M)$ machines, the RMFOP will be independently estimated

for each individual. For instance, the ith degradation path is selected from the $(N - M)$ paths. Once the observation duration ranges from the beginning to for example 10% of life accomplished, the parameter posterior distributions are updated specifically for the ith machine using Equation (4.21) and Equation (4.22). Equation (4.23) is then used to calculate the cdf and pdf of the ith machine's remaining operating period.

Step 6: Repeat Step 5 for the ith machine as more observation times become available. In this case, 30%, 50%, 70%, and 90% are used as the examples of life accomplished percentages. The updated RMFOP distributions for the ith machine under various life accomplished percentages can then be achieved. The life accomplished percentage is calculated as the ratio of elapsed degradation time until the observation time point to the total degradation time. For example, a machine degrading from fault-free to failure takes 100 days. 10% life accomplished percentage means the first 10 days of monitored data are collected for RMFOP predictions. Likewise, the numbers of 30%, 50%, 70%, and 90% separately represent using the first 30, 50, 70, and 90 days of monitored data for predictions.

Step 7: For each residual $(N - M - 1)$ machines, repeat Step 5 and Step 6. As such, the individual specific RMFOPs are estimated.

Step 8: Analyze all $(N - M)$ validation machines at different life accomplished percentages and plot the prediction errors with 95% confidence interval. A conclusion can be made about whether the linear model is suitable for the degradation path modelling.

Step 9: Select another model (i.e., exponential model) and work through Step 4 to Step 8. Unlike the statistics in the linear model that are directly calculated from the amplitude of degradation signals, the statistics in the exponential model are calculated from the logarithm amplitude of degradation signals.

Step 10: Generate models without the Bayesian updating technique in both the linear and exponential model cases and proceed through Step 4 to Step 9. A comparison between models with updating technique and models without updating technique can be used to further evaluate the effectiveness of the incorporated Bayesian theory. A final conclusion about the most appropriate degradation model is then made for the application.

As a summary, fault diagnosis and prognosis methodology has been presented throughout in this chapter. The challenges in state of the art were discussed. The proposed diagnosis and prognosis approaches were explained in theories. In the following Chapter 5, a proof of the concept of the proposed methodology will be designed and implemented to railway switch machines. The experimental results will be demonstrated and analysed.

CHAPTER 5

RESULTS

From the methodologies that have been proposed in Chapter 4, the experiments designed for proof of concept are implemented. The experimental results are demonstrated in this chapter. Section 5.1 is focused on the diagnosis methodology using EBTW and neural networks. Section 5.2 is concerned with the prognosis methodology using RMFOP-based degradation modelling.

5.1 AI-Based Diagnosis Methodology: EBTW and Neural Networks

In order to evaluate the diagnosis methodology proposed in Section 4.1, a group of railway switch systems operating on the Xiliang section of Guangzhou Metro in China were selected from which to collect sensor data. An embedded data acquisition system was developed. The sampling rate was set to be 20 kHz. 160 data sets were collected during the movements of the point machine in the normal-to-reverse direction. Specifically, real-world overdriving fault conditions are imitated by 8 simulated fault conditions with incrementally increasing

severities, which are termed “fault free”, “overdriving 3 faces”, “overdriving 6 faces”, “overdriving 7 faces”, “overdriving 8 faces”, “overdriving 9 faces”, “overdriving 10 faces”, and “overdriving 11 faces”. Twenty sets of data, with each consisted of current, voltage, and force signals, were obtained for each fault condition. In addition to the applied current, voltage and force sensors, a displacement sensor is also a common option to monitor railway POE health conditions. The reason that a displacement sensor is not available in this experiment is that the selection of suitable sensors is limited as a result of the harsh railway environment.

Before showing the detailed implementation results of the proposed methodology, the operational characteristics of a general POE system is described. Original measurement data collected from multiple sensors under both fault-free and faulty conditions are demonstrated and explained. Fig. 5.1 shows that under the fault-free condition, a sharply increased electrical current waveform is observed between 0 and 1.3 sec. This is due to the switch motor providing the mechanical drive to move the switch rails from one place to another. Subsequently, during the period 1.3-3.2 sec, a smooth current waveform is captured, which supplies a constant current value of approximate 0.7 A. After 3.2 seconds, an evident difference is observed between the fault-

free and faulty conditions. The waveform from the fault-free condition instantly drops to approximately zero current amplitude, while the waveform from the faulty condition shows a large current increase, indicating that more electrical power is required to lock the switch rail in its final position after successful actuation.

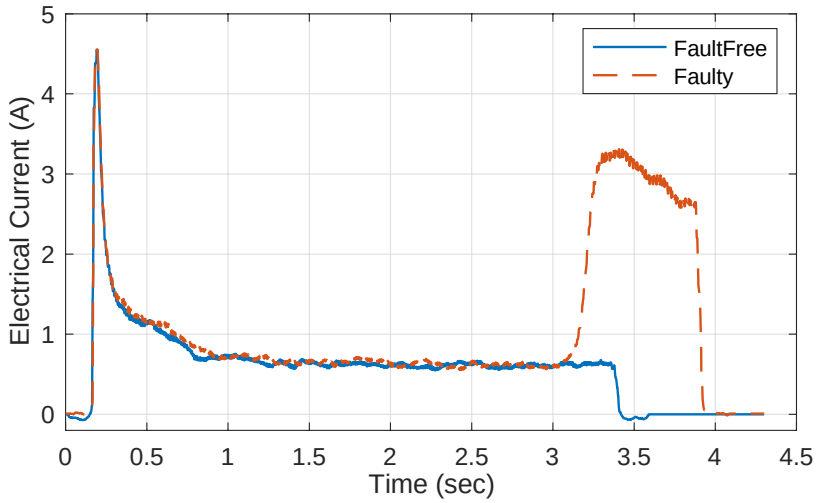

Fig. 5.1 Original electrical current measurements under fault-free and faulty conditions

Likewise, changes in motor voltage signals can be observed in Fig. 5.2. The simulated overdriving fault condition makes the movement of drive rods harder and therefore causes a reduced rod speed and an increased movement duration from 3.6 sec to 4.1 sec, as seen in Fig. 5.2.

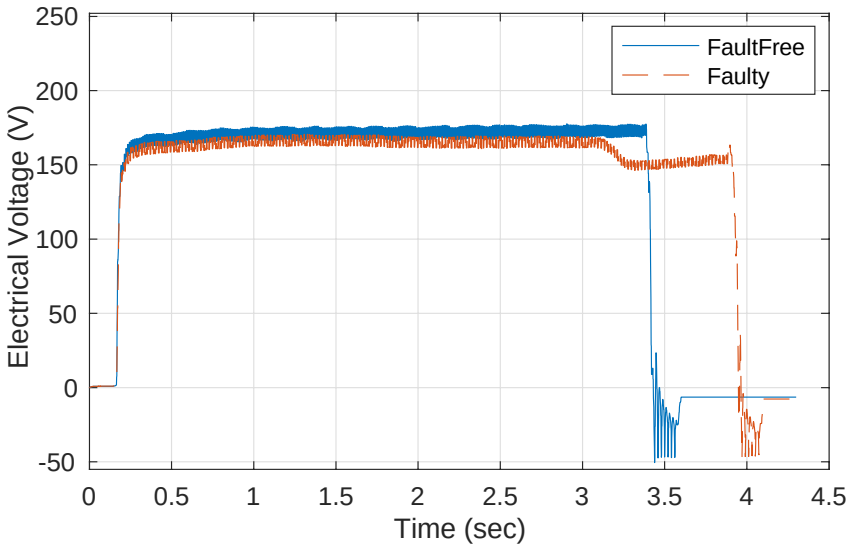

Fig. 5.2 Original electrical voltage measurements under fault-free and faulty conditions

Fig. 5.3 shows the different force sensor behaviours under the conditions of fault-free and faulty. An increased required force is observed when the POE is in faulty condition, which matches the expectation of drive rod overdriving consequences.

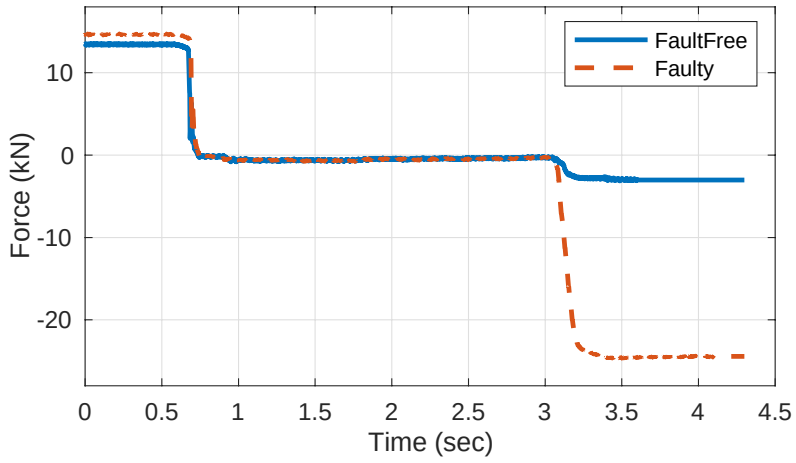

Fig. 5.3 Original force measurements under fault-free and faulty conditions

In the following Section 5.1.1, the EBTW-based feature extraction method will be implemented. The original sensor signals will be decomposed and reconstructed. A parameter selection process will be carried out to determine the values of some key parameters during the signal transformation process by assessing the resolutions of the reconstructed waveforms.

5.1.1 EBTW and Parameter Selection

As described in Section 4.1.2, three individual steps are required to implement the EBTW diagnosis methodology. In Step 1, a discrete wavelet transform (DWT) needs to be applied to the collected sensor data with a pre-defined decomposition level. There are two key points worth noting. First, different decomposition levels will have a significant influence on the decomposition efficiency and diagnosis accuracy. A process of parameter selection is required to decide the most desirable decomposition level. Second, among the three types of sensor data (i.e., current, voltage, and force), current data are selected for detailed demonstration here. The first reason is that the current data shows distinctive waveforms under various fault conditions, which makes it possible to extract effective features for fault classification and diagnosis. The second reason comes from the concern of practical applicability for real industrial application. Current sensors are considered the most cost-effective,

straightforward and non-invasive means of measuring railway switch health condition; refer to Table 3-1. For comprehensiveness of this research, the diagnosis results obtained from current data are later compared with those from voltage and force data, as will be seen in Section 5.1.2.

Fig. 5.4 shows the “Haar” transform of an “Overdriving 6 faces” example current signal. In this figure, the decomposition level is set to be 9, which indicates that the original signal is decomposed at the 9^{th} order of resolution. The results demonstrate that most coefficient amplitudes are approximated as zero. As a result, the subsequent steps regarding signal compression and feature extraction may become more efficient. As different decomposition levels will influence the compression efficiency and diagnosis performance, a comparison is made to identify the most suitable decomposition level, as will be seen later in this subsection of 5.1.1.

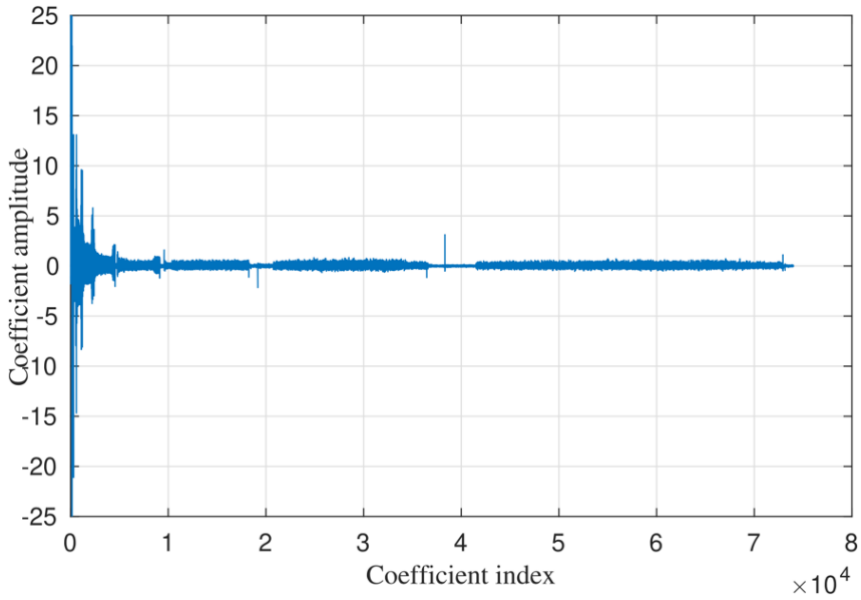

Fig. 5.4 9-level "Haar" transform of "Overdriving 6 faces" fault condition current signal

Step 2 of the methodology is to analyse the energy distribution within the signal by observing the cumulative energy map. As demonstrated in Fig. 5.5, the energy map reaches its maximum of unity very quickly. Finally, in Step 3, the energy percentage to be conserved is determined, and the thresholding technique is applied to the re-ordered coefficients. Fig. 5.6 shows the reconstructed current waveforms when setting the conserved energy percentage to be 99.99%.

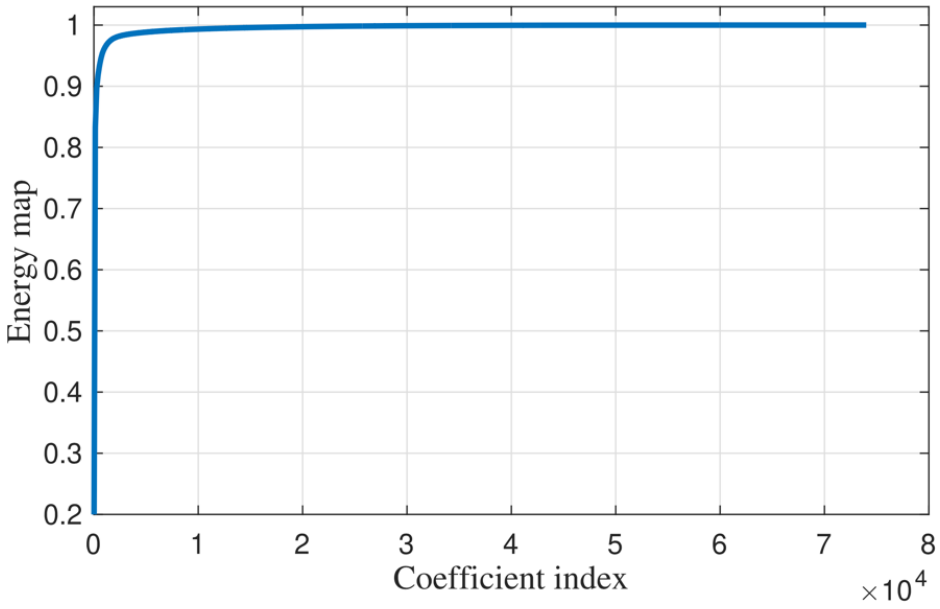

Fig. 5.5 Energy map of the "Haar" transform

It can be observed in Fig. 5.6 that there are four operation phases, which are labelled "inrush", "unlock", "move" and "lock". The four curves separately represent the current signals of different fault severities. When the switch rails are moved out of alignment by turning faces of the adjustment nut, the current curves show distinctive peaks in the "lock" phase. The movement duration is also extended. The force between the switch rails and the stock rails is above the ideal range, and the system suffers an overdriving failure at 11 faces. The four current curves are all smooth without evident noise.

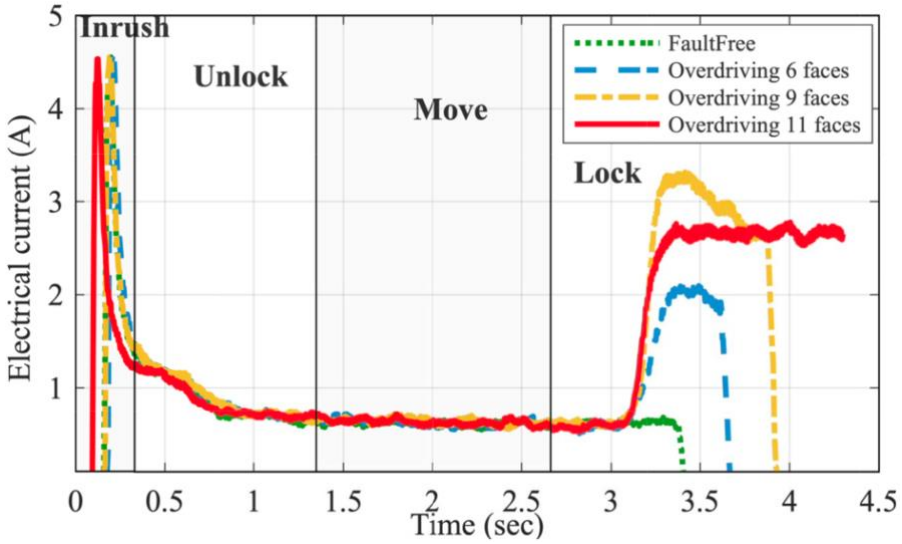

Fig. 5.6 Reconstructed current waveforms with 99.99% energy (dimension reduction ratio = 156:1)

The dimension reduction ratio is calculated as the ratio of original data dimension to the extracted feature dimension. In this case, the initial data dimension of 86,000 is reduced to 548 in feature dimension, indicating a large reduction ratio of 86000:548 ~ 156:1 with respect to current signals.

The reconstructed voltage waveforms shown in Fig. 5.7 also demonstrate a prolonged movement duration due to the overdriven switch. The dimension reduction ratio, with respect to voltage signals, is calculated as 86,000:517 ~ 166:1. This indicates the capability of dimension reduction for current signals and voltage signals is comparable while the performance on voltage signals is slightly better.

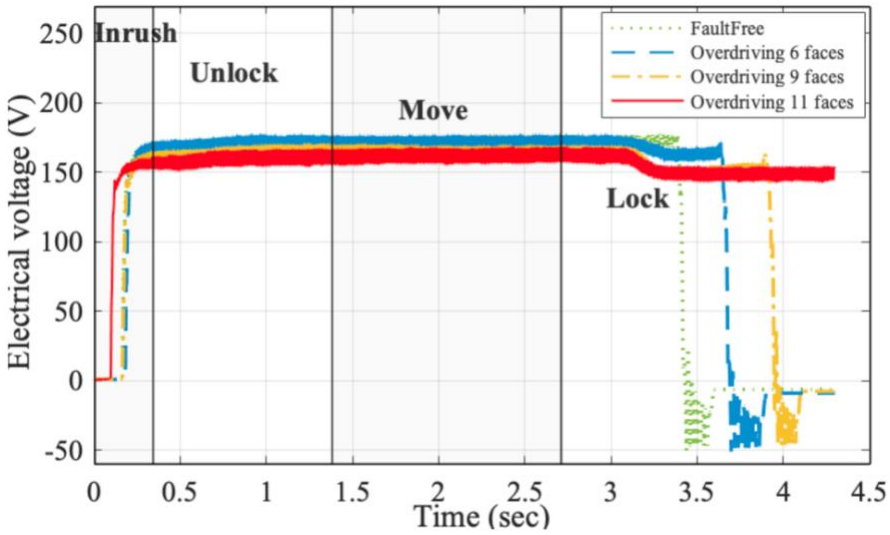

Fig. 5.7 Reconstructed voltage waveforms with 99.99% energy (dimension reduction ratio = 166:1)

The results of the force sensor signals' decomposition and reconstruction are shown in Fig. 5.8. It can be seen that a larger driving rod force is required in the case of a more severe overdriving condition during the locking phase. The force signal dimension ratio is the greatest among the three common types of sensor with 199-dimension feature replacing 86000-dimension original data. The corresponding dimension reduction ratio is thus calculated to be 86000:199 ~ 432:1.

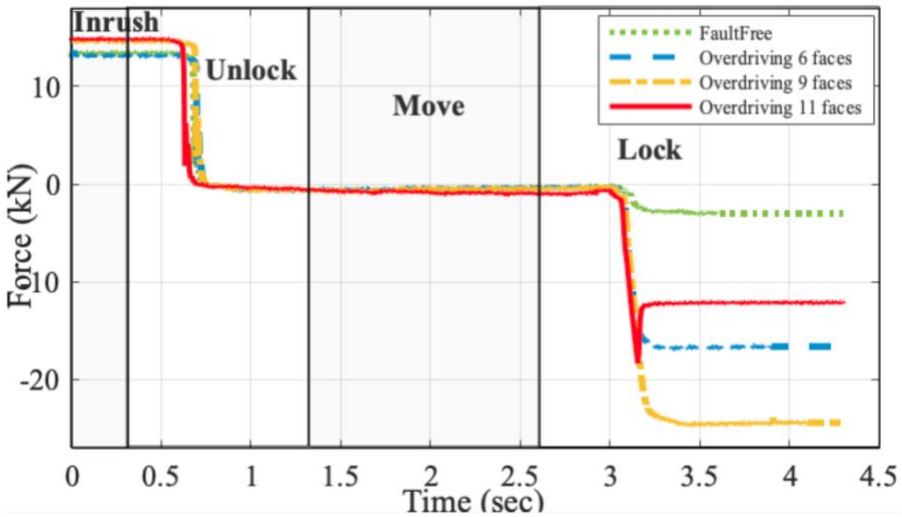

Fig. 5.8 Reconstructed force waveforms with 99.99% energy (dimension reduction ratio = 432:1)

Additionally, a parameter selection process is required in order to find the optimal combination of energy conservation ratio and decomposition level. The waveforms reconstructed from EBTW decomposition are first plotted against various energy conservation ratios, as shown in Fig. 5.9(a)-(d). It can be seen that in this example of reconstructed “Overdriving 6 faces” current waveforms, the results are compared based on the energy conservation ratio changing from 99.99%, 95%, 90% to 85%.

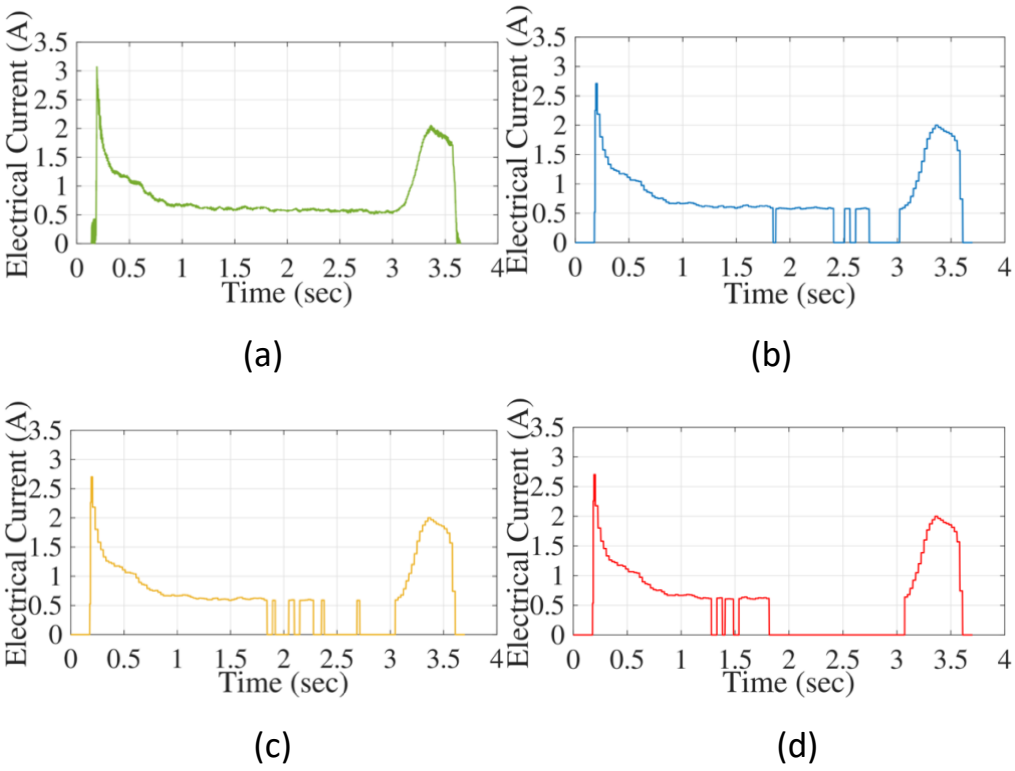

Fig. 5.9 Comparisons of reconstructed "Overdrive 6 faces" current waveforms with different energy conservation ratios. (a)Conserve 99.99%. (b)Conserve 95%. (c)Conserve 90%. (d)Conserve 85%

As is demonstrated in Fig. 5.9(a), the waveform that conserves 99.99% of the original signal energy maintains the most useful characteristics while effectively filtering out the noise components. The peak current in the locking phase, as well as the steady current in moving phase, are clearly captured. As the energy ratio decreases to 95% in Fig. 5.9(b) and to 90% in Fig. 5.9(c), some loss of signal is observed in the moving phase. Based on expert knowledge, it is known that the overdriving fault with different severities is usually identified through the peak

in locking phase. In the cases of 95% and 90% of energy conservation ratio, the peak values during the locking engagement are both well preserved. Nevertheless, as the ratio drops below 85%, as shown in Fig. 5.9(d), the reconstructed current waveform remains at zero for over one second. The continuous loss of signal might be mis-identified as another fault or failure type. From this perspective, the ratio of 90% is taken forward since it preserves an acceptable signal resolution that limits the possibility of mistaking fault or failures, at the same time, benefiting from an efficient signal compression.

Subsequently, a comparison between various decomposition levels is investigated, as shown in Fig. 5.10(a)-(d). According to Eq. 8 and Eq. 9 of the methodology shown in Section 4.1.2.1, the length of scaling coefficients will decrease as the decomposition level increases. At the same time, the wavelet coefficients from more levels will be retained and used for signal reconstruction. Since no evident loss of resolution can be captured from the reconstructed signals at a low level (level 1 - level 7), a discussion on the influence of decomposition level begins at level 8. Fig. 5.10(a)-(d) individually demonstrate the reconstructed current waveforms at decomposition level 8, level 9, level 10 and level 11.

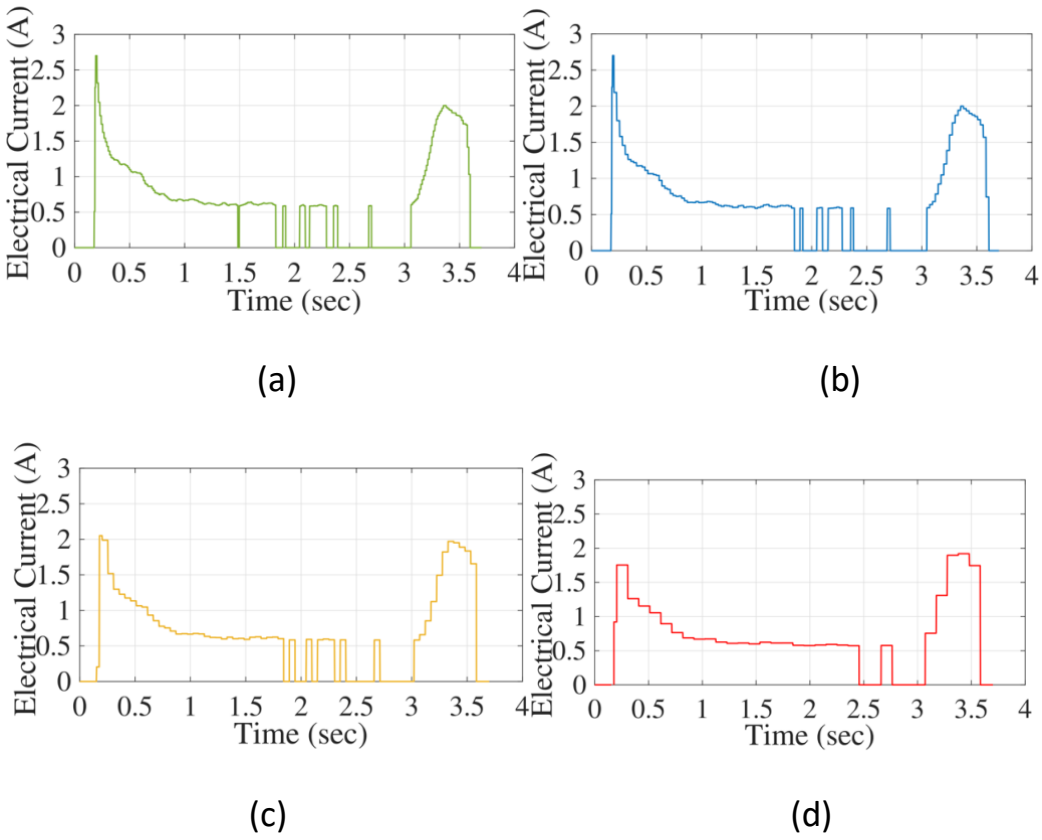

Fig. 5.10 Comparisons of reconstructed "Overdrive 6 faces" current waveforms with different decomposition levels. (a)Level 8. (b)Level 9. (c)Level 10. (d)Level 11

As shown in Fig. 5.10(a)-(d), the reconstructed waveform shows a reduced signal resolution as the decomposition level increases. Clear peaks when the power switches on and lock engages can be seen in the case of level 8 and level 9. As soon as the decomposition level reaches 10, the peak value can longer be observed for fault identification. Thus, the level of decomposition is set to be 9 in this case study. Similar comparisons are also conducted on current waveforms from the other fault, which all end up with the same conclusions.

In summary, the application of the EBTW methodology can effectively remove the noise within raw data sets. The extracted features can well represent the original signals by preserving the key characteristics for fault identification, with a much lower data dimension. A combination of energy conservation ratio of 90% and wavelet decomposition level at 9 is selected in this case to optimize the computation complexity and signal reconstruction resolution. In the following research, comparison results of the EBTW methodology and two well-developed methods (i.e., C-DWT and ST-DWT) are presented, with respect to dimension reduction ratio and diagnosis accuracy (%).

5.1.2 Fault Diagnosis and Comparison Results

In Experiment 1, data from three fault severities (“Fault free”, “Overdriving 6 faces”, and “Overdriving 11 faces”) is used for fault diagnosis. The results will give an indication of the system performance on distinguishing between normal, faulty and failure conditions. However, in practice, an accurate identification of merely three health conditions cannot meet the diagnosis requirement for high reliability and safety of infrastructure. In Experiment 2, one more fault severity (“Overdriving 9 faces”) is included to test the system robustness. As a complementary experiment, the focus of Experiment 3 is a cluster number

sensitivity analysis conducted considering that the continuous degradation data in the real world needs to be first clustered into discrete and observable states. Therefore, Experiment 3 discusses a practical issue when applying the EBTW methodology.

Experiment 1 (Fault Diagnosis for "Fault free", "Overdriving 6 faces" and "Overdriving 11 faces"):

In experiment 1, 60 data sets with each composed of current sensor signals were used. Among the 60 data sets, 20 sets come from "Fault free" state, another 20 sets come from "Overdriving 6 faces" state, and the remaining 20 sets belong to "Overdriving 11 faces" state. In order to train and test the proposed methodology, 48 out of 60 data sets (i.e., 80%) were picked randomly as training sets and the remaining 12 out of 60 sets (i.e., 20%) were for testing.

The results of feature dimension and classification accuracy using current sensor signals from three health conditions are shown in Table 5-1. It can be seen that a comprehensive comparison is made against 5 different classifiers (kNN, SVM, shallow NN, deep NN, and 1-D CNN) working under 3 distinct feature extraction techniques (ST-DWT, C-DWT and the proposed EBTW). The dimension reduction ratio (DRR) is shown as the result of the original data dimension divided by the

dimension of the features. Thus, a higher DRR represents a more efficient process of feature extraction, which will benefit from a shorter computation duration for a fault classifier. The accuracy (%) shown in the following tables (i.e., Table 5-1 to Table 5-5) indicates the percentage of features that are correctly classified. Less corrupting noise within the extracted features leads to a higher diagnosis accuracy. In order to investigate the feature extraction technique that could demonstrate the best diagnosis performance, a comparison of DRR and diagnosis accuracy (%) is conducted between ST-DWT, C-DWT, and the newly proposed EBTW. The fundamental theories and implementation steps of using ST-DWT [89] and C-DWT [87] are provided.

Table 5-1: Comparison of feature dimension and classification accuracy using electrical current signals from 3 health states

Feature Extraction	Dimension Reduction Ratio	Classifier	Accuracy(%)
ST-DWT	86000/2713	kNN	95
		SVM	91.7
		Shallow NN	83.3
		Deep NN	88.89
		1-D CNN	91.67
C-DWT	86000/145	kNN	93.3
		SVM	93.3
		Shallow NN	83.3
		Deep NN	88.89
		1-D CNN	91.67
EBTW	86000/96	kNN	98.3
		SVM	95
		Shallow NN	91.7
		Deep NN	88.89
		1-D CNN	100

As shown in Table 5-1, the DRR is greatest with respect to EBTW methodology, followed by C-DWT and ST-DWT. When analyzing the diagnosis accuracy of a fault classifier using various feature extraction techniques, the classifier combined with the EBTW method always shows the best result. Additionally, it is found that the shallow NN and the deep NN both demonstrate an unsatisfactory accuracy of below 90%. In contrast, two traditional machine learning methods (i.e., kNN and SVM) both have a superior performance demonstrating an accuracy of 90% and above. The performance of the 1-D CNN is the best, which achieves a 100% diagnosis accuracy by employing the EBTW features.

Deep learning techniques can sometimes be implemented without incorporating hand-tuned features for fault diagnosis problems. In other words, no feature extraction process is required, and the deep learning network is capable of dealing with raw data directly. Table 5-2 shows an investigation of neural network classification accuracy without feature extraction.

Table 5-2: Comparison of neural network classification accuracy without feature extractions

Feature Extraction	Classifier	Accuracy(%)
-	Shallow NN	75
	Deep NN	77.78
	1-D CNN	83.33

In Table 5-2, the three listed neural-network-based fault classifiers can all achieve fault diagnosis at a reasonable level. Nevertheless, in comparison with the accuracy shown in Table 5-1 among the EBTW rows, an evident superiority can be observed in the case of the EBTW-supplemented methodology. This strongly demonstrates the importance of the role the EBTW method plays in feature extraction and fault diagnosis.

Compared with ST-DWT and C-DWT, EBTW has demonstrated to possess the most efficient dimension reduction and the most accurate diagnosis results. In the following Experiment 2, the data from one more intermediate faulty

condition is supplemented to validate the robustness of the established diagnosis system.

Experiment 2 (Fault Diagnosis for "Fault free", "Overdriving 6 faces", "Overdriving 9 faces" and "Overdriving 11 faces"):

One intermediate faulty condition between "Overdriving 6 faces" and "Overdriving 11 faces" was added in this experiment, by adding 20 more data sets of "Overdriving 9 faces" to those used in Experiment 1. As such, 80 data sets with each containing 20 data sets from 4 fault conditions were applied in this case. The 80 data sets were randomly divided into 64 sets (80%) and 16 sets (20%) for training and testing purposes, respectively. The significance of adding one more faulty condition that is closer to a failure state instead of a fault-free state is that a diagnosis system is expected to be more sensitive to a more severe faulty condition in reality. Specifically, more concentration needs to be focused on a system when it is close to failure. Detailed diagnosis information and accurate decision support should be provided, which assists the scheduling of maintenance work more accurately and provides a higher chance of preventing railway switch failures.

In Table 5-3, a comparison of DRR and classification accuracy is presented using current sensor signals for 4 health conditions. In comparison with ST-DWT and C-DWT, the EBTW methodology enjoys the most efficient feature dimension reduction and achieves the highest diagnosis accuracy.

Table 5-3: Comparison of feature dimension and classification accuracy using electrical current signals from 4 health states

Feature Extraction	Dimension Reduction Ratio	Classifier	Accuracy(%)
ST-DWT	86000/3553	kNN	96.2
		SVM	93.8
		Shallow NN	81.2
		Deep NN	83.33
		1-D CNN	100
C-DWT	86000/165	kNN	97.5
		SVM	93.8
		Shallow NN	87.5
		Deep NN	91.67
		1-D CNN	100
EBTW	86000/106	kNN	100
		SVM	97.5
		Shallow NN	93.8
		Deep NN	100
		1-D CNN	100

As a result of effective noise removal and feature extraction in EBTW, its combination with kNN, deep NN, and 1-D CNN all reach an accuracy of 100%. This further implies that the proposed EBTW method not only works as an effective fault diagnosis scheme but also surpasses the previously developed ST-DWT and C-DWT in terms of computation speed and diagnosis accuracy. The same conclusions are also drawn for the voltage sensor signals shown in Table 5-4 and force sensor signals shown in Table 5-5.

Table 5-4: Comparison of feature dimension and classification accuracy using electrical voltage signals from 4 health states

Feature Extraction	Dimension Reduction Ratio	Classifier	Accuracy(%)
ST-DWT	86000/4723	kNN	97.5
		SVM	93.8
		Shallow NN	87.5
		Deep NN	83.33
		1-D CNN	95
C-DWT	86000/165	kNN	98.8
		SVM	96.2
		Shallow NN	93.8
		Deep NN	83.33
		1-D CNN	100
EBTW	86000/130	kNN	98.8
		SVM	96.2
		Shallow NN	100
		Deep NN	91.67
		1-D CNN	100

Table 5-5: Comparison of feature dimension and classification accuracy using force signals from 4 health states

Feature Extraction	Dimension Reduction Ratio	Classifier	Accuracy(%)
ST-DWT	86000/359	kNN	98.8
		SVM	93.8
		Shallow NN	81.2
		Deep NN	91.67
		1-D CNN	100
C-DWT	86000/168	kNN	100
		SVM	98.8
		Shallow NN	93.8
		Deep NN	83.33
		1-D CNN	100
EBTW	86000/70	kNN	100
		SVM	100
		Shallow NN	100
		Deep NN	91.67
		1-D CNN	100

When analyzing the performance of EBTW for different sensor types, it can be observed that the current sensor and voltage sensor demonstrate comparable

fault sensitivities, while the DRR results are slightly better for the current sensor. In contrast, the force sensor demonstrates the highest fault sensitivity because its DRR and diagnosis accuracy are both high, no matter which method it uses. The best result is found when using force sensor signals with the EBTW technique, the feature dimension of which has reduced from 86,000 to 70.

Experiment 3 (Cluster Number Sensitivity Analysis):

In the previous two experiments, "well-labelled" data from various health conditions were used. However, this is quite difficult to achieve for a practical application. More specifically, observing the health conditions the machine goes through and knowing the total number of conditions are two huge challenges facing health state diagnosis for complex operational systems. For example, in the case of a railway switch system, it becomes very challenging to classify an incipient overdriving fault with levels of severities in real life. Therefore, Experiment 3 serves as a supplement discussing this practical issue when applying the proposed methodology. Clustering and cluster validity index (CVI) are used in this experiment to explore the total number of health conditions and the data sets belonging to real health conditions. CVI is used in this case to decide the optimal cluster number (health conditions) since the real cluster number

indicating the real number of degradation levels for a railway switch is unknown in practical application.

In Experiment 3, the total number of data sets is expanded to 160 containing 20 data sets from each of the 8 health conditions. The 160 data sets with incremental overdriving fault severities can also be regarded as an incipient fault progression process. Feature extraction is required before clustering using k-means. Subsequently, several CVIs, i.e., Davies-Bouldin (DB), Duun, Calinski-Harabasz (CH) and Silhouette, were applied to decide the appropriate cluster number. Previous work has provided the definitions and full calculation methods in [98]. Among them, the results of the CH index were taken forward as the most robust choice for the data used in this experiment. A higher value of the CH index indicates a good cluster because it represents a larger separation among the different clusters, as well as a smaller separation between the points in a single cluster. The cluster number that corresponds to the maximum CH value is decided to be the optimal. Fig. 5.11 shows the CH index results against different number of clusters changing from 2 to 10 in k-means.

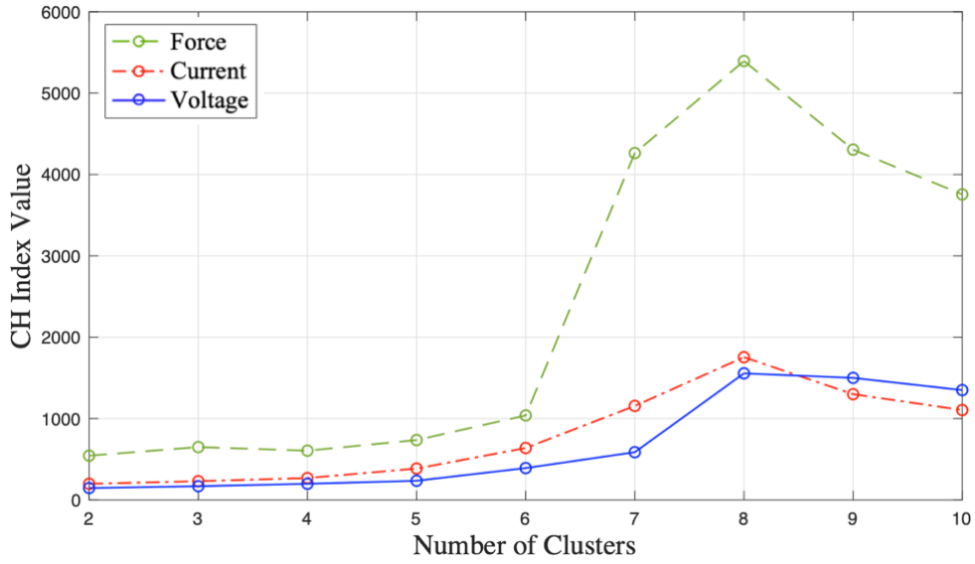

Fig. 5.11 Evaluation of number of clusters in k-means method

As shown in Fig. 5.11, the maximum CH index value appears at 8 number of clusters, for all 3 sensor types. The most distinctive peak is captured for the force sensor, followed by current sensor and voltage sensor. As such, the real incipient overdriving fault progression is estimated to be modelled with 8 health conditions. By setting the cluster number to be 8 in k-means, each simulated health condition has been grouped into one cluster with 100% accuracy. Alternatively, each clustered data group is now represented as real and distinctive fault conditions, rather than simulated fault severities. The aforementioned EBTW-CNN classification method can then be used for detection and diagnosis, which will identify the system's real health condition. As such, not only is the current situation within the overall fault degradation process well understood, but also a future study of fault prognosis and RUL estimations can be prepared for.

As a summary, a condition monitoring system for railway switch fault diagnosis has been presented. Three simulated overdriving fault severities were considered in Experiment 1, the results of which demonstrate that EBTW surpassed the two traditional feature extraction methods in terms of dimension reduction ratio and diagnosis accuracy. Among the five fault classifiers, 1-D CNN demonstrated the best diagnosis performance. An additional faulty condition was considered in Experiment 2, from which the same conclusion is drawn. The comparison of fault sensitivity among various sensor types shows that the force sensor indicates the highest sensitivity with the most efficient dimension reduction. The combination of EBTW and the cost-effective current sensor also shows a 100% accuracy. Experiment 3 applied k-means clustering and Calinski-Harabasz index to investigate the real number of health conditions during a continuous fault progression. The results show that each simulated fault severity defines one health condition in reality. Therefore, the EBTW methodology is capable of diagnosing a machine's real health condition, which also establishes a concrete basis for future study of fault prognostics.

5.2 Data-Driven Prognosis Methodology: RMFOP-Based Degradation Modelling

To demonstrate and evaluate the RMFOP-based prognosis method proposed in Section 4.2, a group of degradation data was collected from operational railway switch systems in Great Britain. Unlike the diagnosis data used in Section 5.1, the prognosis data directly relates to the physical evolution of state degradation. Based on the experimental analysis provided in Section 5.1.2, force sensors and current sensors both demonstrate reliable sensitivity to railway switch faults, when compared to voltage sensors. However, force sensors cost ten times more than current sensors; refer to Table 3-1. Therefore, in the following research on fault prognosis, a current sensor is employed and the data is collected. The results of the data processing are shown in the following Subsection 5.2.1, by implementing the RMFOP methodology in 10 steps.

5.2.1 Experimental Results

Step 1 of the RMFOP methodology is data collection; 50 groups of current sensor data were collected from hydraulic type switches over a 13-month period ranging from January 2018 to February 2019. The data collected from each railway switch is called a data group. Each data group is a series of event-based data records. Each record is a time series describing the current sensor

waveform for one movement of the switch. **Step 2** is to transform the original sensor data into degradation signals. The degradation signals are also known as characteristic patterns, which can directly demonstrate the fault progression. In terms of the railway switch case used in this research, many characteristics including peak, variance, average and median have been estimated. The time-domain average current amplitude is taken forward because of its clear representation of the deterioration process and well-defined maintenance thresholds per movement. The thresholds are provided by Network Rail, which is responsible for the infrastructure management of most railway networks in Great Britain. Once the monitored signals exceed the threshold, a fault alert will be generated in the operation centre for further field inspections to be scheduled.

Step 3 of the methodology is to calculate the degradation signal for each switch movement and plot against the operating period ranging from January 2018 to February 2019. Fig. 5.12 demonstrates the results of average amplitude over the operating period for one railway switch.

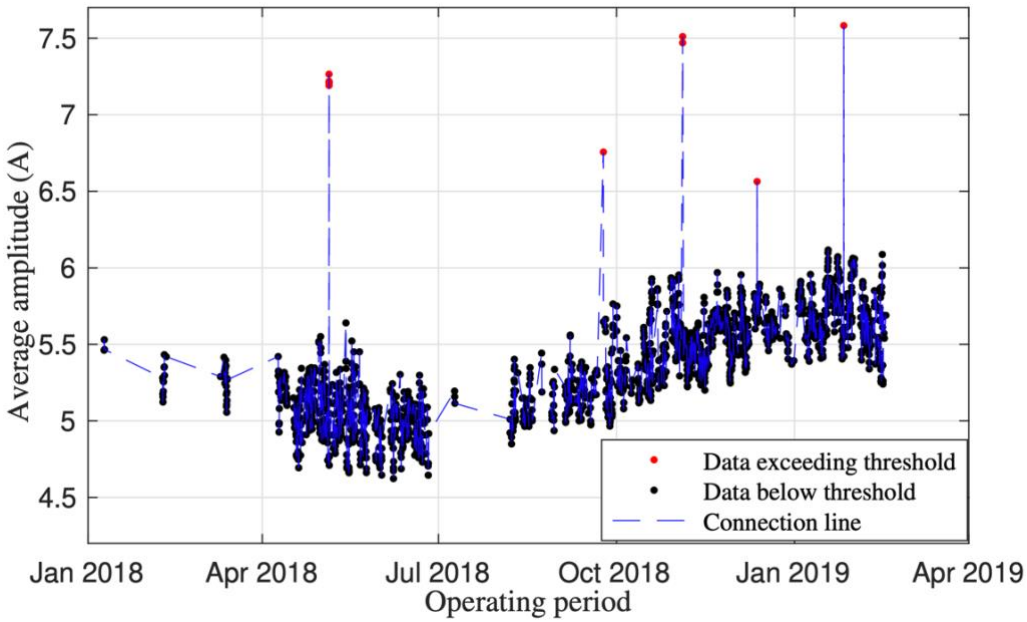

Fig. 5.12 An example of degradation signals: average amplitude over the operating period

The average amplitude of current signals is calculated each time the switch moves and is marked as a data point in Fig. 5.12. The red data points are used to describe the average amplitude exceeding the maintenance threshold, while the residual data points are all marked in black. The connection line represents a blue dashed line linking adjacent data points. Some gaps can be observed in Fig. 5.12 between adjacent data points indicating the switch did not operate during those periods of time.

Subsequently, the degradation signals are truncated to form degradation paths. Five instances of maintenance interventions are marked as red points in Fig. 5.12, which means four degradation paths can be extracted by each time truncating

at the threshold and splitting the degradation signals into several sections. It is assumed the maintenance alerts have all been addressed properly with faults identified and fixed. As such, the degradation signals between two adjacent maintenance records are treated as one degradation path, describing the changes in average current amplitude during switch movements from fault-free to faulty condition. An example of a single degradation path is demonstrated in Fig. 5.13.

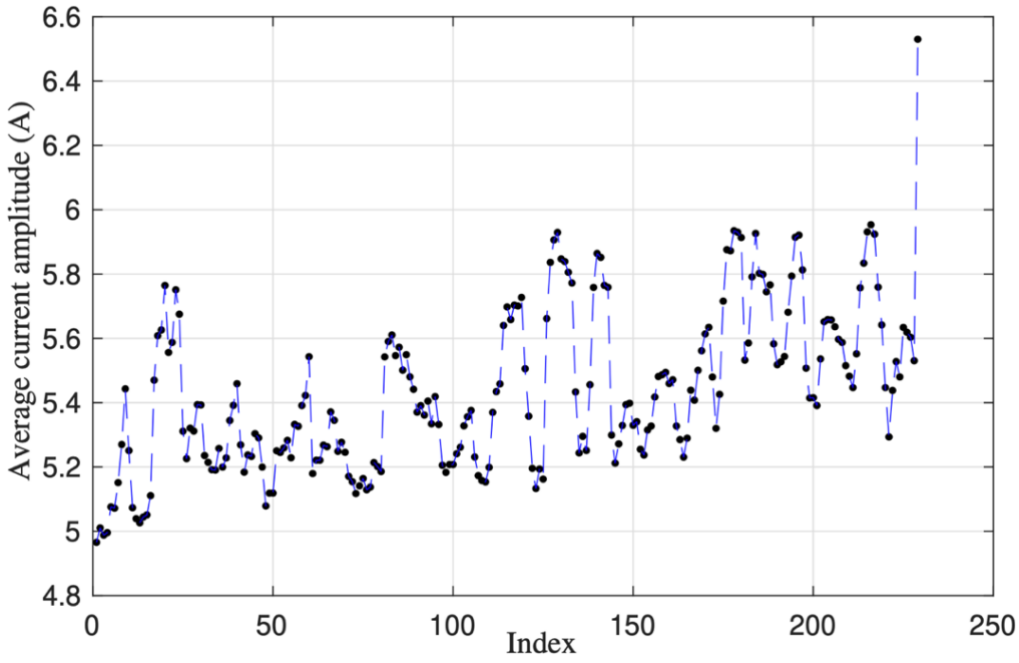

Fig. 5.13 An example of degradation path

It can be seen that the x-axis representing the switch useful life is a time-free index, which refers to the number of events (a single switch movement). In other words, the progression of a machine's fault is recorded only when a movement event happens. The benefit of the time-free measurement is the independence

to the process of fault occurrence. The example degradation path shown in Fig. 5.13 operates approximately 230 times before triggering a fault alarm.

In general, three types of faults will influence the switch operations: intermittent faults, abrupt faults and incipient faults. The progression of intermittent faults and abrupt faults are usually difficult to describe. In contrast, incipient faults could be monitored and predicted as the selected models and parameters are appropriate for the system. In this case study, only the degradation paths demonstrating a gradual increase of average amplitude will be taken forward since they indicate an incipient fault progression and could be predicted by the proposed methodology. A total of 50 degradation paths are extracted from the data groups and will be taken through the following method steps.

To further identify the fault type, examples of current waveforms recorded at a switch's different life percentages are demonstrated in Fig. 5.14. In the case of a railway switch system, the value of the life percentage is determined by the ratio of the number of movements the switch has already performed since the last maintenance intervention (time-free index) to the overall number of switch movements. The overall number of switch movements is counted from the first

time the switch has operated since the last maintenance intervention until the time the switch triggers a fault alarm.

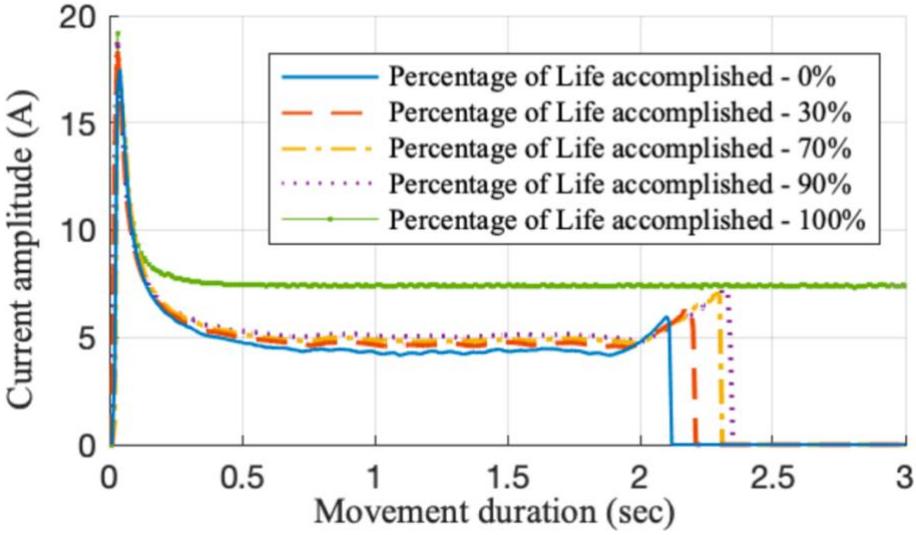

Fig. 5.14 Current waveforms at different degradation percentages

An incipient overdriving fault is observed to be progressing, corresponding to an increased peak at the end of each current waveform. Overdriving defines a fault condition where the force between the switch blade and the stock blade is above the ideal range. This consumes more electric energy to lock the switch after moving it to the proper position. The motor current supply, as well as the overall operation duration, is expected to increase under this circumstance. A continuous current supply is seen in Fig. 5.14 where the switch has reached 100% life, because an overdriving failure occurs and the switch fails to be locked properly.

After detailed analysis of the 50 degradation paths, it is found they can each represent one realization of an incipient overdriving fault progression. Among them, 70% of the paths, i.e., 35 paths, are randomly selected to calculate the model prior distributions, which will be individually updated and validated using the remaining 30% of the paths.

As described in **Step 4** of the methodology, the normalized 35 paths are treated as prior knowledge to estimate parameters within the linear degradation model. According to *Eq. 19*, the prior mean is calculated to be $\mu_\theta = 0.0035892$. The prior variance is calculated as $\sigma_\theta^2 = 1.2781 \times 10^{-5}$. The constant c is estimated to be 0.096942. Variance of the error term is $\sigma^2 = 0.006$, by estimating the highest level of fluctuation. As such, the population-wide characteristics among the monitored switch systems have been identified. Then, the RMFOP distributions are individually estimated for each remaining validation switch system under the various life percentages, by following **Step 5** to **Step 7**. Fig. 5.15 shows the updated RMFOP distributions of one validation switch system.

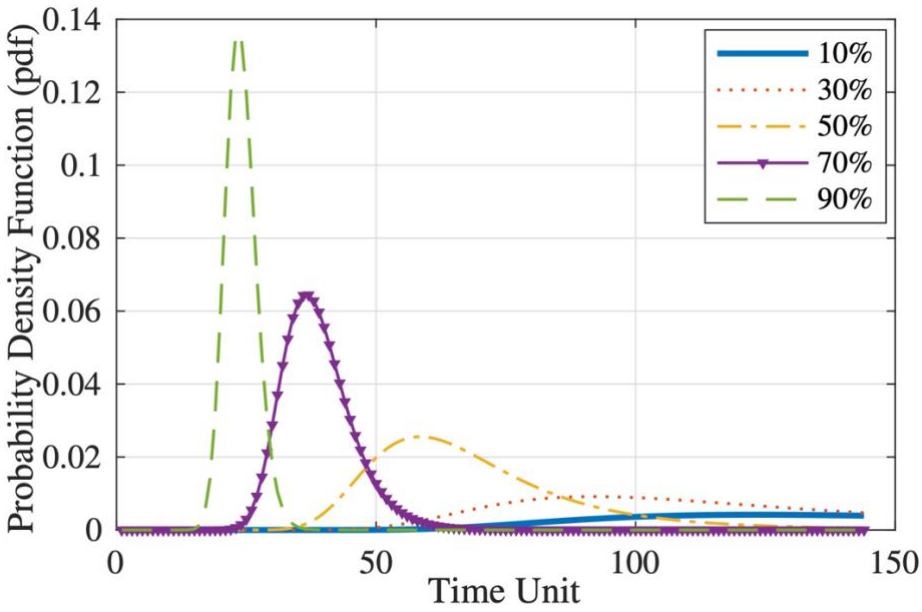

Fig. 5.15 Updated RMFOP distributions of one validation switch system using the linear model

Subsequently, **Step 8** is implemented to calculate the prediction errors at various life percentages with 95% confidence interval for all validation paths. The prediction error is computed as the ratio of the estimated failure time error to the actual failure time. The estimated failure time is the time for which the switch has already been operating since the last maintenance intervention, plus the estimated remaining life. The switch operating time and the estimated remaining life are both measured by number of switch movements, which is a time-free index. The estimated remaining life measuring the predicted number of switch movements before triggering a fault alarm refers to the median of the residual life distribution. The median value is adopted since it can reasonably

measure the central tendency of the distribution. A mathematical expression of prediction error R_i^j is demonstrated in Equation (5.1):

$$R_i^j = \frac{\left|\left(t_o^j + t_{p,i}^j\right) - t_{a,i}^j\right|}{t_{a,i}^j} \tag{5.1}$$

t_o^j is the machine's operating time until the jth observation time. $t_{p,i}^j$ is the predicted residual useful life according to the median value of the ith machine life distribution curve. As such, the summation of t_o^j and $t_{p,i}^j$ represents the predicted time of failure with respect to the ith machine at time j. $t_{a,i}^j$ represents the actual time of failure. The numerator in Eq. 28 therefore represents the absolute prediction error of the machine's failure time. The ratio of the prediction error for the $i$the machine at observation time j is given by R_i^j.

Step 9 is to choose a different model and proceed through Steps 4 to 8 as before. In this case, the prediction results of the exponential degradation model are computed and compared with those derived earlier for linear model; the results are given in Fig. 5.16. For a clear demonstration, the linear curve is slightly offset to the right.

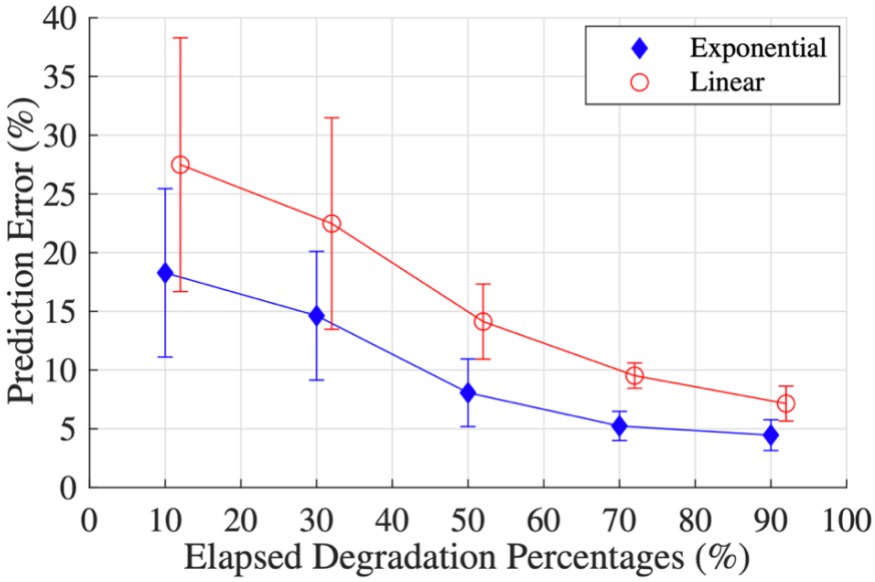

Fig. 5.16 A comparison between the linear model and the exponential model regarding residual life prediction accuracy

It is clear that in both the case of linear and exponential models, the prediction errors will decrease when more life has elapsed. The prediction errors for the exponential model are generally smaller than those for the linear model, which indicates the exponential degradation form may be a better selection for characterizing railway switch fault progression in the real world. This finding can be attributed to the fact that the cumulative damage accelerates the speed of degradation for a switch machine in real life. In addition, the results in Fig. 5.16 demonstrate that the usage of linear and exponential models is a reasonable starting point for RMFOP predictions in terms of complex industrial applications. The proposed RMFOP-based methodology is capable of predicting

and updating the residual useful life distributions when more life percentage has elapsed.

Step 10 is to validate the effectiveness of Bayesian theory on model updating. In Fig. 5.17, the residual life prediction error is plotted for both the 'linear with updating model' and the 'linear without updating model'.

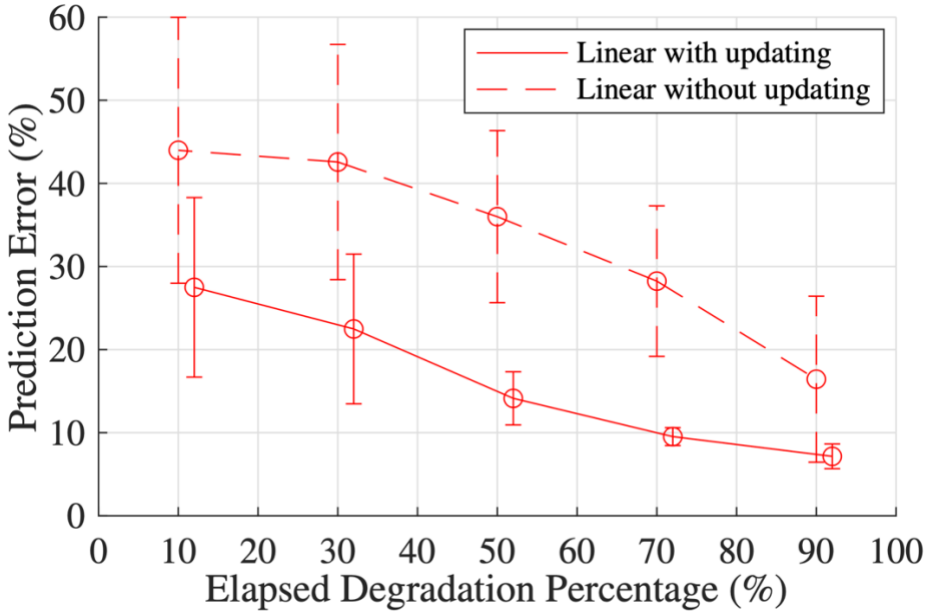

Fig. 5.17 A comparison between the 'linear with updating model' and the 'linear without updating model' regarding residual life prediction accuracy

It can be seen that in the case of 'linear with updating model', the prediction error has been decreased dramatically due to its application of Bayesian theory for updating parameter statistics.

A similar finding is obtained when comparing the ‘exponential with updating model’ and the ‘exponential without updating model’, as shown in Fig. 5.18.

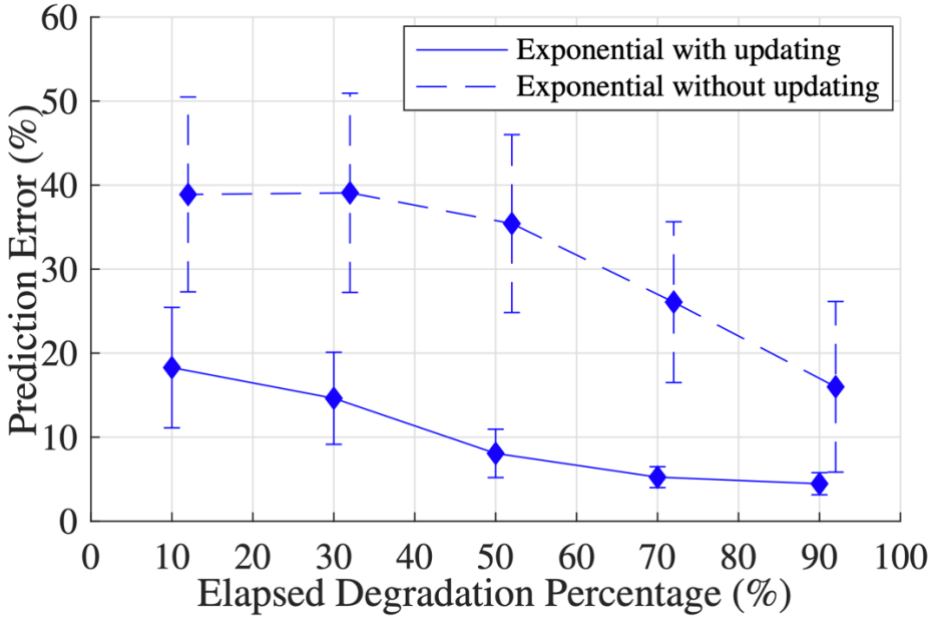

Fig. 5.18 A comparison between the ‘exponential without updating model’ and the ‘exponential with updating model’ regarding residual life prediction accuracy

It is found that the prediction error of ‘exponential without updating model’ fluctuates by around 30%, while the error drops to below 20% when Bayesian theory is applied and parameter distributions are updated.

Thus, a proof of the concept of the RMFOP-based prognosis methodology has been achieved. By implementing the methodology on real-world railway switch systems, the health condition can be recognized and predicted. The probability

distributions of the RMFOP can also be estimated and plotted for any future time.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Fault diagnostics and fault prognostics are the two key components of condition monitoring of railway switch systems. In the most recent two decades, the significance of condition diagnostics and prognostics has been grasped by both the rail industry as well as researchers. Visual inspections that are carried out periodically by personnel may not provide the basis for timely or accurate decision support in these areas. On the other hand, limited research has been conducted using real-world data combined with state-of-the-art methods. The core aim of diagnostics and prognostics is to precisely identify and predict existing / potential faults, respectively, by means of analysing collected sensor data. The benefits of diagnostics and prognostics are the provision of enhanced safety, reliability and availability through better decision support for inspection and maintenance scheduling.

6.1 Summary of Intellectual Contributions

The aim of this research was to explore and enhance the current diagnosis and prognosis approaches in terms of the system performance and the effectiveness for real-world railway switch applications. In order to fulfil the aim, the established literature regarding fault diagnosis and prognosis was categorized. It was found that three types of approach are normally applied to achieve machinery diagnosis results: statistical approaches, AI approaches, and physical approaches. Each type of approach has its own pros and cons. AI diagnosis approaches were identified as showing the highest possibility of being applied to real-world railway switches. Similarly, an analysis and discussion were carried out in terms of three prognosis approach types, the results of which indicate the data-driven prognosis approaches are most likely to be applicable to railway switch systems. Investigating methods' applicability to real-world railway switches is of great importance, hence this research focussed on the topic of AI-based diagnosis and data-driven prognosis, with respect to data collection, method exploration, and result analysis. The intellectual contributions of this research are summarised as follows:

1. This research provided an essential industrial criterion for selecting the most effective sensor that shows high sensitivity to railway switch faults. Specifically,

the data collected from various measurement sensors were explored and compared. The best sensor for monitoring the switch health condition was found when the least computation time was consumed but obtained the highest diagnosis accuracy. The comparisons were made for switch motor current sensor, motor voltage sensor and force sensor; refer to Section 5.1.2 for details.

2. This research proposed a fault diagnosis algorithm that improves the computation complexity and decision accuracy by demonstrating its effectiveness to diagnosing faults in railway switches. Specifically, by localizing and redistributing the signal energy during the wavelet transform process, the derivation of a meaningful descriptor was achieved. The descriptors that replace the raw data were treated as the input for the diagnosis models, which was demonstrated to have achieved a more accurate and quicker diagnosis analysis compared with some state-of-the-art techniques. For related mathematical derivations and methodology explanations, refer to Section 4.1.2. For experimental study and results analysis, refer to Section 5.1.1 - 5.1.2.

3. This research put forward a fault prognosis methodology that is applicable to operational industrial systems in practice. Specifically, a novel prognosis definition and related signal processing approach were proposed and validated

using the real-world switch system data, whose use effectively overcomes the common difficulty in fitting the real-world raw data into established methods for an industrial application. The applicability of the methodology is enhanced when compared with most methods reported in the literature using idealized lab data for proof of concept. Refer to Section 4.2.2 for method explanation. Refer to Section 5.2.1 for experimental results.

The work presented in this thesis has been published in two journals and one academic conference: refer to Appendix A for details.

6.2 Future Work

The main fields for future work will include the exploration of the applicability of the proposed method. As well as the electro-hydraulic type of railway switches applied in this research, other switch machines such as electro-mechanical and electro-pneumatic types should be investigated. In addition, further consideration should be given to the model assumptions in the prognosis algorithm. As seen in Section 4.2.2.2, two monotonically increasing functions (i.e., linear function and exponential function) are applied to model machine faults with different degradation rates. A future study investigating the modelling of a wider set of degradation patterns and rates using various

functions including the Weibull, Gaussian and lognormal can be expected to demonstrate the broadening of the method's applicability. In this study the decision is made to use the median value to represent the residual life value when the residual life probability distribution is obtained. A further improvement could be achieved in prediction accuracy if a close mathematical form is found to describe the residual life distribution, and distribution moments can be computed and replace the median to become a more representative value of the residual life. Furthermore, some consideration can be devoted to investigating deep learning models for prognosis data analysis, which will include collecting an enormous amount of sensor data, establishing and optimizing the model structure for the best outcomes. Also, the collected datasets can serve as a benchmark dataset for testing other newly developed prognosis algorithms in the future.

REFERENCES

1. Office of Rail and Road, *Annual assessment of Network Rail, April 2019-March 2020*. 2020. p. 23-25.
2. Office of Rail and Road, *Delay attribution review: Scoping stage report*. 2019. p. 18-20.
3. Wikipedia contributors. *Railroad switch*. [cited 2020 August 3]; Available from: https://en.wikipedia.org/w/index.php?title=Railroad_switch&oldid=965773459.
4. Branch, R.A.I., *Derailment at Grayrigg, Cumbria 23 February 2007.* Progress report. Report IR2/2007, Department of Transport, UK, 2014.
5. Lu, S., R. Yan, Y. Liu, and Q. Wang, *Tacholess speed estimation in order tracking: A review with application to rotating machine fault diagnosis.* IEEE Transactions on Instrumentation and Measurement, 2019. **68**(7): p. 2315-2332.
6. Nivesrangsan, P. and D. Jantarajirojkul. *Bearing fault monitoring by comparison with main bearing frequency components using vibration signal*. in *2018 5th International Conference on Business and Industrial Research (ICBIR)*. 2018. IEEE.
7. Borovik, S. and Y. Sekisov, *Single-coil eddy current sensors and their application for monitoring the dangerous states of gas-turbine engines.* Sensors, 2020. **20**(7): p. 2107.
8. Dybkowski, M. and K. Klimkowski, *Artificial neural network application for current sensors fault detection in the vector controlled induction motor drive.* Sensors, 2019. **19**(3): p. 571.
9. Sadoughi, M. and C. Hu. *A physics-based deep learning approach for fault diagnosis of rotating machinery*. in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*. 2018. IEEE.
10. Ademujimi, T.T., M.P. Brundage, and V.V. Prabhu. *A review of current machine learning techniques used in manufacturing diagnosis*. in *IFIP International Conference on Advances in Production Management Systems*. 2017. Springer.
11. Network Rail, *Asset Management Policy*, January; Available at: https://cdn. networkrail. co. uk/wp-content/uploads/2018/02/Asset-Management-Policy. pdf (accessed August 8, 2018), 2018.
12. ISO, *13374-1: 2003 Condition Monitoring and Diagnostics of Machines—Data Processing*, in *Communication and Presentation—Part*. 2003.
13. Jardine, A.K., D. Lin, and D. Banjevic, *A review on machinery diagnostics and prognostics implementing condition-based maintenance.* Mechanical systems and signal processing, 2006. **20**(7): p. 1483-1510.

14. Lei, Y., B. Yang, X. Jiang, F. Jia, N. Li, and A.K. Nandi, *Applications of machine learning to machine fault diagnosis: A review and roadmap.* Mechanical Systems and Signal Processing, 2020. **138**: p. 106587.
15. Ciabattoni, L., F. Ferracuti, A. Freddi, and A. Monteriu, *Statistical spectral analysis for fault diagnosis of rotating machines.* IEEE Transactions on Industrial Electronics, 2017. **65**(5): p. 4301-4310.
16. Liu, R., B. Yang, E. Zio, and X. Chen, *Artificial intelligence for fault diagnosis of rotating machinery: A review.* Mechanical Systems and Signal Processing, 2018. **108**: p. 33-47.
17. Bouzid, M.B.K. and G. Champenois, *An efficient simplified physical faulty model of a permanent magnet synchronous generator dedicated to stator fault diagnosis part II: Automatic stator fault diagnosis.* IEEE Transactions on Industry Applications, 2017. **53**(3): p. 2762-2771.
18. Li, Z., Y. Wang, and K.-S. Wang, *Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario.* Advances in Manufacturing, 2017. **5**(4): p. 377-387.
19. Zhong, K., M. Han, and B. Han, *Data-driven based fault prognosis for industrial systems: a concise overview.* IEEE/CAA Journal of automatica sinica, 2019. **7**(2): p. 330-345.
20. Lu, Y., Q. Li, and S.Y. Liang, *Physics-based intelligent prognosis for rolling bearing with fault feature extraction.* The International Journal of Advanced Manufacturing Technology, 2018. **97**(1-4): p. 611-620.
21. Hu, L.-Q., C.-F. He, Z.-Q. Cai, L. Wen, and T. Ren, *Track circuit fault prediction method based on grey theory and expert system.* Journal of Visual Communication and Image Representation, 2019. **58**: p. 37-45.
22. Durazo-Cardenas, I., A. Starr, C.J. Turner, A. Tiwari, L. Kirkwood, M. Bevilacqua, A. Tsourdos, E. Shehab, P. Baguley, and Y. Xu, *An autonomous system for maintenance scheduling data-rich complex infrastructure: Fusing the railways' condition, planning and cost.* Transportation Research Part C: Emerging Technologies, 2018. **89**: p. 234-253.
23. Cai, B., L. Huang, and M. Xie, *Bayesian networks in fault diagnosis.* IEEE Transactions on Industrial Informatics, 2017. **13**(5): p. 2227-2240.
24. Bartram, G.W., *System health diagnosis and prognosis using dynamic Bayesian networks*. 2013.
25. Wu, S., L. Zhang, W. Zheng, Y. Liu, and M.A. Lundteigen, *A DBN-based risk assessment model for prediction and diagnosis of offshore drilling incidents.* Journal of Natural Gas Science and Engineering, 2016. **34**: p. 139-158.
26. Cai, B., H. Liu, and M. Xie, *A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks.* Mechanical Systems and Signal Processing, 2016. **80**: p. 31-44.
27. Jensen, T.B., A.R. Kristensen, N. Toft, N.P. Baadsgaard, S. Østergaard, and H. Houe, *An object-oriented Bayesian network modeling the causes of leg disorders in finisher herds.* Preventive veterinary medicine, 2009. **89**(3-4): p. 237-248.
28. Madeti, S.R. and S. Singh, *Modeling of PV system based on experimental data for fault*

detection using kNN method. Solar Energy, 2018. **173**: p. 139-151.
29. Benmahamed, Y., M. Teguar, and A. Boubakeur, *Application of SVM and KNN to Duval Pentagon 1 for transformer oil diagnosis.* IEEE Transactions on Dielectrics and Electrical Insulation, 2017. **24**(6): p. 3443-3451.
30. Yin, Z. and J. Hou, *Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes.* Neurocomputing, 2016. **174**: p. 643-650.
31. Yang, J., Y. Zhang, and Y. Zhu, *Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension.* Mechanical Systems and Signal Processing, 2007. **21**(5): p. 2012-2024.
32. Widodo, A. and B.-S. Yang, *Wavelet support vector machine for induction machine fault diagnosis based on transient current signal.* Expert Systems with Applications, 2008. **35**(1-2): p. 307-316.
33. Wenyi, L., W. Zhenfeng, H. Jiguang, and W. Guangfeng, *Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM.* Renewable Energy, 2013. **50**: p. 1-6.
34. Bacha, K., S. Souahlia, and M. Gossa, *Power transformer fault diagnosis based on dissolved gas analysis by support vector machine.* Electric power systems research, 2012. **83**(1): p. 73-79.
35. Chinnam, R.B. and P. Baruah, *Autonomous diagnostics and prognostics in machining processes through competitive learning-driven HMM-based clustering.* International Journal of Production Research, 2009. **47**(23): p. 6739-6758.
36. Kouadri, A., M. Hajji, M.-F. Harkat, K. Abodayeh, M. Mansouri, H. Nounou, and M. Nounou, *Hidden Markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems.* Renewable Energy, 2020. **150**: p. 598-606.
37. Huang, L., H. Huang, and Y. Liu, *A Fault Diagnosis Approach for Rolling Bearing Based on Wavelet Packet Decomposition and GMM-HMM.* International Journal of Acoustics & Vibration, 2019. **24**(2).
38. Chen, Q., G. Nicholson, C. Roberts, J. Ye, and Y. Zhao, *Improved Fault Diagnosis of Railway Switch System Using Energy-based Thresholding Wavelets (EBTW) and Neural Networks.* IEEE Transactions on Instrumentation and Measurement, 2020: p. 1-1.
39. Chen, Q., G. Nicholson, J. Ye, and C. Roberts. *Fault Diagnosis Using Discrete Wavelet Transform (DWT) and Artificial Neural Network (ANN) for A Railway Switch*. in *2020 Prognostics and Health Management Conference (PHM-Besançon)*. 2020. IEEE.
40. Dhumale, R. and S. Lokhande, *Neural network fault diagnosis of voltage source inverter under variable load conditions at different frequencies.* Measurement, 2016. **91**: p. 565-575.
41. Ali, J.B., N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, *Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals.* Applied Acoustics, 2015. **89**: p. 16-27.
42. Panda, A., S.M. Satapathy, and S.K. Rath, *Empirical validation of neural network models*

for agile software effort estimation based on story points. Procedia Computer Science, 2015. **57**: p. 772-781.

43. Liu, H., J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, *Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders.* ISA transactions, 2018. **77**: p. 167-178.
44. Shao, S., P. Wang, and R. Yan, *Generative adversarial networks for data augmentation in machine fault diagnosis.* Computers in Industry, 2019. **106**: p. 85-93.
45. Wen, L., X. Li, L. Gao, and Y. Zhang, *A new convolutional neural network-based data-driven fault diagnosis method.* IEEE Transactions on Industrial Electronics, 2017. **65**(7): p. 5990-5998.
46. Khan, S.A., M.D. Equbal, and T. Islam, *A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models.* IEEE Transactions on Dielectrics and Electrical Insulation, 2015. **22**(1): p. 590-596.
47. Kuai, M., G. Cheng, Y. Pang, and Y. Li, *Research of planetary gear fault diagnosis based on permutation entropy of CEEMDAN and ANFIS.* Sensors, 2018. **18**(3): p. 782.
48. Kari, T., W. Gao, D. Zhao, Z. Zhang, W. Mo, Y. Wang, and L. Luan, *An integrated method of ANFIS and Dempster-Shafer theory for fault diagnosis of power transformer.* IEEE Transactions on Dielectrics and Electrical Insulation, 2018. **25**(1): p. 360-371.
49. Mohamed, M.A., A.-A.A. Mohamed, M. Abdel-Nasser, E.E. Mohamed, and M.M. Hassan, *Induction motor broken rotor bar faults diagnosis using ANFIS-based DWT.* International Journal of Modelling and Simulation, 2020: p. 1-14.
50. Cerrada, M., G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, *Fault diagnosis in spur gears based on genetic algorithm and random forest.* Mechanical Systems and Signal Processing, 2016. **70**: p. 87-103.
51. Tyagi, S. and S. Panigrahi, *A hybrid genetic algorithm and back-propagation classifier for gearbox fault diagnosis.* Applied Artificial Intelligence, 2017. **31**(7-8): p. 593-612.
52. Huang, S., K.K. Tan, and T.H. Lee, *Fault diagnosis and fault-tolerant control in linear drives using the Kalman filter.* IEEE Transactions on Industrial Electronics, 2012. **59**(11): p. 4285-4292.
53. Zhang, Q., *Adaptive Kalman filter for actuator fault diagnosis.* Automatica, 2018. **93**: p. 333-342.
54. Cheng, Y. and H. Zhao. *Fault detection and diagnosis for railway switching points using fuzzy neural network*. in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*. 2015. IEEE.
55. Guclu, A., H. Yılboga, O.F. Eker, F. Camci, and I. Jennions. *Prognostics with autoregressive moving average for railway turnouts*. in *Annual Conference of the PHM Society*. 2010.
56. Marquez, F.P.G., D.J.P. Tercero, and F. Schmid, *Unobserved component models applied to the assessment of wear in railway points: A case study.* European Journal of Operational Research, 2007. **176**(3): p. 1703-1712.
57. Garga, A.K., K.T. McClintic, R.L. Campbell, Y. Chih-Chung, M.S. Lebold, T.A. Hay, and C.S. Byington. *Hybrid reasoning for prognostic learning in CBM systems*. in *2001 IEEE*

Aerospace Conference Proceedings (Cat. No.01TH8542). 2001.

58. Majidian, A. and M. Saidi, *Comparison of fuzzy logic and neural network in life prediction of boiler tubes.* International Journal of Fatigue, 2007. **29**(3): p. 489-498.
59. Melli, R. and E. Sciubba, *Diagnostics and Prognostics of Energy Conversion Processes via Knowledge-Based Systems.* 2020.
60. Chen, B., P.C. Matthews, and P.J. Tavner, *Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS.* Expert Systems with Applications, 2013. **40**(17): p. 6863-6876.
61. Li, N., Y. Lei, J. Lin, and S.X. Ding, *An improved exponential model for predicting remaining useful life of rolling element bearings.* IEEE Transactions on Industrial Electronics, 2015. **62**(12): p. 7762-7773.
62. Kundu, P., A.K. Darpe, and M.S. Kulkarni, *Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions.* Mechanical Systems and Signal Processing, 2019. **134**: p. 106302.
63. Van Noortwijk, J.M., *A survey of the application of gamma processes in maintenance.* Reliability Engineering & System Safety, 2009. **94**(1): p. 2-21.
64. Jiang, J., R. Chen, M. Chen, W. Wang, and C. Zhang, *Dynamic Fault Prediction of Power Transformers Based on Hidden Markov Model of Dissolved Gases Analysis.* IEEE Transactions on Power Delivery, 2019. **34**(4): p. 1393-1400.
65. Soualhi, A., H. Razik, G. Clerc, and D.D. Doan, *Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system.* IEEE Transactions on Industrial Electronics, 2013. **61**(6): p. 2864-2874.
66. Eker, O.F., F. Camci, A. Guclu, H. Yilboga, M. Sevkli, and S. Baskan, *A simple state-based prognostic model for railway turnout systems.* IEEE Transactions on Industrial Electronics, 2011. **58**(5): p. 1718-1726.
67. Eker, O.F. and F. Camci, *State-based prognostics with state duration information.* Quality and Reliability Engineering International, 2013. **29**(4): p. 465-476.
68. Wu, Q., K. Ding, and B. Huang, *Approach for fault prognosis using recurrent neural network.* Journal of Intelligent Manufacturing, 2018: p. 1-13.
69. Khan, F., O.F. Eker, I.K. Jennions, and A. Tsourdos. *Prognostics of crack propagation in structures using time delay neural network*. in *2015 IEEE Conference on Prognostics and Health Management (PHM)*. 2015.
70. Hong, J., Z. Wang, and Y. Yao, *Fault prognosis of battery system based on accurate voltage abnormity prognosis using long short-term memory neural networks.* Applied Energy, 2019. **251**: p. 113381.
71. Luo, J., A. Bixby, K. Pattipati, L. Qiao, M. Kawamoto, and S. Chigusa. *An interacting multiple model approach to model-based prognostics*. in *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*. 2003. IEEE.
72. Ray, A. and S. Tangirala, *Stochastic modeling of fatigue crack dynamics for on-line failure prognostics.* IEEE Transactions on Control Systems Technology, 1996. **4**(4): p.

443-451.
73. *Railroad switch*. 2020 8 November 2020 19:58 UTC; Available from: https://en.wikipedia.org/w/index.php?title=Railroad_switch&oldid=987712866.
74. Cope, D. and J. Ellis, *British railway track 7th edition, Volume 5: Switch and crossing maintenance.* Derby, UK: The Permanent Way Institution, 2002.
75. IndiaMART, *Railway Signaling & Coach Applications - Electric Point Machines Manufacturer from Indore*.
76. *In2Rail Innovative Intelligent Rail: Deliverable D2.1 Development of Novel S&C Motion/Locking Mechanisms: Design Concept Report*. 2017, In2Rail.
77. Zhou, F., N. Archer, J. Bowles, M. Duta, M. Henry, M. Tombs, M. Zamora, S. Baker, and C. Burton, *Remote condition monitoring and validation of railway points.* Computing & Control Engineering Journal, 2002. **13**(5): p. 221-230.
78. Camci, F., O.F. Eker, S. Başkan, and S. Konur, *Comparison of sensors and methodologies for effective prognostics on railway turnout systems.* Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2016. **230**(1): p. 24-42.
79. Silmon, J., *Understanding the potential for points condition monitoring as part of an intelligent infrastructure initiative.* Network Rail internal report, 2008.
80. AlShorman, O., M. Irfan, N. Saad, D. Zhen, N. Haider, A. Glowacz, and A. AlShorman, *A review of artificial intelligence methods for condition monitoring and fault diagnosis of rolling element bearings for induction motor.* Shock and Vibration, 2020. **2020**.
81. Kaplan, K., Y. Kaya, M. Kuncan, M.R. Minaz, and H.M. Ertunç, *An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis.* Applied Soft Computing, 2020. **87**: p. 106019.
82. Da Silva, P.R.N., H.A. Gabbar, P.V. Junior, and C.T. da Costa Junior, *A new methodology for multiple incipient fault diagnosis in transmission lines using QTA and Naïve Bayes classifier.* International Journal of Electrical Power & Energy Systems, 2018. **103**: p. 326-346.
83. Jan, S.U., Y.-D. Lee, J. Shin, and I. Koo, *Sensor fault classification based on support vector machine and statistical time-domain features.* IEEE Access, 2017. **5**: p. 8682-8690.
84. Verma, N.K., R.K. Sevakula, S. Dixit, and A. Salour, *Intelligent condition based monitoring using acoustic signals for air compressors.* IEEE Transactions on Reliability, 2015. **65**(1): p. 291-309.
85. Xu, Y., Y. Deng, J. Zhao, W. Tian, and C. Ma, *A Novel Rolling Bearing Fault Diagnosis Method Based on Empirical Wavelet Transform and Spectral Trend.* IEEE Transactions on Instrumentation and Measurement, 2019. **69**(6): p. 2891-2904.
86. Liang, P., C. Deng, J. Wu, G. Li, Z. Yang, and Y. Wang, *Intelligent Fault Diagnosis via Semisupervised Generative Adversarial Nets and Wavelet Transform.* IEEE Transactions on Instrumentation and Measurement, 2019. **69**(7): p. 4659-4671.
87. Asada, T., C. Roberts, and T. Koseki, *An algorithm for improved performance of railway condition monitoring equipment: Alternating-current point machine case study.* Transportation Research Part C: Emerging Technologies, 2013. **30**: p. 81-92.

88. Sheikh, J.A., S. Akhtar, S.A. Parah, and G.M. Bhat, *A new method of speech transmission over space time block coded co-operative MIMO–OFDM networks using time and space diversity.* International Journal of Speech Technology, 2018. **21**(1): p. 65-77.
89. Singh, P., G. Pradhan, and S. Shahnawazuddin, *Denoising of ECG signal by non-local estimation of approximation coefficients in DWT.* Biocybernetics and Biomedical Engineering, 2017. **37**(3): p. 599-610.
90. Saravanan, N. and K. Ramachandran, *Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN).* Expert Systems with Applications, 2010. **37**(6): p. 4168-4181.
91. He, W., Y. He, and B. Li, *Generative adversarial networks with comprehensive wavelet feature for fault diagnosis of analog circuits.* IEEE Transactions on Instrumentation and Measurement, 2020. **69**(9): p. 6640-6650.
92. Jing, L., M. Zhao, P. Li, and X. Xu, *A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox.* Measurement, 2017. **111**: p. 1-10.
93. Khan, S. and T. Yairi, *A review on the application of deep learning in system health management.* Mechanical Systems and Signal Processing, 2018. **107**: p. 241-265.
94. Guo, J., Z. Li, and M. Li, *A review on prognostics methods for engineering systems.* IEEE Transactions on Reliability, 2019. **69**(3): p. 1110-1129.
95. Liao, L. and F. Köttig, *Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction.* IEEE Transactions on Reliability, 2014. **63**(1): p. 191-207.
96. Lo, C.-C., C.-H. Lee, and W.-C. Huang, *Prognosis of bearing and gear wears using convolutional neural network with hybrid loss function.* Sensors, 2020. **20**(12): p. 3539.
97. Gebraeel, N., A. Elwany, and J. Pan, *Residual life predictions in the absence of prior degradation knowledge.* IEEE Transactions on Reliability, 2009. **58**(1): p. 106-117.
98. Ujjwal Maulik, S.B., *Performance Evaluation of some clustering algorithms and validity indices.* IEEE transactions on pattern analysis and machine intelligence, 2002.

APPENDIX A PUBLISHED PAPERS

During the PhD study course, a list of papers has been published as follows:

1. Chen, Q., Nicholson, G., Ye, J. and Roberts, C., 2020, May. Fault Diagnosis Using Discrete Wavelet Transform (DWT) and Artificial Neural Network (ANN) for A Railway Switch. In 2020 Prognostics and Health Management Conference (PHM-Besançon) (pp. 67-71). IEEE.

2. Chen, Q., Nicholson, G., Roberts, C., Ye, J. and Zhao, Y., 2020. Improved Fault Diagnosis of Railway Switch System Using Energy-Based Thresholding Wavelets (EBTW) and Neural Networks. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-12.

3. Chen, Q., Nicholson, G., Ye, J., Zhao, Y. and Roberts, C., 2020. Estimating Residual Life Distributions of Complex Operational Systems Using a Remaining Maintenance Free Operating Period (RMFOP)-Based Methodology. Sensors, 20(19), p.5504.