Prognostics Under the Conditions of Limited Failure Data Availability

Gishan Don Ranasinghe

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Fitzwilliam College August 2021

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, footnotes, tables and equations and has fewer than 150 figures.

Gishan Don Ranasinghe
August 2021

Abstract

Prognostics Under the Conditions of Limited Failure Data Availability
Gishan Don Ranasinghe

With the scale and need for industrial operations increasing, industrial systems are becoming far more complex than they had ever been. As a consequence, traditional maintenance strategies are not only failing to prevent unplanned downtime, they introduce additional costs due to over-maintenance and false alarms. To overcome this challenge, predictive maintenance aims to deliver on the promise of an optimal maintenance policy that balances the cost of maintenance, the risk of failure and performance of the equipment. To this end, it aims to produce equipment prognostics with minimal bias and uncertainty, and then plan and schedule predictive maintenance actions to prevent the unexpected consequences of failures: unplanned downtime and collateral damage to equipment and processes.

Despite its popularity, the data-driven prognostics approach has been unsuccessful in enabling the effective implementation of predictive maintenance due to the limited failure data availability for prognostics. The handful of existing techniques available for addressing this problem has been unsuccessful since they either duplicate existing failure data or randomly generate data, hence the fundamental lack of failure data issue is not addressed. When failure data are limited for data-driven prognostics, the use of physics model-based and knowledge-based prognostics have also been unsuccessful in most industrial scenarios since it is difficult and expensive to develop them due to the stochasticity and complexity of degradation processes. Hence, a systematic approach for prognostics modelling under the conditions of limited failure data availability remains a research gap.

The aim of this thesis is to address the aforementioned research gap, and thus enable the effective implementation of predictive maintenance for industrial organisations. To this end, the thesis presents a methodology that allows developing prognostics models under the conditions of limited failure data availability. The methodology integrates sound engineering knowledge of failure mechanisms, the limited amount of failure data samples available and noise into a conditional generative modelling framework to estimate a generative model. In contrast to the existing techniques which duplicate existing failure data or randomly generate data, this generative model is capable of generating new and plausible failure data samples.

The thesis concludes that the methodology outperforms existing techniques used in the literature for prognostics under the conditions of limited failure data availability by a large margin. More specifically, the methodology increases the prognostics performance by approximately 40% compared to the existing techniques. Thus, it enables the effective implementation of predictive maintenance for industrial organisations allowing them to prevent costs due to under-maintenance and over-maintenance of equipment and false alarms.

I would like to dedicate this thesis first and foremost to my wife and my parents. Ruvini, you have been wonderful and immensely supportive, and I know you gave me the priority over the things you wanted to do in life. For this, I am ever so grateful. Father, thank you for hiding your happiness every time I got high marks for the exams and complaining about the remaining marks. This kind of thinking made me who I am and drives me to this day to achieve more in life. To my wonderful mother, thank you for teaching me how to be a gentle and loving person.

Secondly, to my brothers, Sanjaya and Ashan, I am lucky to have you both and thank you for supporting me in the time of need, and for being two wonderful brothers.

To my best friends, Jorge, Sangeet, Mike, Samy and Freeman, I am ever so grateful for your support during the hardships I had, which I will never forget.

To Dr Ajith Kumar Parlikad, for being a flawless and flexible supervisor. Your constructive feedback has always been immensely helpful and it has been a wonderful experience working with you.

To Prof Duncan McFarlane and Prof Mark Girolami, for all the insightful discussions and feedback you gave me during this PhD.

To Prof Dario Landa-Silva and Prof Rong Qu, for your invaluable guidance during my masters and bachelors.

I would also like to dedicate this thesis to people all around the world who lost their lives and loved ones due to the COVID-19 pandemic.

...

Acknowledgements

This PhD Thesis is supported by the Engineering and Physical Sciences Research Council through a Doctoral Training Partnership Grant, with reference EP/M508007/1.

The support of British Telecom (via the Next Generation Converged Digital Infrastructure project) and Scania Commercial Vehicles who contributed useful data and information to the case studies of this thesis is acknowledged.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

This thesis focuses on addressing the problem of equipment prognostics under the conditions of limited failure data availability and thus enabling the effective implementation of predictive maintenance (PdM) for industrial organisations. To address this problem, the thesis presents a methodology that integrates sound engineering knowledge of failure mechanisms (hereinafter referred to as auxiliary information) and the limited amount of failure data samples available to estimate a model that is capable of generating new and plausible failure data for prognostics under the conditions of limited failure data availability.

The auxiliary information is identified from expert knowledge provided by maintenance and reliability engineers, component designers, equipment operators and applied research scientists, and includes expert knowledge of the effect of environmental conditions on failure mechanisms, the effect of harsh use of equipment on failure mechanisms and similarity between equipment that has failed under a single failure mode. These auxiliary information are used to condition the noise being added to newly generated failure data samples, making the failure data generation process controlled and directed. The estimated data generation model is then used to produce new and plausible failure data samples, which allows augmenting the historical datasets used for prognostics modelling to include an increased amount of failure data samples for estimating effective prognostics models under the conditions of limited failure data availability.

In this introductory chapter, an overview of the background and motivation of the research presented in this thesis is provided first. Then a description of the problem of prognostics under the conditions of limited failure data availability is provided. The chapter is concluded by stating the research questions, research methodology and the outline of the thesis.

1.1 Prognostics and effective implementation of predictive maintenance

Prognostics involves predicting the time-to-failure (TTF) of equipment or predicting the probability that equipment operates without failure up to some future time (e.g. until the next inspection time or the end of the current mission window) [1]. Prognostics approaches can be classified as physics model-based, knowledge-based, data-driven and hybrid prognostics [1]. The data-driven prognostics approach has become popular over the other prognostics approaches since it does not require explicit mathematical representations or detailed knowledge bases to estimate prognostics models [2, 3].

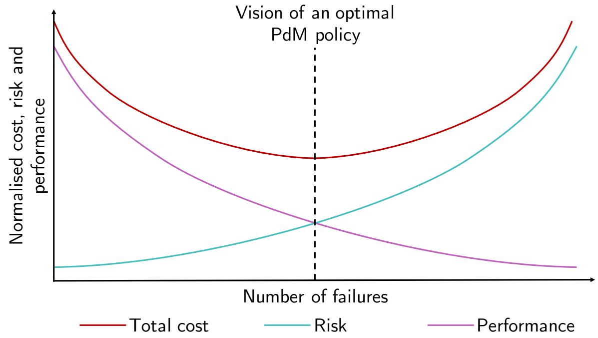


Fig. 1.1 Diagram depicting an optimal PdM policy.

One of the key applications of prognostics is PdM [4]. PdM involves first producing equipment prognostics with minimal bias and uncertainty, and then planning and scheduling PdM actions to prevent the unexpected consequences of failures (e.g. unplanned downtime and collateral damage to equipment and processes) [4]. When implemented effectively, PdM can deliver on the promise of an optimal maintenance policy that balances the cost of maintenance, the risk of failure and performance of the equipment [5]. This is achieved by finding the optimal balance between the cost of maintenance, the risk of failure and performance of the equipment as shown in Fig. 1.1. The effective implementation of PdM also ensures that the equipment is always in ready and reliable condition whilst producing several cost savings: minimises production hours lost to unplanned downtime, minimises costs of spare parts, supplies and labour, and prevents collateral damage to equipment and processes [5].

1.2 Problem of prognostics under limited failure data availability

Despite its popularity, the long-lasting problem with data-driven prognostics is that it relies on large amounts of historical failure data to estimate prognostics models [2]. Nevertheless, historical failure data are limited in industrial scenarios due to two major reasons: (i) overprotective maintenance and replacement regimes; (ii) highly reliable equipment [6, 7]. These reasons cause failures to become rare (i.e. infrequent occurrence of failures under a single failure mode), and thus cause historical datasets used for prognostics modelling to consist of a limited amount of failure data samples. This makes it difficult for data-driven algorithms to estimate prognostics model parameters (i.e. model weights and biases) from degradation patterns and characterise system performance for prognostics modelling [7]. Hence, predictions produced by these models are associated with high bias and uncertainty. When these prognostics predictions are incorporated into a PdM policy, that policy will be far from an optimal maintenance policy and therefore introduces high costs due to under-maintenance and over-maintenance of equipment and false alarms [7]. In other words, the limited failure data availability problem leads to the ineffective implementation of PdM.

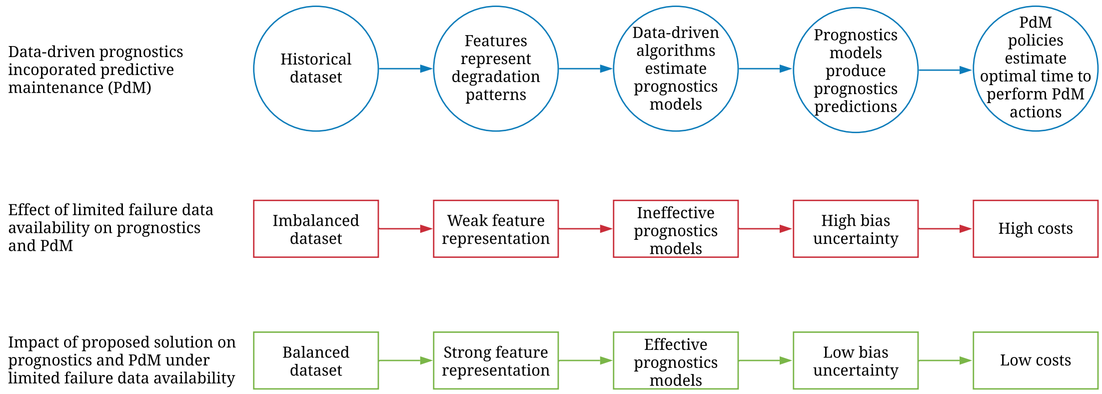


Fig. 1.2 Diagram depicting the effect of the problem of limited failure data availability for prognostics and the impact of the solution presented in this thesis on the effective implementation of PdM.

The problem of prognostics under the conditions of limited failure data availability, its effect on the effective implementation of PdM and the impact of the solution presented in this thesis are illustrated Fig. 1.2. When failure data are limited, historical datasets used for prognostics modelling become imbalanced [8]. In the prognostics domain, this means there is a relatively large number of normal data samples (i.e. data pertaining to the normal condition of equipment) compared to the failure data samples (i.e. data pertaining to the failure condition of equipment). Data imbalance causes features and samples in datasets to weakly represent degradation patterns, which makes it difficult for data-driven algorithms used for prognostics modelling (i.e. statistical and machine learning algorithms) to extract degradation patterns and characterise equipment performance for estimating effective prognostics models [7]. The prognostics predictions produced by these models are associated with high bias and uncertainty, and when incorporated into a PdM model, the time to perform PdM actions produced by the PdM model is far from optimal. Hence when failure data are limited, it is no longer possible to effectively plan and schedule PdM actions. This leads to high costs due to under-maintenance and over-maintenance of equipment and false alarms for industrial organisations [9, 8].

The solution presented in this thesis is capable of generating new and plausible failure data samples for treating the data imbalance in historical datasets used for prognostics modelling. Thus, the methodology allows developing effective prognostics models under the conditions of limited failure data availability to enable the effective implementation of PdM for industrial organisations.

In the remainder of this section, a concrete description of the problem of prognostics under the conditions of limited failure data availability is provided.

Problem description and statement

There have been a few techniques proposed in the literature for addressing the limited failure data availability problem for data-driven prognostics. They can be classified as data level and algorithmic level techniques [10]. Oversampling, undersampling and generative modelling-based data augmentation are data level techniques and have been unsuccessful in addressing the problem sufficiently since they either duplicate existing failure data or randomly generate data, and hence do not introduce new and plausible failure data samples to treat the data imbalance [11, 10]. The major shortcoming of all the algorithmic level techniques (i.e. cost-sensitive learning, ensemble learning and grey prediction model) is that they do not increase the quantity of failure data samples, hence the fundamental lack of failure data issue is not addressed [11, 10].

When failure data are limited for data-driven prognostics, the use of physics model-based, knowledge-based and hybrid prognostics approaches have been unsuccessful in most industrial scenarios. More specifically, once a physics model-based prognostics model is developed, it needs to be validated prior to its application using large amounts of historical failure data pertaining to the failure mode that needs predicting [2, 12]. However, under the conditions of limited failure data availability, it is not possible to obtain large amounts of failure data needed for validating these prognostics models. Moreover, physics-model-based prognostics models are based on the assumption that system behaviour can be described analytically and accurately which makes it difficult and expensive to develop them for equipment prognostics since equipment degradation is too stochastic and complex to model [2, 13].

Knowledge-based prognostics involves obtaining domain knowledge and converting it into rules which is also difficult in most industrial scenarios due to the stochasticity and complexity of equipment degradation processes [2]. And once built, they do not generalise into new situations that are not covered explicitly in their knowledge bases [2]. Even though the hybrid approach aims to leverage the strengths of multiple prognostics approaches, it has also been unsuccessful since under the conditions of limited failure data availability, this prognostics approach inherits the aforementioned limitations of data-driven, physics model-based and knowledge-based approaches.

It can be observe that even though there is an increasing motivation in the industry to implement PdM effectively, it has so far been stalled due to the absence of a systematic approach to developing prognostics models under the conditions of limited failure data availability. More specifically, a systematic approach for prognostics modelling under the conditions of limited failure availability is a research gap. This thesis aims to address this research gap and thus allow industrial organisations to implement PdM effectively for preventing costs due to under-maintenance and over-maintenance of equipment and false alarms. The problem statement of this thesis follows from this problem description:

Problem statement. To devise and implement a systematic and pragmatic approach for prognostics modelling under the conditions of limited failure data availability.

1.3 Research questions

In order to address the problem of prognostics under the conditions of limited failure data availability, this thesis answers the following two research questions:

- **Research Question 1:** How can equipment failure be predicted under the conditions of limited failure data availability?
  This research question aims to develop a methodology that allows generating new and plausible failure data for prognostics under the conditions of limited failure data availability.

- **Research Question 2:** What impact does the proposed solution have on prognostics compared to existing techniques under the conditions of limited failure data availability?
  This research question aims to quantify the impact of methodology in terms of improved prognostics performance under the conditions of limited failure data availability compared to the state-of-the-art techniques proposed in the literature, and produce insights into the key factors influencing the effectiveness of the methodology.

1.4 Research methodology

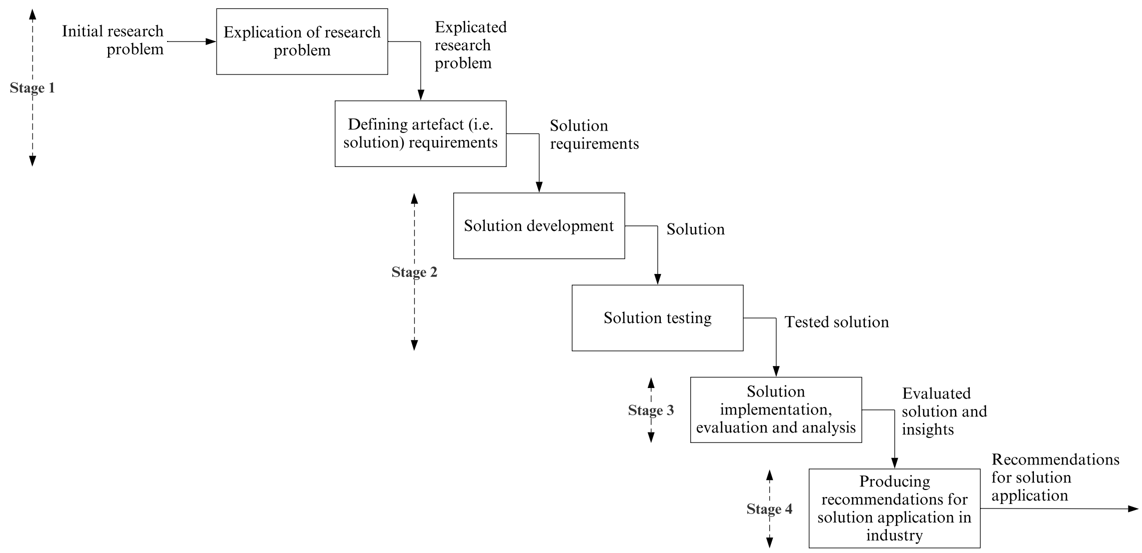


Fig. 1.3 Diagram depicting the stages of research methodology. This research methodology is based on the framework for design science research proposed in [14].

In order to address the research questions, the research methodology outlined in Fig. 1.3 is followed. First, the research problem is formulated by identifying the research gap, solution requirements and industrial rationale using findings generated from a systematic literature review and an industry review as shown in Fig. 1.4. In the literature review, prognostics approaches and methods are reviewed to identify the state-of-the-art of prognostics research field. Then the topic of prognostics under the conditions of limited failure data availability is reviewed to identify existing techniques used to address the problem and reasons for them to be unsuccessful.

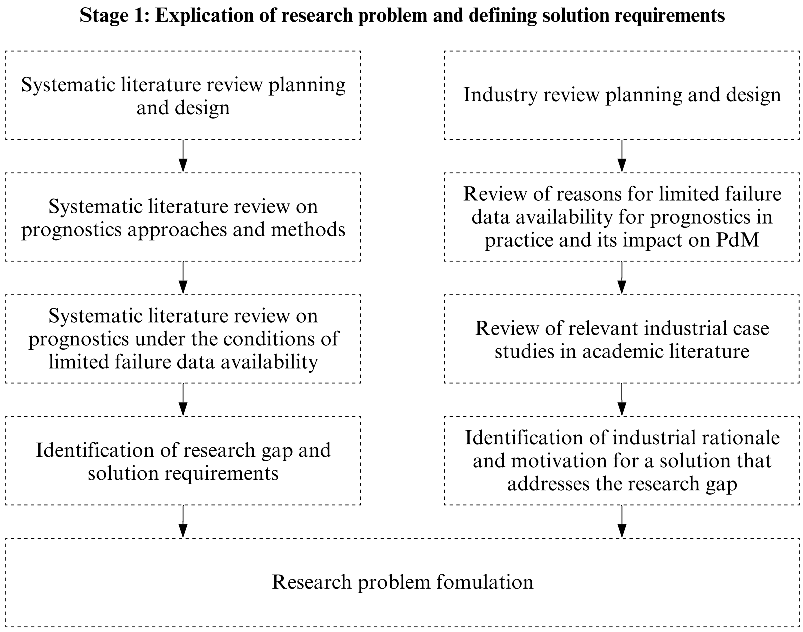


Fig. 1.4 Diagram depicting the elements composing the Stage 1 of research methodology.

The solution which addresses the research problem is developed in Stage 2 (see Fig. 1.5). In Stage 3, the solution is implemented in three industrial scenarios that have the problem of limited failure data availability for prognostics: BT residential broadband line prognostics, Scania heavy-truck air processing system prognostics, and Scania heavy-truck turbocharger prognostics (see Fig. 1.6). At the end of this stage, insights are produced to identify the impact of solution on prognostics compared to the existing techniques under the conditions of limited failure data availability, and to identify key factors influencing the effectiveness of the solution.

As shown in Fig. 1.7, Stage 4 of the research methodology involves producing recommendations for solution application in the industry. These recommendations are produced using the insights gained from the analyses performed in Stage 2 and Stage 3. The recommendations include the strengths and weaknesses of the solution, risks and potential risk mitigation actions and application criteria that allow identifying suitable and unsuitable industrial applications for the solution.

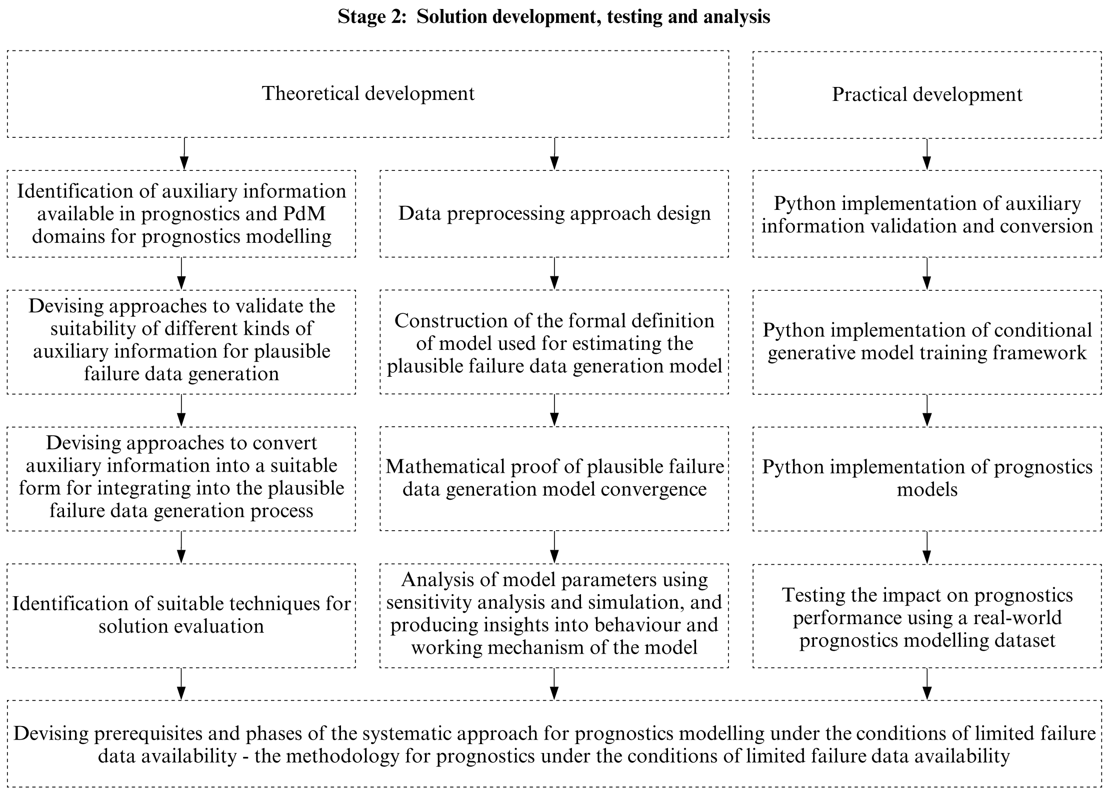


Fig. 1.5 Diagram depicting the elements composing the Stage 2 of research methodology.

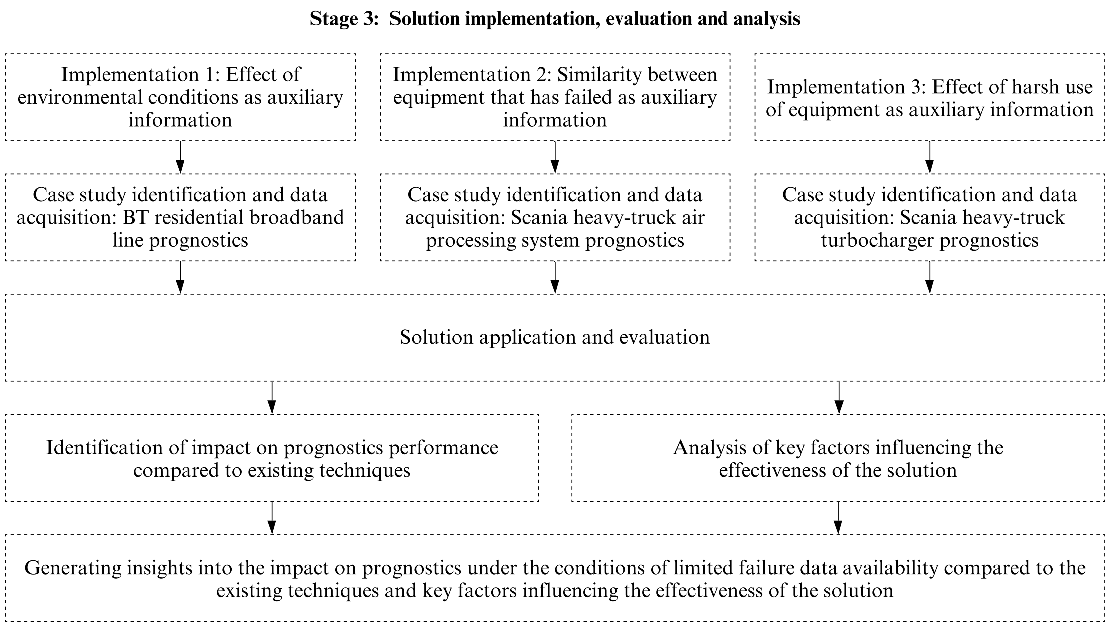


Fig. 1.6 Diagram depicting the elements composing the Stage 3 of research methodology.

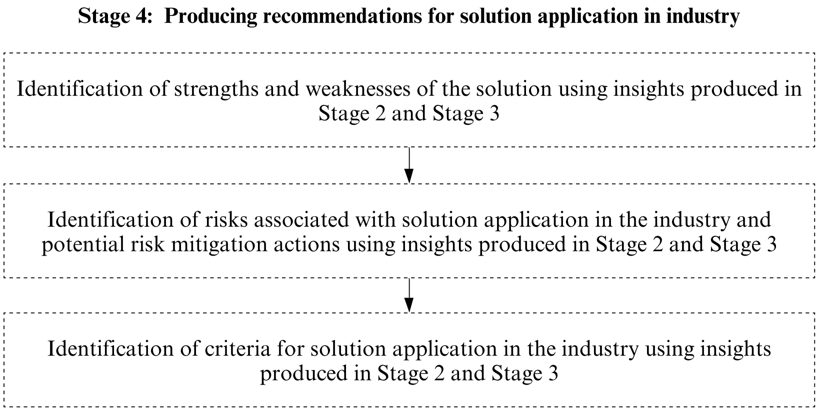


Fig. 1.7 Diagram depicting the elements composing the Stage 4 of research methodology.

1.5 Thesis outline

This thesis consists of six chapters. The first chapter outlines the purpose of this thesis, and presents the research methodology and the two research questions this thesis will address. The second chapter contains the research background, including a systematic literature review and an industry review.

The third chapter answers the first research question of this thesis by presenting a methodology for prognostics under the conditions of limited failure data availability. Moreover, insights are generated from a theoretical perspective to identify the behaviour and working mechanism of the methodology in this chapter. The fourth chapter presents implementations of the methodology using industrial case studies. This chapter also answers the second research question of this thesis by quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques proposed in the literature, and by producing insights into the key factors influencing the effectiveness of the methodology.

The fifth chapter uses the insights generated in the third and fourth chapters to provide recommendations for methodology application in the industry. The last chapter presents conclusions, awards and recognition the research has received in academia and industry, and recommendations for future research.

Chapter 2

Research background

In this chapter, findings from the literature review of this thesis are used to provide discussions on the state-of-the-art of prognostics research field, existing techniques used to address the problem of limited failure data availability for prognostics and reasons for them to be unsuccessful. In addition to these discussions which show that the research presented in this thesis is addressing a research gap, the importance of addressing the research gap for industrial practice is discussed using the findings obtained from an industry review.

2.1 Literature review

The literature review is performed using the methodology for the systematic literature review proposed in [15]. In this section, the planning protocol used for the systematic literature review of this thesis is outlined first. The publications selected after executing the planning protocol are used to review the state-of-the-art of prognostics research field, existing techniques used to address the problem of limited failure data availability for prognostics and reasons for them to be unsuccessful. The review is concluded by stating the research gap the research presented in this thesis aims to address.

2.1.1 Systematic literature review planning and execution

The systematic literature review planning protocol consists of a set of literature review questions, literature search databases and exclusion criteria [15]. The literature review questions set out the framework in which the literature is searched, literature search databases are the knowledge sources used for the literature search, and the exclusion criteria define what publications will be excluded from the review. The planning protocol designed for this thesis is executed once every 6 months of the PhD research and the findings are accumulated over time.[1] In the following, the design and execution of the planning protocol used in this thesis are summarised.

[1] During the PhD research, a systematic literature review is conducted once every 6 months. However, automatic alerts are set to receive notifications about new and relevant publications to keep up-to-date with the research field at all times.

Planning protocol

The planning protocol consists of the below literature review questions (LRQ). LRQ-1 aims to identify the state-of-the-art of prognostic research field. LRQ-2 aims to identify the existing techniques used to address the problem of limited failure data availability for prognostics and the reasons for them to be unsuccessful.

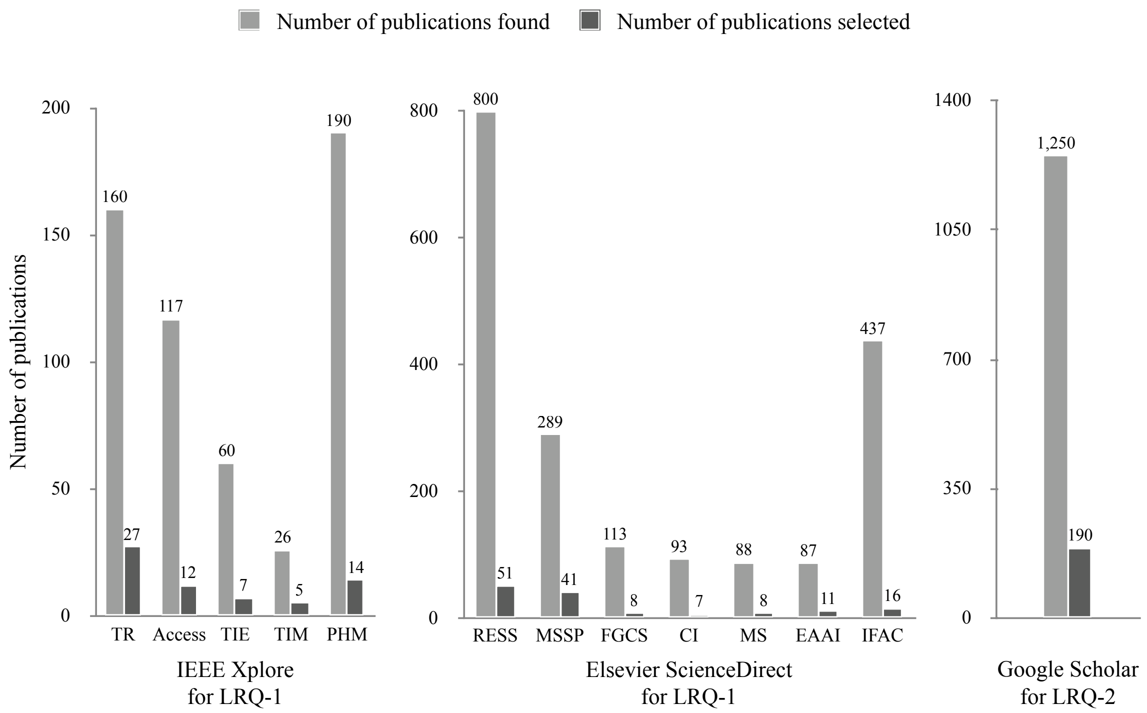
- LRQ-1: What are the prognostics approaches and predictive modelling methods (i.e. first principles, statistical and machine learning methods) used in each approach?
- LRQ-2: What are the existing techniques used to address the problem of limited failure data availability for prognostics and why have they been unsuccessful?

IEEE Xplore and ScienceDirect databases are used for searching the literature for answering LRQ-1 since they provide sufficient literature for reviewing the state-of-the-art of prognostics research field. Google Scholar is used for searching the literature for answering LRQ-2 since it allows surveying over multiple literature databases and the limited failure data availability problem reduces the search space which makes it feasible to survey over multiple literature databases. The exclusion criteria are: works that do not present any type of experimentation or comparison results and make only propositions; publications outside the first 50% of the most cited publications per year.

In addition to the exclusion criteria, for LRQ-1, a set of journals and conferences which is the most suitable for answering this literature review question is selected based on the guidance received from the peers. IEEE Access, Transactions on Reliability, Industrial Electronics, Instrumentation and Measurement journals and the conference paper archive of the International Conference on Prognostics and Health Management are selected for the IEEE Xplore literature search. Reliability Engineering and System Safety, Mechanical Systems and Signal Processing, Engineering Applications of Artificial Intelligence, Computers in Industry, Future Generation Computer Systems and Manufacturing Systems journals and the IFAC conference paper archive are selected for the ScienceDirect literature search. Even though these journals and conferences are the core sources used for literature search, suitable literature from other journals and conferences are also used in the literature review.

The planning protocol is executed using the search strings (SS) outlined below and they are developed using keywords that are commonly found in the literature and related to the literature review of this thesis. SS-1 is for searching the literature published on the aforementioned IEEE and ScienceDirect journals for answering LRQ-1. SS-2 is for searching the literature using Google Scholar for answering LRQ-2. The number of publications found using the search strings and the number of publications selected after applying the exclusion criteria are shown in Fig. 2.1. In the following sections, the selected set of publications is used to review the state-of-the-art of prognostics research field, and existing techniques used to address the problem of limited failure data availability for prognostics and reasons for them to be unsuccessful.

- SS-1: *"Abstract": "prognostics" OR "failure prediction" OR "remaining useful life" OR "time-to-failure" OR "predictive maintenance"*
- SS-2: *"Abstract": ("prognostics" OR "failure prediction" OR "remaining useful life" OR "time-to-failure" OR "predictive maintenance") AND ("data imbalance" OR "class imbalance" OR "limited data" OR "lack of data" OR "limited failure data" OR "lack of failure data" OR "small-sample size" OR "rare failures" OR "rare events")*



Key: TR=Transactions on Reliability; Access=IEEE Access; TIE=Transactions on Industrial Electronics; TIM=Transactions on Instrumentation and Measurement; PHM=Prognostics and Health Management Conference Paper Archive; RESS=Reliability Engineering and System Safety; MSSP=Mechanical Systems and Signal Processing; FGCS=Future Generation Computer Systems; CI=Computers in Industry; MS=Journal of Manufacturing Systems; EAAI=Engineering Applications of Artificial Intelligence; IFAC=IFAC Conference Paper Archive.

Fig. 2.1 Number of publications found and selected for literature review using the planning protocol.

2.1.2 Prognostics approaches and methods

Fig. 2.2 shows the importance of the prognostics research field, and the popularity of prognostics approaches (i.e. physics model-based prognostics, knowledge-based prognostics, data-driven prognostics and hybrid prognostics). It can be observed that the importance of the prognostics research field is growing, and the data-driven prognostics approach is becoming increasingly popular compared to other prognostics approaches.

In this section, prognostics approaches and predictive modelling methods (i.e. first principles, statistical and machine learning methods) used in each approach are reviewed.

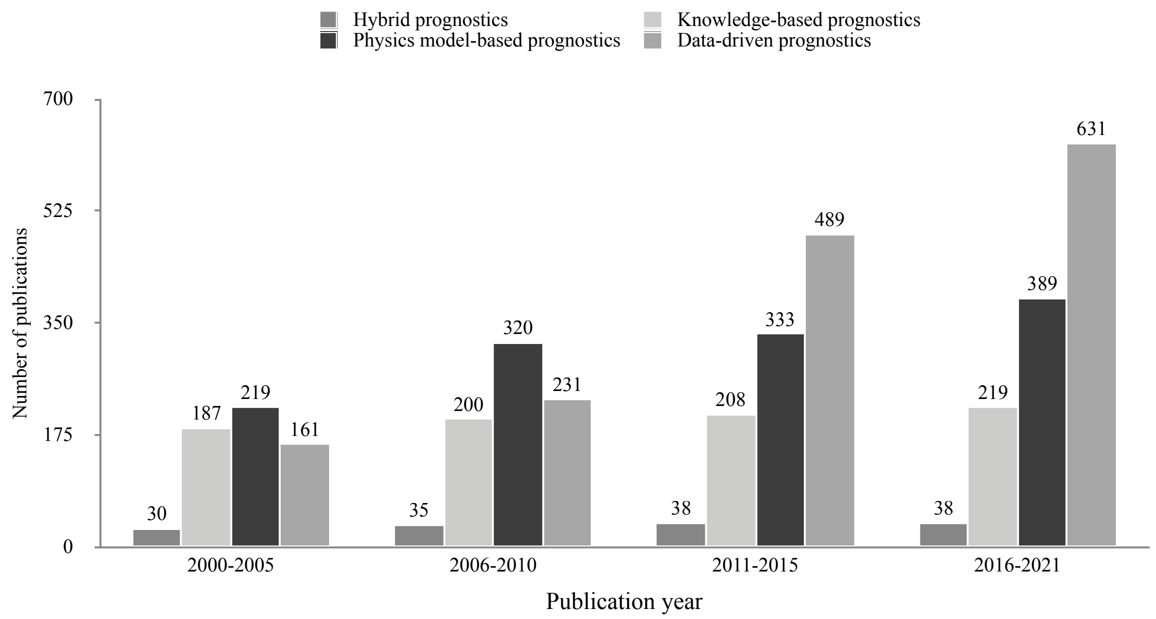


Fig. 2.2 Plot summarising importance of prognostics research field, and increasing popularity of data-driven prognostics. The publications found using the planning protocol are used in this analysis.

Physics model-based prognostics

Physics model-based prognostics uses explicit mathematical representations to create degradation models by formalising the physical understanding of failure modes that need predicting [2]. The physical understanding is based on the principle that failure occurs from fundamental processes (i.e. electrical, chemical, mechanical, thermal and radiation) [16]. Once a degradation model is formulated, its parameters are calculated using an estimation algorithm (e.g. Kalman filter, extended Kalman filter and particle filter). Prior to application, the model is validated on large amounts of historical data pertaining to the failure mode that needs predicting [2, 12]. The degradation model is then combined with condition monitoring data and/or event data captured from the equipment to predict the future degradation behaviour, and thus estimate prognostics [2]. The popular physics model-based prognostics models include the following [12]:

Paris fatigue crack growth model. This model is used to calculate fatigue crack growth in materials due to cyclic loads [17]. It is usually combined with state estimation which is used to project material crack growth into steps ahead, and thus estimate the future distribution of a health indicator to produce prognostics [18]. Before employing this model for material fatigue crack growth prediction, one needs to identify whether the fatigue crack growth in materials under the study is governed by the Paris law [19].

Archard wear model. This model is used for describing material mass loss due to abrasive wear [20]. It is popular for modelling blade mass loss caused by blade interaction with shot particles in shot blasting turbines [20]. According to the Archard law, each time a shot particle enters a turbine it travels along the turbine blade length removing a small layer of blade material of a certain volume.

In Boškoski et al. [20], abrasive wear in shot blasting turbine blades is measured indirectly using vibration produced by rotating blades. The vibration is caused when the blades are rotating and different volumes of material being removed from the blades. This relationship between abrasive wear in turbine blades (defined by the Archard law) and vibration produced when the blades are rotating is modelled as a stochastic process, and the turbine failure is predicted using a hidden Markov model.

Bearing L10 model. This model describes a 10% probability of failure (or a 90% probability of survival) for a bearing unit in terms of millions of revolutions or hours of operation [21]. For instance, Ewing et al. [22] employed the bearing L10 model to predict the TTF of a pitch system bearing unit in a tidal stream turbine. First, tidal flow data are measured and fed into a turbine hydrodynamic model to generate synthetic loading regimes. The loading regimes are then used as the input to the Bearing L10 model to predict the remaining hours of operation as probability of failure increases.

Knowledge-based prognostics

In contrast to physics model-based prognostics, knowledge-based prognostics does not require explicit mathematical representations of degradation processes for prognostics modelling [2]. Instead, it involves assessing the similarity between an observed situation and a knowledge base of historical failures to estimate prognostics [2]. Typical predictive modelling methods used for knowledge-based prognostics are expert systems and fuzzy logic [2].

Expert systems. These are software programs that simulate how a human expert do the thinking and inference in solving a particular domain problem [2]. An expert system stores domain knowledge extracted by human experts into software programs as rules. The rules are formulated as precise IF-THEN statements (i.e. classical predicate logic) and are based on heuristic facts acquired by the human experts over a number of years. Creating an expert system for prognostics involves acquiring knowledge pertaining to the failure mode, representing the acquired knowledge in classical predicate logic and model verification and validation [12]. There are also open standards and domain-specific models that can be used to design expert systems for prognostics: MIMOSA open system architecture standard, Failure Mode and Effect Analysis (FMEA) and Fault Tree Analysis (FTA) [23, 24]. FMEA and FTA allow reviewing components, assemblies and subsystems to identify potential failure modes in an industrial system and their causes and effects [24]. Thus, they provide knowledge on failure modes and failure mechanisms for designing expert systems for prognostics.

Fuzzy logic. This is a superset of conventional Boolean logic and can be used to incorporate imprecise and qualitative reasoning statements with knowledge bases to produce simpler and intuitive predictive models compared to the expert systems [25]. The knowledge bases in expert systems are developed using classical predicate logic in which a statement can be either true or false. However, it is not always feasible to represent knowledge so precisely, especially for equipment prognostics since knowledge related to equipment degradation is too complex to acquire and represent [12]. In this case, knowledge needs to be represented in linguistic terms and this is enabled by the fuzzy logic. A typical fuzzy process logic statement may look like the following: *IF (process is too hot) AND (process* *is heating rapidly) THEN (cool the process quickly).* Developing a fuzzy logic-based prognostics model involves creating a library of reference degradation trajectories using historical failure instances. Then the TTF is predicted using fuzzy similarity analysis in which data pertaining to an ongoing equipment degradation are compared to the reference trajectory patterns, and by aggregating the TTF of reference trajectories in a weighted sum which accounts for the similarity between patterns in reference trajectories and patterns in data captured from the ongoing degradation process [26].

Data-driven prognostics

In contrast to physics model-based and knowledge-based approaches which estimate prognostics models from explicit mathematical representations and detailed knowledge bases respectively, data-driven approaches estimate prognostics models directly from large amounts of historical condition monitoring and/or event data pertaining to the failure modes [2]. According to the literature, they can be classified into statistical, error-based learning, instance-based learning and probabilistic machine learning approaches [16, 3].

(I) Statistical approach

The statistical approach to prognostics modelling involves establishing a statistical model based on empirical knowledge and presenting prognostics results as a conditional probability density function [3]. The prognostics are estimated by fitting past observations into random coefficient or stochastic process models under a probabilistic method without relying on physics and expert rules. Statistical methods used for prognostics modelling include autoregression, random coefficient regression, Wiener process, gamma process, Markov models and proportional hazards models [3].

Autoregression. Autoregressive models estimate future values using a linear function and random errors, hence they are widely used for time series forecasting [27]. These models can be used for prognostics by extrapolating the behaviour of equipment condition over time [27]. The WatchdogAgent is a toolbox that can be used to quantitatively assess and predict performance degradation levels of equipment using autoregression [27]. In here, time series data collected from the equipment are processed into joint time-frequency distributions. Then patterns in equipment condition data are extracted using principal component analysis and the principal components are predicted using autoregression to identify whether the future equipment condition stays within the expected thresholds for the normal equipment condition.

Random coefficient regression. In a normal regression model (e.g. linear, polynomial and logistic regression) the slope and intercept model parameters are fixed to a single value. In contrast, random coefficient regression models allow varying these model parameters according to a distribution [28]. This enables them to be used for prognostics since stochasticity in degradation processes can be modelled to a certain degree using random coefficients, assuming the parameters related to stochasticity are normally distributed [3]. Once a random coefficient regression-based prognostics model is estimated, a TTF distribution is computed as the distribution of times required for a failure precursor or a health index to violate a predefined failure threshold.

Wiener process. Formulated as a drift term plus a diffusion term following Brownian motion to model the stochastic degradation process, Wiener process is considered as a classical data-driven method for prognostics modelling [29, 12]. Often, the Bayesian method is used to update the drift parameter in these prognostics models so that real-time degradation monitoring information can be incorporated into prognostics modelling [12]. However, due to the Markov property of the Wiener process, the Bayesian updated drift parameter only utilises the current degradation measurement and not the accumulated measurements which could improve the prognostics performance [30]. This issue can be addressed by employing the sequential Bayesian update method [30].

Gamma process. Gamma process models assume that a particular process is monotonic and evolving only in one direction (i.e. either increasing or decreasing) [31]. Therefore, they are a natural selection for modelling degradation processes in which the deterioration is supposed to take place gradually over time in a sequence of positive increments (e.g. material fatigue crack growth) [31]. A comparison of the Wiener process and gamma process for predicting material fatigue that exhibits monotone-increasing patterns is presented in [32]. It is shown that the gamma process outperforms the Wiener process for material fatigue prognostics. Moreover, it is shown that wrongly treating a degradation process as a Wiener process instead of a gamma process causes the prognostics model to be misspecified and the negative impact on prognostics model performance is not negligible.

Markov models. These are stochastic methods and are based on the assumption that future states only depend on the present state and not on sequences of events that preceded it (i.e. Markov property) [29]. Markov model-based prognostics involves estimating probabilities of future failure states by defining probabilities associated with each state and probabilities associated with transitioning from one state to another. Variations of Markov models used for prognostics modelling include semi-Markov models, hidden Markov models and semi-hidden Markov models [29, 12].

Proportional hazards models. These are used to model the way covariates affect equipment life [33, 34]. They model deterioration as the product of a baseline hazard rate and a covariate function which reflects the effect of the operating environment on baseline hazard. This function is described by a vector of covariates and an associated vector of regression parameters. In order to estimate the TTF using a proportional hazards model, a survival function is often used [34]. The traditional approach to proportional hazards modelling was based on the assumption that the effects of covariates are time-independent [35]. An alternative approach was proposed by Makis and Jiang [35] where a generalisation of the original proportional hazards model equation is used to take the time-dependent covariates into account.

(II) Error-based learning approach

Error-based learning approach uses computational systems (e.g. artificial neural network (ANN)) to estimate typically a non-linear model that establishes the functional relationship between input stimuli and desired output [36]. It uses historical data to adjust the parameters of the functional relationship in order to obtain the optimal model possible for the computation task.

Multilayer perceptron (MLP). As shown in Fig. 2.3, MLP is a type of ANN that contains at least one hidden layer in addition to the input and output layers [37]. It can be used to develop prognostics models based on the assumption that there would be a certain relationship between the degradation process and corresponding prognostics of the equipment [38]. More specifically, if the degradation process remains consistent and follows the Markov property, the MLP can be used to extract the mapping relationship between degradation patterns and corresponding TTF of equipment [38].

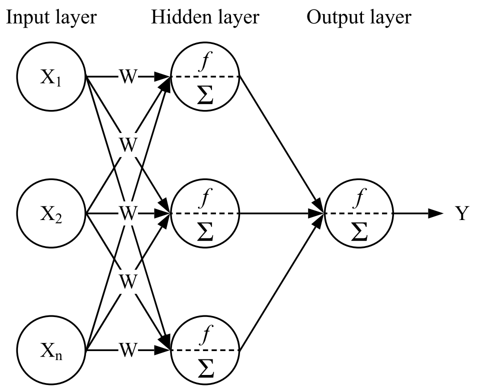


Fig. 2.3 Diagram depicting the multilayer perceptron with one hidden layer. The weights are represented with W and the activation function is represented with f. The $\sum$ represents the calculation of the weighted sum of inputs which is used as the input to the activation function.

Long short-term memory network (LSTM). This is a recurrent neural network (RNN) that is explicitly designed to avoid the long-term dependency problem (also referred to as the "fading memory" problem) in the vanilla RNN (see Fig. 2.4) [39]. Compared to the vanilla RNN which contains a simple structure for storing and maintaining information, the LSTM has four layers that interact with each other to maintain information in memory for long periods. It can be used to forecast the multivariate time series of condition monitoring and/or event data pertaining to the failure modes to estimate prognostics. More specifically, multiple features containing multivariate time series are given as inputs to the LSTM and the desired output is the forecast of the input features into a fixed number of future time steps. Then detecting when the forecasts first violated a predefined threshold the TTF is estimated.

Neuro-fuzzy systems (NFS). These combine fuzzy systems with ANNs to allow the estimation of fuzzy membership function parameters using an ANN, thus the manual optimisation of the fuzzy system is eliminated (also see Fig. 2.5) [25]. Since condition monitoring and event data pertaining to failures are linguistic, ambiguous or approximate in the real-world, NFS is popular for developing prognostics models [12]. To develop a NFS-based prognostics model, expert knowledge pertaining to the failure modes is converted into fuzzy rules first. Then the parameters of fuzzy membership functions are estimated by training an ANN using historical data pertaining to the failure mode [12].

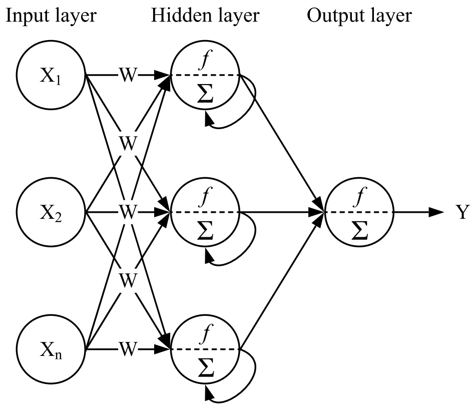


Fig. 2.4 Diagram depicting the recurrent neural network with one hidden layer. In contrast to the multi-layer perceptron, in a recurrent neural network the information cycles through a loop allowing the previously computed information to be utilised for the present computation task.

The objective of a NFS is to estimate the parameters of fuzzy membership functions using an ANN and extract the non-linear relationship between historical input data and corresponding prognostics of the equipment.

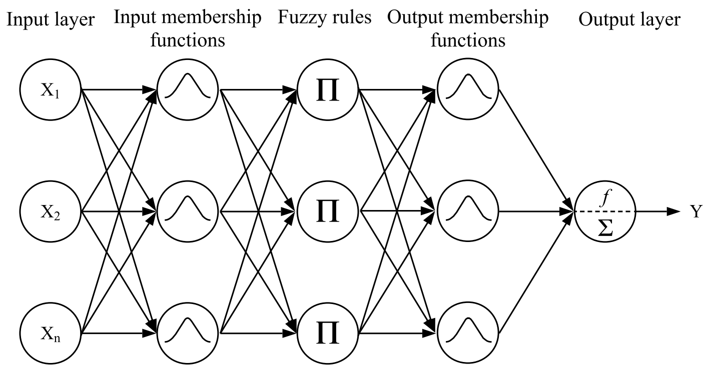


Fig. 2.5 Diagram depicting a neuro-fuzzy system.

(III) Instance-based learning approach

In contrast to the error-based approach which performs explicit generalisation using historical condition monitoring and/or event data, the instance-based learning approach to prognostics involves comparing data samples captured from ongoing degradation processes with historical condition monitoring and/or event data samples seen in training for prognostics modelling [40].

Random forest (RF). This is an instance-based ensemble decision tree learning algorithm for predictive modelling [41]. It uses historical data to estimate a model consists of a set of decision rules which is then used to produce predictions when a new data sample is given as the input [41]. The structure of a random forest model is shown in Fig. 2.6. It can be used to develop prognostics models by formulating a multi-class or a binary classification problem in which the task is to predict whether an input data sample captured from degrading equipment is pertaining to the failure mode that needs predicting [42]. The random survival forest is an extension to the RF and it combines survival analysis with instance-based learning for prognostics modelling [43]. Compared to other survival models (e.g. proportional hazards models), random survival forest does not rely on a specific hazard function parametrisation [44]. Instead, it can be used to cluster the equipment with similar degradation characteristics and then estimate a nonparametric reliability function for each equipment to predict their failures.

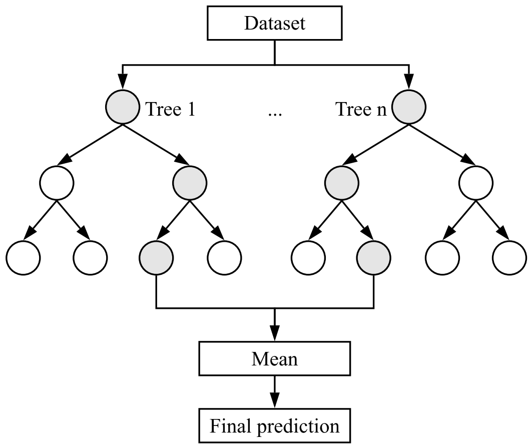


Fig. 2.6 Diagram depicting the random forest with *n* number of randomly uncorrelated decision trees. The nodes in each tree represent the features, links represent the decision rules and the leaves represent the predictions.

Support vector machine (SVM). This implements the principle of structural risk minimisation by constructing an optimal separating hyperplane in the hidden feature space of a dataset (also see Fig. 2.7) [45]. SVM classifier can be used to model prognostics by formulating a classification problem in which the task is to detect unhealthy states of equipment prior to their failures [46]. To this end, time series of condition monitoring and/or event data are converted into discrete data (i.e. histograms). The SVM classifier is then trained on the discrete dataset to learn a decision surface that separates healthy and unhealthy states of the equipment. When a new data sample captured from degrading equipment is given as the input, the classification model classifies it as healthy or unhealthy, and thus predicts imminent equipment failures. SVM can be used as a non-linear estimator for regression problems by extrapolating an optimal hyperplane that best fits training samples [47]. The extrapolation of the optimal hyperplane is used to make the TTF predictions using time-dependent features [48]. Apart from this distinction, SVM regression uses the same principle as the SVM classifier, that is, minimise model error using structural risk minimisation whilst individualising the hyperplane by maximising the margin.

(IV) Probabilistic machine learning approach

The probabilistic framework to machine learning is based on the idea that learning can be considered as inferring plausible models to explain the data [49]. More specifically, if the mathematics

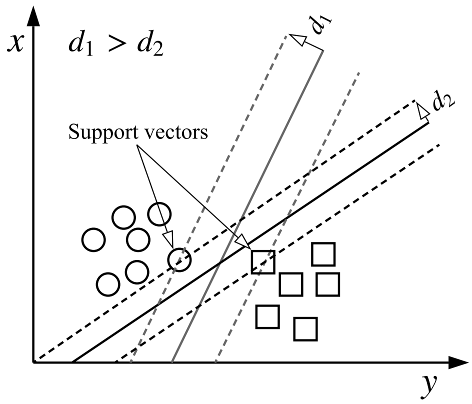


Fig. 2.7 Diagram depicting the support vector machine. The closest layer of vectors to the separating hyperplane is called support vectors. The space between the support vectors of different classes is the margin (d_1 and d_2). The best fit hyperplane is the one with the maximum margin (in this case d_1).

of probability theory is used to express all forms of uncertainty and noise associated with a model, then the inverse probability (i.e. Bayes theorem) can be used to adapt that model and make predictions [49]. The key difference between probabilistic machine learning methods and the rest of the machine learning methods (i.e. error-based and instance-based learning methods) is that the main computation problem of the former is integration, whereas the computation problem of the latter is the optimisation of model parameters [49]. The common probabilistic machine learning methods used for prognostics modelling includes the following [50, 2]:

Gaussian process regression (GP regression). GP regression is a kernel-based fully Bayesian regression algorithm that provides a natural way to introduce kernels into a regression modelling framework [51]. It can be used to develop prognostics models that estimate the TTF of equipment [52]. The approach involves using the GP regression algorithm to compute posterior degradation estimates by constraining a prior distribution to fit a historical dataset consisting of data pertaining to a failure precursor of the equipment. Then the regression model is used to trend the failure precursor and by detecting when a predefined failure threshold is violated the TTF is predicted [52].

Relevance vector machine (RVM). RVM uses Bayesian inference for predictive modelling and its functional form is identical to the SVM [53]. However, it allows regression tasks to be performed within a probabilistic machine learning framework [53]. The approach to use RVM regression for prognostics modelling is similar to the approach used for SVM regression, although there are some notable differences. First, the likelihood function and prior knowledge need to be defined. The former describes the features in the historical dataset used for model training and the latter captures the complexity of the prognostics model to enhance the generalisation capability on new data samples captured from degrading equipment [54]. Then the training algorithm to estimate a RVM regression-based prognostics model needs to be identified. Suitable training algorithms include the MacKay iterative algorithm, sequential sparse Bayesian learning and Expectation Maximisation [53].

Hybrid prognostics

Hybrid prognostics models aim to leverage the strengths of multiple prognostics approaches [13]. According to the literature, they can be classified as follows [13]:

Hybrids of physics model-based and knowledge-based models. In these hybrid models, prognostics is modelled using the physics model-based approach and a knowledge-based model is used to enhance the prognostics performance by integrating expert knowledge pertaining to failure modes and equipment operating conditions [13]. More specifically, using a set of rules developed by the domain experts, the knowledge-based model supports the physics model-based prognostics model to determine fault states and categorise degradation levels for producing prognostics.

Hybrids of data-driven and knowledge-based models. In here, prognostics is modelled using the data-driven approach and similar to the aforementioned hybrid models, a knowledge-based model is used to enhance the prognostics performance by integrating expert knowledge pertaining to failure modes and equipment operating conditions [13]. Introducing expert knowledge into data-driven prognostics models aims to account for the dynamics of degradation due to various operating conditions and failure modes, and thus reduce the uncertainty associated with predictions produced by purely data-driven models [13].

Hybrids of data-driven and physics model-based models. According to the literature, there are two methods to combine data-driven and physics model-based models: series method and parallel method [16]. The former involves combining a physics model-based model that has prior knowledge about the degradation process and a data-driven model that serves as a state estimator of unmeasured degradation process parameters that are hard to model using first principles. The latter involves fusing predictions produced by data-driven and physics model-based prognostics models to reduce the uncertainty associated with the final prediction.

Summary of findings and concluding remark

In this review, a discussion on prognostics approaches and predictive modelling methods used in each approach is provided. The findings discussed in the review and merits and limitations of prognostics approaches are summarised below.

1. Physics model-based prognostics models are applicable in industrial scenarios in which accurate mathematical models can be constructed from first principles [2, 12]. The key limitation of these models is that in most industrial scenarios physics pertaining to equipment degradation is too stochastic and complex to model, which makes it difficult and expensive to develop physics model-based prognostics models [2, 55].

2. Knowledge-based prognostics models assess the similarity between an observed situation and a knowledge base of historical failures to estimate prognostics. The key limitations of knowledge-based prognostics models are: obtaining domain knowledge and converting it into rules is difficult in most industrial scenarios due to stochasticity and complexity of degradation

processes [2, 12]; they do not generalise into new situations that are not covered explicitly in their knowledge bases [2, 12].

3. Even though the hybrid approach aims to leverage the strengths of multiple prognostics approaches, it has also been unsuccessful since under the conditions of limited failure data availability, this prognostics approach inherits the limitations of data-driven, physics model-based and knowledge-based approaches [2, 13].

4. Data-driven prognostics approach has become popular over the other prognostics approaches since it does not require explicit mathematical representations or detailed knowledge bases to estimate prognostics models [2, 3]. The key limitation of the data-driven approach to prognostics is that it requires large amounts of historical failure data to estimate prognostics models [2]. Table 2.1, 2.2 and 2.3 outline the other limitations and merits of data-driven prognostics methods using the findings obtained from this literature review.

Table 2.1 Merits and limitations of error-based learning methods used for prognostics modelling.

Method	Merits	Limitations
MLP	• Able to model the non-linear behaviour of industrial systems using historical condition monitoring and/or event data [56]. • Does not require prior knowledge pertaining to failure modes [55].	• Requires large amount of historical condition monitoring and/or event data [56]. • No standard method to determine network structure and these methods are "black box" systems [2, 56].
LSTM	• Good for prognostics modelling using sequential data as it can capture and utilise long-term dependencies [57]. • Provides different approaches to prognostics modelling (e.g. many-to-many prediction and sequence-to-sequence prediction) [58].	• Can be difficult to train and implement [57]. • Often requires more data to train than other error-based learning methods [59].
Neuro-fuzzy systems	• Good for prognostics in scenarios with ambiguous and/or approximate historical data [12]. • More compatible with human reasoning process than other error-based learning methods [56].	• Linguistic terms may compromise the accuracy of the prognostics [56]. • Not feasible in situations where fuzzy membership functions are complex to determine [56].

Table 2.2 Merits and limitations of instance-based learning methods used for prognostics modelling.

Method	Merits	Limitations
Random forest	• Easy to use, trains faster and scalable [60]. • Requires less data preprocessing since normalisation is not required as it only needs the absolute values for branching and creating decision rules [60].	• A large number of decision trees can make the algorithm slow and ineffective for estimating prognostics in real-time [60]. • Models are not easily interpretable [61].
Random survival forest	• Allows prognostics modelling using right-censored survival data [43].	• Inferior in identifying predictors with a less ratio of population, hence it can underperform compared to the other survival models [62].
SVM	• Compared to the error-based learning methods, SVM can prevent overfitting effectively [63].	• Standard methods and theoretical guidance for choosing the kernel function for SVM do not exist [56, 63].

Table 2.3 Merits and limitations of probabilistic machine learning methods used for prognostics modelling.

Method	Merits	Limitations
GP regression	• Can estimate model hyperparamters and covariances directly from training data, hence does not involve hyperparameter tuning [64]. • Can adapt to environments (i.e. dynamic processes and variable operating conditions) [56].	• A large number of historical data samples are required to avoid the degeneracy problem [12]. • Only suitable for Gaussian likelihood [56].
RVM	• Can estimate model hyperparamters directly from training data, hence does not involve hyperparameter tuning [53]. • As with other probabilistic machine learning methods, RVM can produce probabilistic predictions allowing the uncertainty quantification, instead of point predictions [53].	• The storage and computation complexity of RVM during the training phase grows as the number of data samples increases, thus it can be computationally costly for large training datasets [53]. • Requires prior assumptions of model structure [53].

Concluding remark. Data-driven prognostics approach has become popular over the other prognostics approaches, and yet, it is challenging to develop effective data-driven models for industrial equipment prognostics due to limited failure data availability in industrial scenarios.

2.1.3 Prognostics under the conditions of limited failure data availability

In this section, the reasons for the use of physics model-based, knowledge-based and hybrid prognostics approaches to be unsuccessful when failure data are limited for data-driven prognostics are outlined first. The existing techniques used to address the limited failure data availability problem for data-driven prognostics and the reasons for them to be unsuccessful are discussed next.

When failure data are limited for data-driven prognostics, the use of physics model-based, knowledge-based and hybrid prognostics approaches have been unsuccessful in most industrial scenarios due to the following: once a physics model-based prognostics model is developed, it needs to be validated prior to its application using large amounts of historical failure data pertaining to the failure mode that needs predicting [2, 12]. However, under the conditions of limited failure data availability, it is not possible to obtain large amounts of failure data required for validating physics model-based prognostics models. Moreover, they are based on the assumption that system behaviour can be described analytically and accurately which makes it difficult and expensive to develop them for prognostics since equipment degradation is too stochastic and complex to model [2, 13].

As discussed in the previous section, knowledge-based prognostics involves obtaining domain knowledge and converting it into rules which is also difficult in most industrial scenarios due to the stochasticity and complexity of equipment degradation processes [2]. More importantly, for a knowledge-based prognostics model to be effective under the conditions of limited failure data availability, the knowledge base needs to be as complete and accurate as possible [2]. This leads to the number of rules to increase, which causes the combinatorial explosion problem [2]. The hybrid prognostics approach has also been unsuccessful for prognostics under the conditions of limited failure data availability since it inherits the aforementioned shortcomings of data-driven, physics model-based and knowledge-based prognostics approaches.

A set of techniques that aimed to address the problem of limited failure data availability for data-driven prognostics has emerged over the years. According to the literature, these existing techniques can be classified as data level and algorithmic level techniques [10].

Data level techniques

Data level techniques aim to address the limited failure data availability problem by treating data imbalance in prognostics modelling datasets [65, 10]. They include oversampling, undersampling, generative modelling-based data augmentation, equipment clustering-based prognostics and simulation-based run-to-failure data generation.

(I) Oversampling

This is the most popular existing technique used for treating data imbalance [10, 11]. To this end, it introduces data samples to the minority class (i.e. failure data class) by duplicating existing failure data samples or creating new data samples randomly [10]. Oversampling methods include Random

Oversampling (ROS), Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthesis (ADASYN).

Even though ROS is a naive oversampling method, it is the most effective existing oversampling method since it makes no assumptions about the data and no heuristics are used [10]. It reduces data imbalance in prognostics modelling datasets by randomly selecting existing failure data samples and duplicating them [10]. In contrast to ROS, SMOTE generates new failure data samples to treat data imbalance by interpolating existing failure data samples, hence data are not duplicated [10]. More specifically, a real failure data sample is randomly selected and k of the nearest neighbours for that sample are found. Then a neighbour sample is chosen and a synthetic failure data sample is created at a randomly selected point between the two samples in the feature space. ADASYN is based on SMOTE and the difference is that it adaptively generates failure data samples according to their distributions [10]. In other words, more synthetic data are generated for the failure data samples that are harder for a prognostics model to learn than the failure data samples that are easier to learn.

The major shortcoming of oversampling methods is that they do not introduce new and plausible failure data samples into prognostics modelling datasets [11, 10]. Hence, the fundamental lack of failure data issue is not addressed [11]. More specifically, ROS makes exact copies of real failure data samples which may lead to overfitting, hence a prognostics model can only learn to generalise for a small, duplicated set of failure data samples [11]. Advanced oversampling methods (i.e. SMOTE and ADASYN) are based on data interpolation which is less effective in synthesising data variants that are subjected to the underlying distribution of real failure data samples [11]. In other words, these methods generate data randomly, hence there is no control over the modes of data being generated. This leads implausible failure data samples to be introduced into prognostics modelling datasets [11].

(II) Undersampling

This technique treats data imbalance by removing intrinsic samples from the majority class (i.e. normal data class) [10]. Random Undersampling (RUS) is a naive yet the most effective existing undersampling method since similar to ROS, it makes no assumptions about the data and no heuristics are used [10]. To treat data imbalance in prognostics modelling datasets, RUS can be used to remove normal data samples randomly. Advance undersampling methods include nearest neighbour undersampling and One-Sided Selection (OSS) [66]. The nearest neighbour undersampling method aims to treat data imbalance by randomly removing normal data samples based on their distance to other samples in the normal data class [66]. OSS applies Tomek links followed by the Condensed Nearest Neighbour rule to treat data imbalance [10]. Tomek links are used to remove noisy and borderline normal data samples. Then Condensed Nearest Neighbour rule is used to remove normal data samples that are distant from the decision boarder.

Since undersampling methods discard data samples pertaining to the normal condition of the equipment, potentially useful non-failure data samples could also be discarded [11]. This can degrade the discriminating power of predictive algorithms [11]. More importantly, undersampling does not increase the quantity of failure data samples in imbalanced prognostics modelling datasets, hence the fundamental lack of failure data issue is not addressed [11].

(III) Generative modelling-based data augmentation

Recently, a few generative modelling-based techniques have emerged for treating data imbalance in prognostics modelling datasets (e.g. [67, 68]). These techniques employ generative adversarial networks (GAN) to generate data that resemble the distributional properties of real failure data and thus augment imbalanced prognostics modelling datasets to include an increased number of failure data samples. More specifically, the GAN is used to estimate a generative model that is capable of replicating the real failure data distribution in an adversarial training framework. The adversarial training framework allows a model to estimate its parameters by competing with another model [69]. The estimated generative model is used to sample new data from a data distribution containing random noise (i.e. from a Gaussian distribution) [67]. The generated samples are then added to the imbalanced prognostics modelling dataset to treat data imbalance.

The major shortcoming of generative modelling-based data augmentation techniques is that the noise being added to newly generated data samples is not based on sound engineering knowledge of failure mechanisms or the context of the prognostics problem [70, 8]. This leads to different modes of data being generated since the data generation process is not controlled and directed [70, 8]. Hence, similar to the data interpolation used by SMOTE and ADASYN, generative modelling-based data augmentation techniques cause implausible failure data samples to be introduced into prognostics modelling datasets.

(IV) Equipment clustering-based prognostics

This technique uses the equipment similarity information to create clusters of equipment and then estimate prognostics models for individual equipment or a cluster of equipment in a collaborative learning framework [71]. The equipment similarity information can range from metrics such as equipment age, type and model to calculated metrics such as health indicators [71]. Dynamic metrics which adapt to the equipment operating condition are also proposed in the literature [72]. The collaborative learning framework allows Digital Twins (i.e. dynamic digital representations of the equipment) to share data with each other [71, 73]. When equipment has a well-defined range of similarity and a cluster of equipment collectively has a sufficient amount of failure data samples for prognostics modelling, this technique allows treating data imbalance as follows: equipment with sufficient amount of failure data samples share them with similar equipment that has a limited amount of failure data samples so that data imbalance in datasets used for their prognostics modelling is reduced or eliminated [74, 71].

This technique has been successful in industrial scenarios in which it is possible to accumulate sufficient amounts of failure data for individual clusters of equipment [71, 74]. However, as stated in the problem statement, this thesis defines the limited failure data availability problem as the lack of failure data samples collectively available for all the equipment in an industrial scenario for prognostics modelling. This is often the case in practice due to two major reasons: (i) overprotective maintenance and replacement regimes; (ii) highly reliable equipment [6, 7]. These cause failures to become rare for all the equipment, and thus make it difficult to accumulate a sufficient amount of failure data using equipment clustering for prognostics modelling.

Moreover, for equipment clustering-based prognostics to be successful under the conditions of limited failure data availability, the optimal balance between the size of equipment clusters and the amount of failure data collectively available for individual equipment clusters needs to be established. However, under the conditions of limited failure data availability, this is difficult due to the following: (i) accumulating a sufficient amount of failure data samples by abstracting the similarity criteria to increase the number of equipment belongs to individual clusters will cause the clusters to overlap. Hence, the fundamental objective of equipment clustering-based prognostics is not achieved; (ii) improving the similarity between equipment to improve the suitability of data shared between them could cause the equipment to be divided into small clusters [75]. This causes individual clusters of equipment to have a small amount of failure data samples for prognostics modelling, especially under the conditions of limited failure data availability,

(V) Simulation-based run-to-failure data generation

When failure data are limited for data-driven prognostics, simulation is used to generate run-to-failure data using physics-based simulation models or test rigs [76, 77]. Physics-based simulation models generate response surfaces of sensors using a simulation model as a function of the variation of condition and performance of the equipment of interest [76]. Popular physics-based simulation models include thermo-dynamical models, Thevenin model and lumped-element models [76, 78]. There also exists commercial simulation environments which can be used to generate run-to-failure data by varying certain parameters affecting equipment degradation (e.g. operating environment conditions and load): Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) environment [76] and Gas Turbine Simulation environment [79].

Test rigs allow using an apparatus specifically designed for performing controlled experiments that are suitable for generating run-to-failure data [80, 77]. The experiments are performed using accelerated life testing (ALT), which involves making the target equipment subject to conditions (e.g. stress, temperature, vibration and pressure) in excess of its normal service parameters in a short amount of time. Test rigs are incorporated with a condition monitoring system which allows capturing run-to-failure data during ALT. These data are then used to estimate prognostics models which can be applied to predict failures in similar equipment in real-world industrial scenarios that are subjected to the same conditions used during ALT [80].

Despite its success in scenarios where accurate mathematical models can be constructed from first principles, physics-based simulation has been unsuccessful in most industrial scenarios due to the following: physics-based simulation is based on the assumption that equipment behaviour and degradation processes can be described analytically and accurately, which is difficult and expensive since they are too stochastic and complex to model [2, 13]. Hence, when prognostics models that are trained on data generated by physics-based simulation have been applied in the real-world, they often underperform. The use of test rigs has been successful when the effect of equipment operating environment on failure modes is well understood and can replicate physically [77]. However, this is also difficult and expensive in most industrial scenarios since the effect of equipment operating environment on failure modes is difficult and expensive to replicate physically [2, 13]. Hence, prognostics models that are trained on data generated by test rigs also often underperform when applied in the real-world.

Algorithmic level techniques

Algorithmic level techniques include cost-sensitive learning, boosting and bootstrap aggregation (also known as bagging) [10]. They aim to overcome data imbalance by adapting predictive algorithms to account for the skewness in class distributions during model estimation [10]. However, algorithmic level techniques are seldom used to overcome data imbalance caused by the limited failure data availability problem since they do not increase the quantity of failure data samples in imbalanced prognostics modelling datasets [11]. In addition to this major shortcoming of all the algorithmic level techniques, their other shortcomings when employing to develop prognostics models under the conditions of limited failure data availability are discussed below.

(I) Cost-sensitive learning

This technique takes the cost of prediction errors into account when training classification-based prognostics models [81]. More specifically, instead of each data sample being either correctly or incorrectly classified, failure data and normal data classes are given a misclassification cost. Thus, instead of trying to optimise model accuracy, the objective is to minimise the total misclassification cost. Cost-sensitive learning can be used to overcome data imbalance in prognostics modelling datasets by biasing the classifier to perform well on the failure data class by assigning it a higher misclassification cost compared to the normal data class [81].

However, the issue with this technique is that cost information is rarely available for industrial equipment prognostics since they often depend on multiple considerations that are not easily compared [82]. In other industrial scenarios, the cost associated with prognostics is often a complex, multidimensional function that includes monetary costs, safety costs and reputation costs [82].

(II) Boosting and bagging

Boosting iteratively increases the accuracy of classification-based models by placing different weights on data samples in each training iteration [10]. To overcome data imbalance in prognostics modelling datasets, boosting allows increasing weights assigned to the failure data class and decreasing weights assigned to the normal data class. This forces classification-based prognostics models to focus on learning to classify failure data samples accurately in consecutive training iterations. Bagging allows using multiple learners to make a single prediction [10]. More specifically, random samples from a training dataset (i.e. bootstrap samples) are selected and a single weak learner model is fitted on each sample. Then predictions produced by all the models are combined (i.e. aggregation) to make a single prediction. However, bagging does not overcome data imbalance alone and requires combining it with an oversampling or an undersampling method [83, 11].

The issue with boosting is that it causes prognostics models to overfit since it tends to weigh data samples belongs to the failure data class than those belonging to the normal data class [11]. Moreover, it performs poorly if the base learner always achieves poor precision and recall [11]. Since bagging needs to be combined with an oversampling or an undersampling method to overcome data imbalance in prognostics modelling datasets, it inherits their shortcomings which are discussed under the data level techniques [83, 11].

Summary of findings and concluding remark

In this review, a discussion on existing techniques used to address the problem of limited failure data availability for data-driven prognostics and the reasons for them to be unsuccessful in addressing this problem is provided. The findings discussed in the review are summarised below.

1. Oversampling, undersampling and generative modelling-based data augmentation have been unsuccessful in addressing the problem of limited failure data availability for data-driven prognostics sufficiently. This is since they either duplicate existing failure data or randomly generate data, hence do not introduce new and plausible failure data samples to treat data imbalance in prognostics modelling datasets [11, 10].

2. Often in industrial scenarios there is limited failure data available for all the equipment collectively due to the rare failures problem which is caused by overprotective maintenance and replacement regimes and highly reliable equipment [6, 7]. Equipment clustering-based prognostics techniques have been unsuccessful in these industrial scenarios since it is impossible to accumulate sufficient amounts of failure data for prognostics modelling through data sharing.

3. Simulation-based run-to-failure data generation has been successful in industrial scenarios in which accurate mathematical models can be constructed from first principles (in the case of the physics-based simulation) or the effect of equipment operating environment on failure modes is well understood and can replicate physically (in the case of test rigs). However, they have been unsuccessful in most industrial scenarios since the aforementioned luxuries are seldom available for industrial equipment prognostics due to the stochasticity and complexity of equipment degradation processes [2, 13].

4. The major shortcoming of all the algorithmic level techniques is that they do not increase the quantity of failure data samples, hence the fundamental lack of failure data issue is not addressed [11, 10].

Concluding remark. Due to the major shortcomings of existing techniques, the problem of limited failure data availability remains a challenge for developing effective data-driven models for industrial equipment prognostics. Moreover, when failure data are limited for data-driven prognostics, the use of other prognostics approaches has been unsuccessful.

In order to address this problem, this thesis presents a methodology developed using the conditional generative adversarial network (CGAN) which is an extension to the previously discussed GAN (see Generative modelling-based data augmentation in Sec. 2.1.3) [70]. The methodology uses the CGAN to condition the failure data generation process on auxiliary information, and thus allows estimating a generative model that is capable of generating new and plausible failure data samples.

2.1.4 Research gap

It can be observed from the literature review that there is a growing research trend to develop prognostics models that produce prognostics with minimal uncertainty to enable the effective implementation PdM. The motivation for this trend is that the effective implementation of PdM is capable of delivering on the promise of an optimal maintenance policy that balances the cost of maintenance, the risk of failure and performance of the equipment. Thus, maintenance can be delayed to take the maximum value out from the equipment until the risk of failure is not critical and equipment performance remains unaffected.

The data-driven prognostics approach has become popular for industrial equipment prognostics over the other approaches since it does not require explicit mathematical representations or detailed knowledge bases. Instead, it uses statistical and machine learning methods to estimate prognostics models from large amounts of condition monitoring and/or event data relating to past failures. Indeed, a growing number of data-driven models have been proposed in the literature to exploit how to produce prognostics with minimal uncertainty for enabling the development of optimal PdM policies. However, they have been unsuccessful due to the problem of limited failure data availability for prognostics, which is often the case in industrial scenarios due to overprotective maintenance and replacement regimes and highly reliable equipment.

Since employing other prognostics approaches to develop prognostics models under the conditions of limited failure data availability has been unsuccessful, a set of techniques that aimed to address the limited failure data availability problem for data-driven prognostics has emerged. However, the use of these existing techniques has been stalled due to the following two reasons: (i) they either duplicate existing failure data or randomly generate data, hence new and plausible failure data are not introduced into prognostics modelling datasets; (ii) or they do not increase the quantity of failure data samples available for prognostics modelling, hence the fundamental lack of failure data issue is not addressed. Namely, the development of prognostics models under the conditions of limited failure data availability is not fully realised. Hence, the effective implementation of PdM remains a vision rather than a reality. The following research gap follows from this analysis:

Research gap. A systematic approach for prognostics modelling under the conditions of limited failure availability is a research gap.

The importance of a solution that addresses the aforementioned research gap is that it enables the effective implementation of PdM, and hence realise the vision of an optimal maintenance policy for industrial organisations. In other words, industrial organisations will be able to prevent costs due to under-maintenance and over-maintenance of equipment and false alarms. To address the research gap, this thesis presents a methodology for prognostics under the conditions of limited failure data availability. More specifically, the research presented in this thesis answers the following research questions:

1. *How can equipment failure be predicted under the conditions of limited failure data availability?* This research question aims to develop a methodology that allows generating new and plausible failure data for prognostics under the conditions of limited failure data availability.

2. *What impact does the proposed solution have on prognostics compared to existing techniques under the conditions of limited failure data availability?* This research question aims to quantify the impact of methodology in terms of improved prognostics performance under the conditions of limited failure data availability compared to the state-of-the-art techniques proposed in the literature, and produce insights into the key factors influencing the effectiveness of the methodology.

2.2 Industrial rationale

The literature review established the academic background for this research by providing a detailed review of existing literature related to the topic of interest. The review identified the need for a systematic approach to develop prognostics models under the conditions of limited failure data availability for enabling the effective implementation of PdM. Apart from the findings obtained from an entirely theoretical point of view, it is also essential to look at the problem from a practical angle to reinforce the academic gap in knowledge with an industrial rationale. The aim of this section is therefore to show the necessity of the methodology for prognostics under the conditions of failure data availability in practice.

2.2.1 Limited failure data availability for prognostics in practice

The airline, logistics, nuclear power and telecommunications industries are examples of industries that have the problem of limited failure data availability for prognostics, and hence so far failed to implement PdM effectively [84–86]. The motivation for PdM in the airline, logistics and nuclear power industries is since it can increase the uptime and safety whilst maximising the value of the equipment [84, 87]. The telecommunications industry is also motivated to adopt PdM since it is able to ensure that the performance of the equipment used in telecommunications networks (e.g.

core routers, edge routers and broadband lines) remains unaffected and unplanned downtime of telecommunications services (e.g. telephony, internet, television broadcasting and radio) is prevented whilst maximising the value of the equipment [85].

Despite the increasing motivation, the effective implementation of PdM in industries such as airline, logistics, nuclear power and telecommunications has been unsuccessful due to highly reliable equipment and overprotective maintenance and replacement regimes [6, 7]. More specifically, industrial organisations are committed and responsible for designing, manufacturing, testing and maintaining equipment according to standards that enforce high reliability since its failures often lead to catastrophic consequences: loss of lives due to aircraft component failures, disruptions to medical and food supply chains due to heavy-truck component failures, collateral damage to lives, equipment and environment due to nuclear power plant equipment failures) [84, 88, 87]. However, highly reliable equipment and overprotective maintenance and replacement regimes pose a challenge to prognostics modelling since the equipment is rarely allowed to run to failure once degradation has been detected [86]. In other words, they reduce the number of degradation trajectories pertaining to failure modes, and hence cause failure data to be limited for developing prognostics models [6, 7].

2.2.2 Industrial case studies in academic literature

In this section, a set of industrial case studies from the aforementioned industries is summarised to show the motivation for the effective implementation of PdM in industry and how the problem of limited failure data availability for prognostics has been the bottleneck for accomplishing it.

(I) PdM of Airbus aircraft

Airbus is a world-leading aircraft manufacturer and interested in PdM since it allows maintenance to be performed only when necessary so that additional costs due to over-maintenance actions (i.e. production loss due to planned downtime of aircraft and costs of spare parts, supplies and labour) which are resulted from the time-based maintenance strategy can be prevented [84]. To this aim, Alestra et al. [84] presented an industrial case study that involved developing a prognostics model for a critical subsystem of Airbus aircraft. Since the target failure mode is rare (the failure mode only occurs with a frequency of 0.2-1% compared to other failure modes of the subsystem), condition monitoring and event data pertaining to it are limited for developing a prognostics model. Hence, the authors developed an event prediction model to raise alarms when the condition of the equipment deviates from the norm.

However, in order to develop a PdM model one must predict failures and not events [4]. More specifically, PdM models need the TTF of equipment or the probability that equipment operates without failure up to some future time as the input to estimate the optimal time to perform PdM actions [4]. Predicting events will cause PdM policies to produce over-maintenance actions which lead to the aforementioned additional costs. This is since event predictions are false alarms for PdM models which aim to estimate the optimal time to perform PdM actions in order to prevent the unexpected consequences of failures.

(II) PdM of United States nuclear power reactor fleet

In the United States (US), a majority of nuclear power reactors (NPR) is moving to extended operation from the initial operating license period of 40 years to an extended period of 60 years [86]. This is since it is challenging for the current technical, manufacturing, economic and regulatory infrastructures to support the building of replacement power generation (while continuing to build new NPRs) if operating NPRs were to be shut down and decommissioned at the end of their initial operating period [86]. The maintenance of safety-critical and high-value components of NPRs operating under the extended license period is given priority since they are coming to the end of their lifetime, which causes an increasing number of wear out failures [86]. However, since the maintenance and replacement of these components lead to high monetary and safety costs, maintenance actions need to be postponed until warranted [89, 86]. The PdM strategy is chosen to accomplish this as it allows increasing the uptime and safety of NPRs whilst maximising the value of their components [87].

Ramuhalli et al. [86] presented an industrial case study that involved assessing the feasibility of developing prognostics models for predicting crack growth in austenitic and ferritic steels used in reactor pressure vessels in the US NPR fleet. It is found that even though condition monitoring data captured from degrading reactor pressure vessels are showing “trendable” patterns, the absence of a sufficient amount of failure data is a barrier for developing effective prognostics models. The reason for this challenge is, safety-critical and high-value components in NPRs are rarely allowed to run to failure once degradation has been detected since their failures often lead to catastrophic consequences (e.g. collateral damage to lives, equipment and environment) [89, 86].

(III) PdM of AT&T telecommunications network equipment

AT&T is a world-leading telecommunications service provider and its motivation for adopting PdM is to increase the availability of telecommunications services for their customers whilst preventing additional costs due to over-maintenance and false alarms [90]. Weiss and Hirsh [85] presented an industrial case study which aimed at predicting electronic equipment failures in the Class-4 Electronic Switching System (4ESS) of the AT&T telecommunications network. Since 4ESS failures are rare, the authors formulated the problem as an event prediction problem in which the objective is to predict events using historical alarm event data [85]. A genetic algorithm-based machine learning model is proposed and it is capable of identifying temporal and sequential patterns in historical alarm event data for predicting impending events. Multiple prediction patterns are identified and the best pattern is selected by fusing multiple prediction patterns in different ways using the genetic algorithm.

However, as discussed in the aforementioned Airbus aircraft maintenance case study, predicting events is not the same as predicting failures and it will cause PdM policies to produce over-maintenance actions. This leads to additional costs due to planned downtime of telecommunications services and costs of spare parts, supplies and labour. Hence, in order to develop PdM models, one must predict failures and not events [4].

In this section, the industrial rationale for the research presented in this thesis is provided. It is shown that highly reliable equipment and overprotective maintenance and replacement regimes cause failure data to be limited for developing prognostics models in practice. The diverse set of industrial case studies summarised in this section demonstrates the motivation for the effective implementation of PdM in industry and how the problem of limited failure data availability for prognostics has been the bottleneck for accomplishing it. The purpose of the methodology for prognostics under the conditions of limited failure data availability is to address this bottleneck, and hence enable the effective implementation of PdM for industrial organisations.

2.3 Conclusion

Using the findings obtained from a systematic literature review, this chapter provided discussions on the state-of-the-art of prognostics research field, existing techniques used to address the problem of limited failure data availability for prognostics and reasons for them to be unsuccessful. It was shown that a systematic approach to prognostics modelling under the conditions of limited failure data availability is a research gap. Moreover, the findings obtained from the industry review showed that there is an increasing motivation for addressing this research gap. In the next chapter, the solution this thesis presents to address this research gap is provided.

Chapter 3

Methodology for prognostics under the conditions of limited failure data availability

This chapter presents the methodology for prognostics under the conditions of limited failure data availability which addresses the first research question of this thesis. The chapter commences by introducing the concept of conditional generative modelling since it is used to integrate auxiliary information pertaining to failure mechanisms, limited amount of failure data available and noise for estimating a generative model that is capable of generating new and plausible failure data. Further, a method that can be used to measure the extent of limited failure data availability problem for prognostics is introduced. The description of the methodology and its theoretical results are presented next. Then an analysis is performed to gain deeper insights into the behaviour and working mechanism of the methodology.

3.1 Concept of conditional generative modelling

The key components of the methodology are: (i) identification, validation and conversion of sound engineering knowledge of failure mechanisms (i.e. auxiliary information); (ii) using the converted auxiliary information and the available historical failure data to estimate a generative model that is capable of generating plausible failure data samples. These two components are integrated using the conditional generative adversarial network (CGAN). In this section, a description on how the CGAN can be used as a platform to estimate generative models is provided using the definitions and equations from the literature.

Generative adversarial network

The CGAN is an extension to the GAN (commonly known as the vanilla GAN) which provides a framework for training deep generative models in a two-player minimax game [70]. More specifically, a generative model is trained in an adversarial training framework to generate data similar to the real data distribution. The adversarial training framework allows a model to estimate its parameters by competing with another model according to a value function of a two-player minimax game [69]. Then the trained generative model is used to sample new data from a distribution containing random noise (i.e. from a standard multivariate normal distribution) [69].

The vanilla GAN consists of two ANNs: a generator G and a discriminator D. Given a dataset X with samples $\{x \in X\}$, its goal is to estimate a generative model that captures the generator distribution $P_G(x)$ that matches the real data distribution $P_{\text{data}}(x)$. The vanilla GAN estimates this generative model by first enabling the model to sample data from P_G by transforming a prior noise variable $z \sim P_{\text{noise}}(z)$ into a new data sample $G(z)$. Then the discriminator network D is used to discriminate between whether the generated sample $G(z)$ is a real data sample (i.e. $G(z)$ is sampled from the real data distribution P_{data}) or a fake data sample (i.e. $G(z)$ is sampled from the generator distribution P_G). Thus, the discriminator outputs a single scaler indicating the probability of whether a given data sample is real or fake without knowing the sample is generated by the generator. The generator G uses this scaler as the feedback to minimise its loss function $\log(1-D(G(z)))$ whilst the discriminator D tries to minimise the loss function $\log(D(x))$ to improve its discriminating power. The training of the GAN is stopped when an equilibrium for the minimax game is achieved. In the GAN, the equilibrium is achieved when the probability $D(x) = 1/2$ (i.e. the discriminator is no longer able to detect whether a given data sample is real or fake) [69]. Formally, the value function $V(G,D)$ of the two-player minimax game for estimating a generative model using the vanilla GAN is as follows:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim \text{noise}}[\log(1-D(G(z)))] \tag{3.1}$$

Conditional generative adversarial network

In the vanilla GAN, the generative model is trained without conditioning the noise being added to the newly generated data samples [70]. Thus, the generation of plausible data is not guaranteed, since there is no control over the modes of data being generated [70]. This issue is addressed by extending the vanilla GAN into the CGAN [70]. In contrast to the vanilla GAN, CGAN conditions the generator and discriminator on auxiliary information related to the prediction task [70]. This is done by feeding a conditioning vector (i.e. vector representation of auxiliary information) into the generator and discriminator ANNs as additional inputs.

More specifically, using auxiliary information vector Y with samples $\{y \in Y\}$ and noise vector Z with samples $\{z \in Z\}$, the generator G in CGAN is modified to generate data samples $G(z \mid y)$ compared to $G(z)$ in the vanilla GAN. This means G generates a fake data sample $G(z \mid y)$ from the joint distribution of noise and auxiliary information $Pr(Z,Y)$. Similarly, the discriminator is extended to $D(x \mid y)$ compared to $D(x)$ in the vanilla GAN. Thus, the discriminator tries to discriminate between real and fake data samples by detecting whether a given sample is sampled from the joint distribution of real data and auxiliary information $Pr(X,Y)$. The feedback from the discriminator at each training iteration allows the generator to improve its performance when conditioning noise on auxiliary information since it now needs to generate fake data samples that fool the discriminator in two cases: (i) when discriminating against the real data distribution; (ii) when discriminating against auxiliary information related to the prediction task. In other words, the CGAN allows achieving an equilibrium for the minimax game with the probability $D(x \mid y) = 1/2$ [70]. Formally, the value function $V(G,D)$ of the minimax game for estimating a generative model using the CGAN is as follows:

$$\min_G \max_D V(G,D) = \mathbb{E}_{x \sim p_{\text{data}}}[\log D(x \mid y)] + \mathbb{E}_{z \sim p_{\text{noise}}}[\log(1 - D(G(z \mid y)))] \tag{3.2}$$

The reason for using the conditional generative modelling approach in this thesis is since it has become popular in other domains (e.g. image recognition domain) for augmenting training datasets used for predictive modelling [91]. Despite its success in other domains, the following challenges need to be addressed prior to using it for generating plausible failure data: (i) systematically identifying auxiliary information pertaining to failure mechanisms that is useful for generating plausible failure data; (ii) systematically converting different kinds of auxiliary information that are in complex and different forms (e.g. aural, visual and text entries) into a form that is suitable for integrating into the plausible failure data generation process. The solution developed by this research overcomes these challenges, and thus enable the effective implementation of PdM under the conditions of limited failure data availability for prognostics.

3.2 Measuring limited failure data availability

Shannon entropy can be used to measure the balance of a dataset and it calculates the average rate at which the information is produced by a stochastic dataset [92]. Formally, for a dataset D with N number of total data samples and K number of classes with size C_1 to C_K, the normalised Shannon entropy $H'(D)$ is given by Eq. 3.3. This means, when H' gets closer to 0 the extent of the limited failure data availability problem increases, and conversely, when it gets closer to 1 the extent of the problem decreases. This method is used to measure the extent of limited failure data availability problem in different applications studied in this thesis which are discussed later in Chapter 4.

$$H'(D) = \frac{-\sum_{i=1}^{K} \frac{C_i}{N} \log \frac{C_i}{N}}{\log K} \tag{3.3}$$

3.3 Description of methodology and theoretical results

The methodology for prognostics under the conditions of limited failure data availability consists of four phases (see Fig. 3.1). In this section, the prerequisite and assumption of the methodology are stated first. Then the four phases, theoretical results and methods for evaluating the outcome of each phase are described.

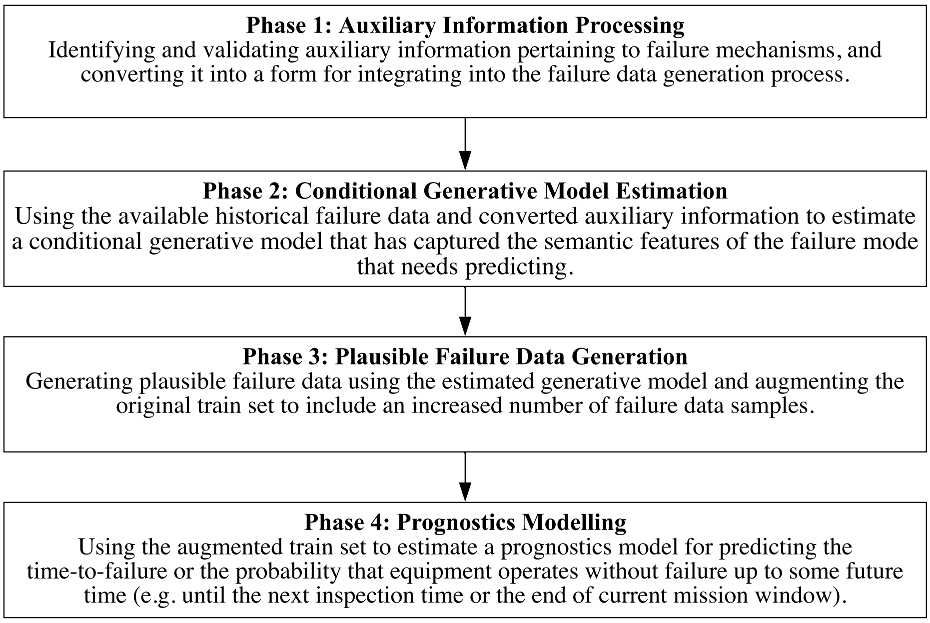


Fig. 3.1 Flowchart outlining the four phases of the methodology.

3.3.1 Prerequisite and assumption

The prerequisite of the methodology is that at least one of the following three kinds of auxiliary information needs to be available in the industrial scenario in addition to condition monitoring and event data which are limited: (i) effect of environmental conditions on failure mechanisms; (ii) effect of harsh use of equipment on failure mechanisms; (iii) similarity between equipment that has failed under a single failure mode. The prerequisite is reasonable in practice since these kinds of information are widely used for maintenance decision making and risk analysis of industrial systems, and available through expert knowledge and historical maintenance records (i.e. inspection, service, repair and replacement records) in prognostics and PdM domains [1, 93].

It is worth mentioning at this point that the effect of improper equipment configuration and customer satisfaction were also considered as auxiliary information in this research. However, it was found that these auxiliary information kinds are not suitable for the methodology as their relationship to the failure mode that needs predicting is complex to identify and difficult to measure statistically.

The following assumption is made in the methodology: the failure mode that needs predicting causes equipment to fail due to gradual degradation and is not a sudden failure. This is a reasonable assumption in practice since the dominant failure modes of industrial equipment cause the hazard rate to be increased with the equipment age [94]. By making this assumption the following is implied: (i) data-driven prognostics using condition monitoring and/or event data is feasible since the evolution of fault into failure causes monotonic trends in equipment condition and performance, and hence data-driven predictive algorithms can use these trends to estimate prognostics model parameters; (ii) since the failure can be predicted, predicting equipment failure is useful as maintenance before failure affects the probability that the equipment will fail in the next instance, hence the unexpected consequences of failures (e.g. unplanned downtime and collateral damage to equipment and processes) can be prevented [94].

3.3.2 Phase 1: Auxiliary information processing

The methodology uses sound engineering knowledge of failure mechanisms as auxiliary information to control and direct the failure data generation process. More specifically, the sound engineering knowledge of failure mechanisms is used to condition the noise being added to the newly generated failure data samples by constructing a conditional generator and discriminator for the CGAN. The aim of the auxiliary information processing phase of the methodology is to identify, validate and convert the sound engineering knowledge of failure mechanisms available in the industrial scenario for the failure mode that needs predicting (also see Fig. 3.2).

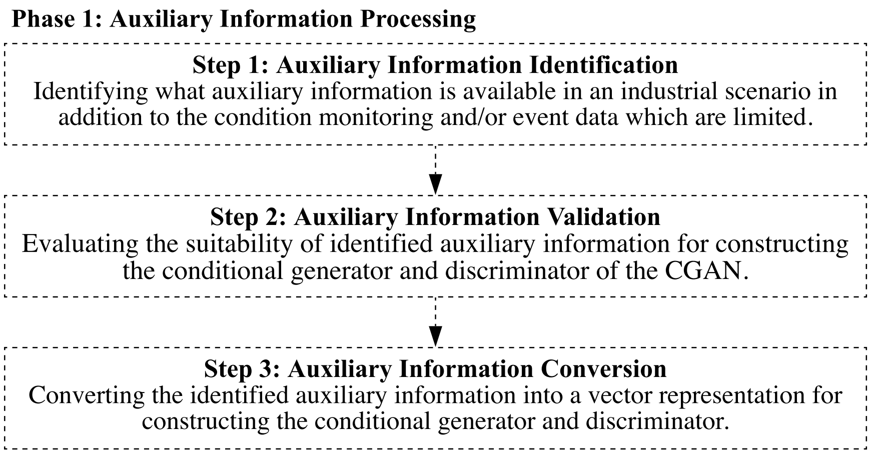


Fig. 3.2 Flowchart outlining the steps of auxiliary information processing phase.

Step 1: Auxiliary information identification

To reiterate, the methodology supports the following three kinds of auxiliary information: (i) effect of environmental conditions on failure mechanisms; (ii) effect of harsh use of equipment on failure mechanisms; (iii) similarity between equipment that has failed under a single failure mode. The objective of this step is to identify potentially suitable pieces of auxiliary information from expert knowledge and historical maintenance records captured during the equipment failure process shown in Fig. 3.3.

The information on the effect of environmental conditions on failure mechanisms refers to the information related to failure mechanisms that are initiated by the deviation from specified

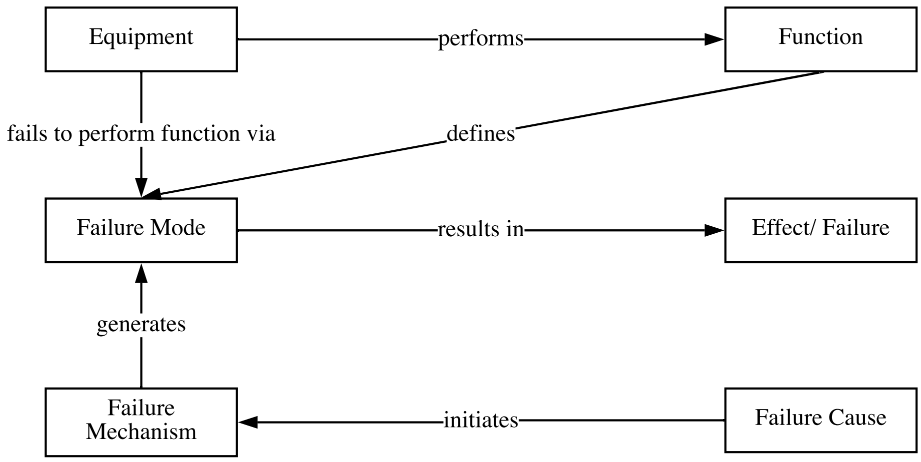


Fig. 3.3 Diagram depicting an example model of the equipment failure process. A failure mode is the effect by which the failure is observed to occur and it is associated with the function of equipment [95]. A failure mechanism is the physical, chemical, thermodynamic or another process that generates a failure mode [95]. A failure cause is an external process, design or environmental condition that initiates a failure mechanism, and whose removal will prevent the failure [95]. A failure mode eventually results in the failure of the equipment (i.e. the state of the equipment in which it is no longer able to perform its indented function with the expected performance) [96].

environmental conditions under which the equipment has to operate. Typically, more than one natural environmental factor (e.g. temperature, rain, salt spray and dust) and/or induced environmental factor (e.g. mechanical and acoustic vibration, radiation and thermal shock) act on the equipment initiating failure mechanisms such as corrosion, thermal fatigue, creep and electromigration [95]. In the methodology, the information on the effect of environmental conditions is identified from on-site maintenance engineers and historical maintenance records. An example of an identified auxiliary information on the effect of environmental conditions on failure mechanisms is as follows: *when the rainfall is increasing and when the driving rain (i.e. rain propelled by a driving wind) is coming from the east, the failure rate of broadband lines is observed to be increasing.* The first case study which is discussed later in Chapter 4 demonstrates how to use this auxiliary information kind in the methodology for generating plausible failure data.

The harsh use of equipment is a common cause of vehicle component failures [97]. Harsh acceleration and braking (i.e. the application of more force than normal to the brake and accelerator systems of a vehicle) which are due to aggressive driving behaviour or “lead foot syndrome” increase wear-and-tear of components [97]. Moreover, excessive use of exhaust brake in heavy vehicles (e.g. heavy-trucks and buses) leads to high wear of turbocharger components since it causes frequent and abrupt changes in turbocharger operating condition [98]. In order to identify the effect of harsh use of equipment on failure mechanisms, the information obtained from equipment operators and on-site maintenance engineers is used. An example of this kind of identified auxiliary information is as follows: *when engine load of a heavy truck is greater than 85% and the engine speed is between 1320* *and 2200 revolutions per minute (RPM), the turbocharger failure rate is observed to be increased*. The third case study of this thesis demonstrates how to use the effect of harsh use of equipment auxiliary information kind in the methodology.

Equipment similarity information can range from metrics such as equipment age, type, model and job type to calculated metrics such as health indicators and natural groupings [71]. Often, a relationship between the similarity between equipment that has failed and degradation patterns can be established since there is an effect of equipment characteristics on the failure process [71, 72]. The information on the similarity between equipment that has failed under a single failure mode is obtained by identifying natural groups of the failed equipment. To this end, cluster analysis and Silhouette analysis (for identifying the number of clusters) are applied to historical condition monitoring and/or event data pertaining to the failure mode that needs predicting [99]. An example of this kind of identified auxiliary information is as follows: *there are 2 groups of heavy-trucks that have failed due to air purge valve surface crack.* The second case study of this thesis demonstrates the use of this auxiliary information kind in the methodology.

Step 2: Auxiliary information validation

The objective of this step is to validate the identified auxiliary information using the statistical test framework (i.e. significance-based hypothesis testing). To this end, data that are suitable for representing the identified auxiliary information is obtained since it is the data that needs to be used as inputs to statistical tests. Table 3.1 outlines data sources and metrics used for representing the identified auxiliary information as data for the corresponding statistical tests[1].

Table 3.1 Data sources and metrics used for representing the identified auxiliary information as data to the statistical test framework.

Auxiliary information	Data sources	Metrics	Statistical test
Effect of environmental conditions on failure mechanisms	Online weather databases: OpenWeatherMap API [101] and Met Office Data Point [102].	The failure rate in different thresholds of environmental factors (e.g. low, medium and high temperature, humidity, rainfall, snow, atmospheric air pressure and wind speed).	Welch's t-test
Effect of harsh use of equipment on failure mechanisms	Equipment load matrix.	The failure rate in different regions of the load matrix.	Welch's t-test
Similarity between equipment that has failed under a single failure mode	Natural groups of failed equipment identified using cluster analysis and Silhouette analysis.	The number of natural groups of equipment that has failed.	Welch's t-test or cluster analysis and Silhouette analysis

[1] Since the objective is to measure the statistical similarity between samples drawn from two distributions, Welch's t-test does not make any assumptions on the modality of the distributions (i.e. unimodal, bimodal and multimodal) [100].

In order to apply the statistical test framework to validate the identified auxiliary information, a null hypothesis or an alternative hypothesis needs to be defined. In the methodology, the null hypotheses outlined in Table 3.2 are used and adapted to the equipment and failure mode of interest. Once the null hypothesis is defined, the Welch's t-test is performed to identify whether the corresponding null hypothesis can be rejected. If the p-value (probability value) of the statistical test is greater than 0.05, the null hypothesis is retained since there is weak evidence against it. In other words, the identified auxiliary information is considered to be valid. If the p-value is less than or equal to 0.05, the null hypothesis is rejected and the identified auxiliary information is considered to be invalid since there is strong evidence against the null hypothesis.

Table 3.2 Null hypotheses used for validating the identified auxiliary information using the statistical test framework.

Auxiliary information	Null hypothesis
Effect of environmental conditions on failure mechanisms	There is an increase in equipment failures when a certain environmental condition is increasing or decreasing.
Effect of harsh use of equipment on failure mechanisms	There is an increase in equipment failures when the equipment is operating in certain regions of the load matrix.
Similarity between equipment that has failed under a single failure mode	There is a certain number of groups of equipment that has failed under the failure mode that needs predicting (for Welch's t-test) or the natural groupings of equipment have a low intra-cluster distance and a high inter-cluster distance (for cluster analysis).

Step 3: Auxiliary information conversion

The validated auxiliary information pertaining to failure mechanisms needs to be converted into a form that can be integrated into the CGAN described in Sec. 3.1. More specifically, it needs to be represented as a vector Y with samples $\{y \in Y\}$ so that the conditional generator $G(Z \mid Y)$ which generates plausible failure data samples from the joint distribution of noise and auxiliary information $Pr(Z,Y)$, and the conditional discriminator $D(X \mid Y)$ which detects whether a given failure data sample is sampled from the joint distribution of real failure data and auxiliary information $Pr(X,Y)$ can be constructed. In the methodology, two approaches are taken to convert the validated auxiliary information pertaining to failure mechanisms into vector representations.

(I) Constructing an auxiliary information vector from a continuous distribution

This approach is applicable to auxiliary information on the effect of environmental conditions on failure mechanisms and the focus is on monotonic trends (i.e. trends that are either increasing or decreasing) in data captured for natural and induced environmental factors. It involves first representing data pertaining to a monotonic trend of an environmental factor in the statistical form (i.e. a continuous random variable which takes on the all values in an interval of numbers). Then representing the continuous random variable as a continuous distribution to construct a vector Y

containing some values that are either increasing or decreasing. More specifically, Y contains some values $\{y \in Y \mid y_0 < y < y_1,$ and y increases$\}$ or $\{y \in Y \mid y_0 > y > y_1,$ and y decreases$\}$. This approach is further explained using the below example which involves using expert knowledge on the effect of environmental conditions on failure mechanisms as auxiliary information.

Imagine that there is a set of outdoor electronic equipment that has failed. Over time, maintenance engineers have gained knowledge that the failure rate of this set of equipment increases during the rainy season. The reason for this is, driving rain causes water ingress (i.e. failure cause) into outdoor electronic equipment and initiates failure mechanisms such as corrosion and electrical shorts. The identified auxiliary information is validated by performing a statistical hypothesis test using data extracted from historical maintenance records and weather reports.

In order to convert the validated auxiliary information, it is first converted into a form in which all the equipment-specific information is removed (hereinafter referred to as the abstract form). This allows auxiliary information to be generalised to all the equipment that has failed under the failure mode that needs predicting. For instance, if the rainfall in a particular location where the equipment placed at during their degradation period is recorded as *the rainfall at the location where the equipment A, B and C placed at increased from 43 mm to 65 mm*, once converted into the abstract form this information becomes *some variable X increases*. Thus, specific terms such as equipment A, B and C, rainfall and numerical thresholds are ignored. Then the abstracted information is converted into the statistical form by representing it as some continuous variable C. The continuous variable C can be converted into a distribution between some values y_0 and y_1. This distribution can be represented as a vector Y containing some values $\{y \in Y \mid y_0 < y < y_1,$ and y increases$\}$ to obtain the auxiliary information vector which will be used to construct the conditional generator and discriminator of the CGAN.

(II) Constructing an auxiliary information vector from class labels

This approach is applicable to the effect of harsh use of equipment on failure mechanisms and similarity between equipment that has failed under a single failure mode auxiliary information kinds. It involves using critical regions in equipment load matrix (in the case of harsh use of equipment) or natural groups of equipment that has failed (in the case of similarity between equipment that has failed) to construct an auxiliary information vector. More specifically, each critical region in an equipment load matrix (i.e. regions in which the failure rate of the equipment is high) is given the following labels: 1 for the first region, 2 for the second region, 3 for the third region and so on. The natural groups of equipment that has failed under a single failure mode are identified by performing cluster analysis on condition monitoring and/or event data captured during the past degradation processes. Then each natural group is given the following labels: 0 for the first group, 1 for the second group, 2 for the third group and so on.

After the class labels are identified, each data sample in the dataset used for prognostics modelling is labelled using the critical region label the sample belongs to (in the case of harsh use of equipment) or the natural group label the equipment belongs to (in the case of similarity between equipment that has failed). In the case of critical regions, 0 is used to label all the data samples that do not belong to any of the critical regions. Thus, the prognostics modelling dataset now has a new label column which will be used as the auxiliary information vector to construct the conditional generator and discriminator of the CGAN.

3.3.3 Phase 2: Conditional generative model estimation

This phase of the methodology involves integrating the converted vector of auxiliary information pertaining to failure mechanisms and the limited amount of real failure data samples available to the CGAN to estimate a generative model that is capable of generating plausible failure data samples. To this end, the steps outlined in Fig. 3.4 are performed.

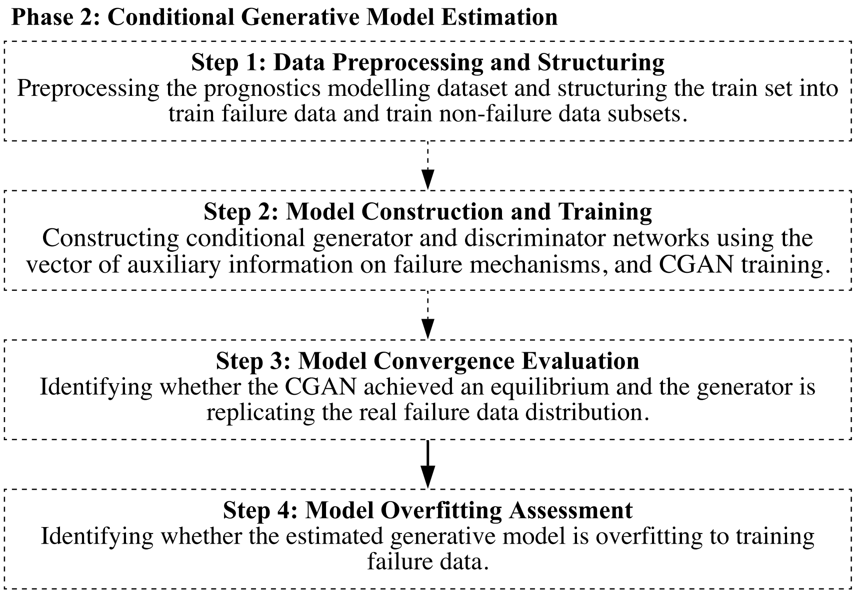


Fig. 3.4 Flowchart outlining the steps of conditional generative model estimation phase.

Step 1: Data preprocessing and structuring

The objective of this step is to preprocess and structure the prognostics modelling dataset into a form suitable for integrating into the failure data generation process. As shown in Fig. 3.5, preprocessing starts with removing the low variance features due to their low predictive power. Then the dataset is converted into a run-to-failure dataset by removing the parts of time series that belong to the time after the failure and retaining a sufficient amount of data pertaining to the normal condition of the equipment.

Splitting a dataset into train, validation and test sets without stratification is not suitable for evaluating predictive models under the data imbalance since this will not maintain the skewness between class distributions across the splits [103]. This means, some splits are likely to have few or no examples from the minority class (i.e. failure data class), and hence the model performance will be misleading as the model needs to only predict the majority class (i.e. non-failure data class) correctly. Whilst k-fold cross-validation is widely used for evaluating predictive models when data belonging to all the classes are limited, it is not suitable for evaluating predictive models under the data imbalance since it is likely the skewness between class distributions are not maintained across the folds [103].

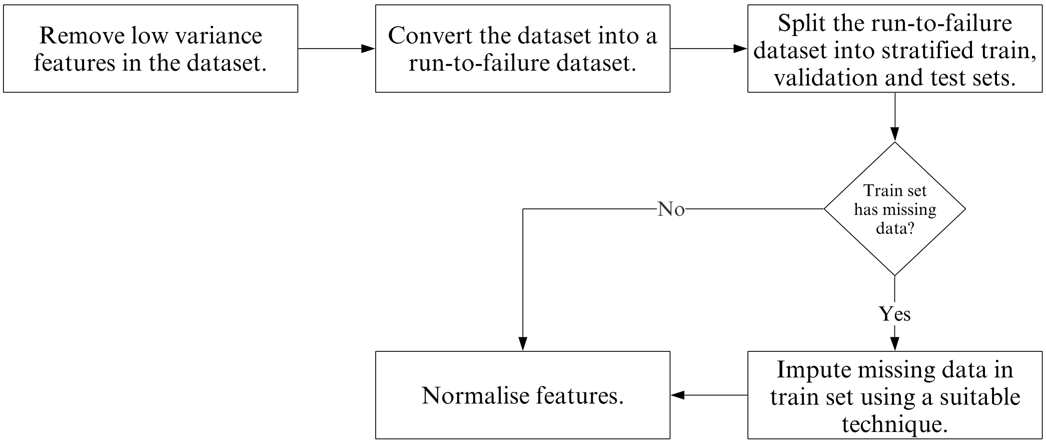


Fig. 3.5 Flowchart of the general steps used for data preprocessing in the methodology.

Hence in this methodology, prognostics modelling datasets are divided into stratified train, validation and test sets using stratification which allows splitting a dataset into multiple subsets whilst preserving the skewness between class distributions [103][2]. The stratified train set contains 60% of the data and is used for training prognostics models. The stratified validation set which contains 20% of the data is used for model hyperparameter tuning. The stratified test set contains 20% of the data and is used to evaluate prognostics models on previously unseen data.

As shown in Fig. 3.5, if the train set contains missing data, it needs to be imputed using a suitable missing data imputation technique (e.g. mean imputation, time interpolation and k-nearest neighbour imputation). The most suitable technique depends on the domain and it is selected by observing the prognostics performance obtained after employing different imputation techniques. Finally, the data are normalised in order to transform all the features into a comparable scale. The transformation model for normalisation is calculated using the train set and then applied to all the three sets. In order to improve the prognostics performance obtained at the end of methodology application, data preprocessing steps are revisited. For example, the low variance features removed at the first iteration are used in the second iteration to identify whether they improve the predictive power of prognostics models at least by a small margin.

After the prognostics modelling dataset is preprocessed, it needs to be structured as shown in Fig. 3.6. More specifically, the objective of generating plausible failure data is to augment the stratified train set A (hereinafter referred to as the original train set) so that the number of failure samples

[2] In this methodology, the Scikit learn Split function is used to split datasets using stratification [104].

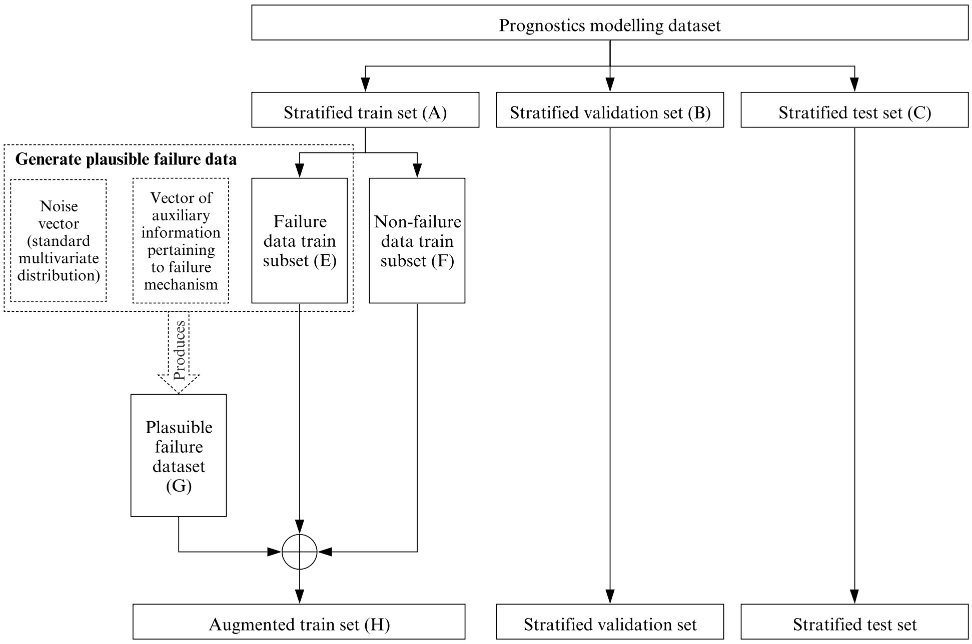


Fig. 3.6 Diagram depicting how the prognostics modelling dataset is structured in the methodology for plausible failure data generation, prognostics model estimation and evaluation.

available for training prognostics models is increased. To this end, the original train set is further split into failure data train subset E and non-failure data train subset F. The former is used with the vector of auxiliary information on failure mechanisms and a noise vector to estimate a generative model that has captured the semantic features of the failure mode that needs predicting (i.e. a generative model that is capable of replicating the real failure data distribution). After the new dataset containing plausible failure data samples is generated (dataset G in the figure), it is combined with subsets E and F to obtain the augmented train set H. The stratified validation set B and stratified test set C are left unchanged for model hyperparameter tuning, and prognostics performance evaluation and prognostics model uncertainty quantification respectively.

Step 2: Model construction and training

In this step, the CGAN architecture presented in Fig. 3.7 is implemented and used as the platform to integrate the vector of auxiliary information pertaining to failure mechanisms and the train failure data subset. Then it is used to estimate a generative model that has captured the semantic features of the failure mode that needs predicting. The architecture is implemented using the theoretical aspects and the value function $V(G,D)$ of CGAN discussed in Sec. 3.1. The key difference of this architecture to the standard CGAN architecture is the use of auxiliary information pertaining to failure mechanisms as the conditioning vector Y for constructing the conditional generator $G(Z \mid Y)$ and discriminator $D(X \mid Y)$.

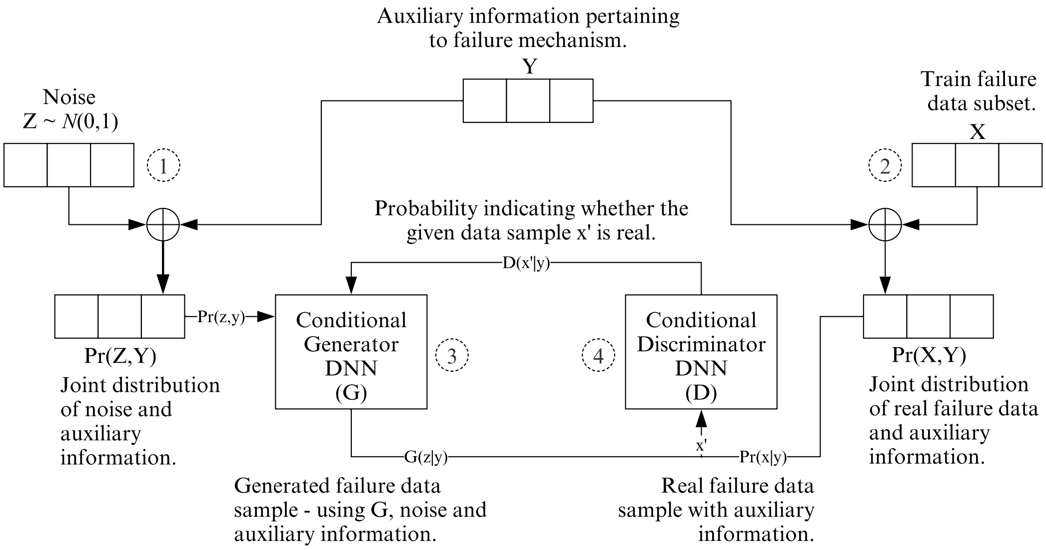


Fig. 3.7 Diagram depicting the architecture of the conditional generative adversarial network implemented for the methodology. The generator and discriminator are deep neural networks (DNN).

As shown in Fig. 3.7, the first step is to combine the noise vector Z with the auxiliary information vector Y into the joint distribution $Pr(Z,Y)$. Z is formed from a standard multivariate normal distribution such that $Z \sim N(0,1)$, where 0 is the mean and 1 is the variance. In order to accomplish this, the approach proposed in Xu et al. [105] for developing $G(Z \mid Y)$ and $D(X \mid Y)$ using tabular data is used. More specifically, each row in the noise vector is given a label from the auxiliary information vector. Then the entire row which consists of noise data and a label is provided as the input data sample (i.e. $Pr(z,y)$) to the generator DNN. Similarly, the data samples in the train failure data subset X are combined with the auxiliary information vector Y into the joint distribution $Pr(X,Y)$. This is used as the input to the discriminator DNN. The objective of the generator G is to learn to fool the discriminator D into believing that a generated failure data sample is real (i.e. the generated failure data sample is sampled from the real failure data distribution) during each training iteration. Thus, the generator produces a fake failure data sample $G(z \mid y)$ by conditioning the noise z on auxiliary information pertaining to failure mechanisms y. More specifically, the generator aims to minimise its loss function $\log(1 - D(G(z \mid y)))$ (i.e. learn to fool the discriminator the most) [70].

The objective of the discriminator D is to detect whether a given failure data sample is real. It produces a probability $D(x' \mid y)$ indicating how much it believes the given failure data sample x' is sampled from the joint distribution $Pr(X,Y)$. More specifically, the discriminator tries to minimise its loss function $\log(D(x' \mid y))$ (i.e. learn to discriminate between real and fake failure data samples better) [70]. During the training, the probability $D(x' \mid y)$ is given as the feedback to the generator to allow it to converge the generated failure data distribution to the real failure data distribution. At the end of the training, that is, when $D(x' \mid y) = 1/2$, the generated failure data distribution is converged to the real failure data distribution, and hence the generator is capable of generating plausible failure data samples that the discriminator cannot discriminate as real or fake. In practice however, it is reasonable to expect $D(x' \mid y)$ to be closer to $1/2$ due to the uncertainty associated with hyperparameter tuning of generator and discriminator DNNs [69].

The CGAN is trained using the minibatch stochastic gradient descent algorithm. The loss functions of discriminator and generator DNNs are obtained by decomposing the value function of the minimax game in CGAN. More specifically, for the joint distribution of the real failure data and auxiliary information pertaining to the failure mechanisms $p_{\text{data}}(x,y)$, and the joint distribution of noise and auxiliary information pertaining to the failure mechanisms $p_{\text{noise}}(z,y)$, decomposing the value function $V(G,D)$ into the loss functions of discriminator D and generator G gives Eq. 3.4 and 3.5 respectively.

$$\max_D V(D) = \overbrace{\mathbb{E}_{x \sim p_{\text{data}}}[\log D(x \mid y)]}^{\text{Recognise real failure data samples better}} + \underbrace{\mathbb{E}_{z \sim p_{\text{noise}}}[\log(1 - D(G(z \mid y)))]}_{\text{Recognise generated failure data samples better}} \tag{3.4}$$

$$\min_G V(G) = \underbrace{\mathbb{E}_{z \sim p_{\text{noise}}}[\log(1 - D(G(z \mid y)))]}_{\text{Optimise G to fool the discriminator the most}} \tag{3.5}$$

Using the loss function of discriminator D given in Eq. 3.4, the total loss of D for a minibatch of m examples $\{xy^{(1)}, ..., xy^{(m)}\}$ from $p_{\text{data}}(x,y)$, and for a minibatch of m examples $\{zy^{(1)}, ..., zy^{(m)}\}$ from $p_{\text{noise}}(z,y)$ is,

$$\frac{1}{m}\sum_{i=1}^{m}[\log D(xy^{(i)}) + \log(1 - D(G(zy^{(i)})))]$$

Similarly, using the loss function of generator G given in Eq. 3.5, the total loss of G for a minibatch of m examples $\{zy^{(1)}, ..., zy^{(m)}\}$ from $p_{\text{noise}}(z,y)$ is,

$$\frac{1}{m}\sum_{i=1}^{m}\log(1 - D(G(zy^{(i)})))$$

The loss functions are used as the objective functions of the minibatch stochastic gradient descent algorithm (see Algorithm 1). The discriminator DNN is executed twice per training iteration before it calculates the total loss: once for real failure data and once for generated failure data so that it learns to discriminate between real and fake data samples (also see Eq. 3.4). The generator DNN is executed only once per training iteration. When the discriminator and generator losses are known, the gradients with regard to their parameters are calculated and backpropagated through the discriminator and generator DNNs to optimise their model parameters (i.e. model weights and biases).

Step 3: Model convergence evaluation

Unlike other ANNs which typically fail to converge when the model loss does not stabilise during training, a failure to converge in GANs refers to not finding an equilibrium between the generator and discriminator [69]. In order to identify whether the CGAN implemented for the methodology has converged, learning curves of generator and discriminator DNNs are analysed first. Then using the Kolmogorov-Smirnov test (K-S test), the statistical similarity between generated failure data samples

Algorithm 1: Minibatch stochastic gradient descent training of the conditional generative adversarial network implemented for the methodology [70, 69].

for *number of training iterations* **do**
  for *number of steps to apply to the discriminator* **do**
  - Sample a minibatch of m examples $\{zy^{(1)},...,zy^{(m)}\}$ from the joint distribution of noise and auxiliary information pertaining to failure mechanisms $p_{\text{noise}}(z,y)$.
  - Sample a minibatch of m examples $\{xy^{(1)},...,xy^{(m)}\}$ from the joint distribution of real failure data and auxiliary information pertaining to failure mechanisms $p_{\text{data}}(x,y)$.
  - Update the discriminator model weights and biases by ascending its stochastic gradient:
  $$\bigtriangledown\theta_d \frac{1}{m}\sum_{i=1}^{m}[\log D(xy^{(i)}) + \log(1 - D(G(zy^{(i)})))]$$

  end
  - Sample a minibatch of m examples $\{zy^{(1)},...,zy^{(m)}\}$ from the joint distribution of noise and auxiliary information pertaining to failure mechanisms $p_{\text{noise}}(z,y)$.
  - Update the generator model weights and biases by descending its stochastic gradient:
  $$\bigtriangledown\theta_g \frac{1}{m}\sum_{i=1}^{m}\log(1 - D(G(zy^{(i)}))]$$

end

and real failure data samples is measured to identify whether the generator is capable of replicating the real failure data distribution. This is since at the equilibrium, the generator is expected to have learnt the real failure data distribution.

If estimating a generative model by conditioning the generator and discriminator on auxiliary information pertaining to failure mechanisms has been successful, the CGAN is expected to produce learning curves similar to the ones shown in Fig. 3.8. More specifically, the learning curves are expected to have the following properties: (i) discriminator loss on real and generated failure data samples is expected to remain around $1/2$ which indicates that the discriminator cannot discriminate between real and generated failure data samples; (ii) generator loss is expected to stabilise as the number of training iterations increases which indicates that the generator is consistent; (iii) variance of generator and discriminator loss is expected to remain modest [69].

Whilst the analysis of learning curves provides a visual indication of whether the convergence is achieved, a more robust evaluation needs to be performed from a theoretical point of view. To this end, this methodology recommends the K-S test for statistically identifying whether the convergence is achieved since it is widely used for GAN convergence evaluation [106]. The K-S test provides a statistical test framework for detecting whether two samples are drawn from the same distribution. The null hypothesis of this test is that the two distributions are statistically similar. Thus, the objective of this evaluation is to statistically identify whether the following null hypothesis can be rejected: *the*

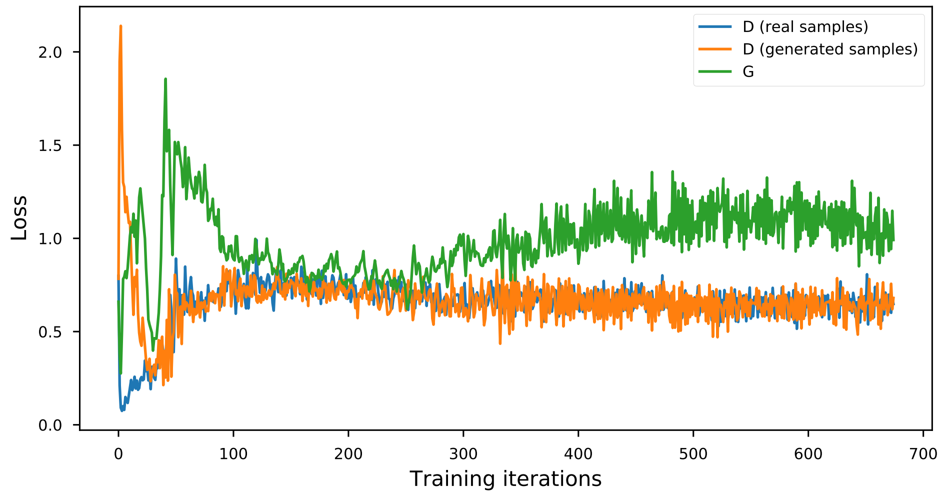


Fig. 3.8 An example of generator (G) and discriminator (D) learning curves expected of a generative adversarial network that achieved an equilibrium between G and D [69].

generated failure data distribution produced by the generator and the real failure data distribution are statistically similar. If the null hypothesis can be retained, convergence is achieved. This means, the generator has captured the semantic features of the failure mode (i.e. capable of replicating the real failure data distribution), and hence capable of generating plausible failure data samples for predicting the failure mode. If the null hypothesis cannot be retained, the convergence is not achieved. In this case, one needs to retune the hyperparameters of the CGAN and may also need to increase the number of training iterations until the convergence is achieved.

Since its ability to replicate the real failure data distribution is integral to the methodology, a theoretical grounding for this claim is also developed in this thesis. To this end, the mathematical proof of GAN provided in [69] is adopted and extended to include the joint probability distribution which allows the conditioning of noise on auxiliary information pertaining to failure mechanisms.

Using Lemma 1, the following is theorised first: the CGAN implemented for the methodology allows estimating the optimal discriminator which can perfectly discriminate between real and generated failure data samples for the fixed generator.

Lemma 1. *Given the joint probability distribution of real failure data $p_{data}(x,y)$ and generated failure data $p_G(x,y)$, where $\{x \in X\}$ consists of real and generated input failure data samples and $\{y \in Y\}$ is the vector of auxiliary information on failur mechanisms, the optimal discriminator D^*_G for the fixed generator is,*

$$D^*_G(x \mid y) = \frac{p_{data}(x,y)}{p_{data}(x,y) + p_G(x,y)} \tag{3.6}$$

Intuitively, this lemma states that when the real failure data distribution and generated failure data distribution are given, the optimal discriminator should be able to identify the real failure data fraction. The proof of Lemma 1 is as follows:

Proof. Consider the value function $V(G,D)$ of the minimax game in CGAN provided in Sec. 3.1,

$$V(G,D) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x \mid y)] + \mathbb{E}_{z\sim p_{\text{noise}}}[\log(1 - D(G(z \mid y)))]$$

Using the law of unconscious statistician (LOTUS) theorem,

$$\mathbb{E}_{z\sim p_{\text{noise}}}[\log(1 - D(G(z \mid y)))] = \mathbb{E}_{x\sim p_G}[\log(1 - D(x \mid y))]$$

Therefore $V(G,D)$ can be written as,

$$V(G,D) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x \mid y)] + \mathbb{E}_{x\sim p_G}[\log(1 - D(x \mid y))]$$

For a pair of continuous random variables X and Y with a joint probability distribution $Pr(x = X, y = Y)$, the expected value $\mathbb{E}$ can be found using an arbitrary function of the continuous variables $g(X,Y)$ such that,

$$\mathbb{E}[g(X,Y)] = \int_x \int_y p(x,y)\; g(x,y)\; dx\; dy$$

Therefore, $V(G,D)$ can be written as,

$$V(G,D) = \int_x \int_y p_{\text{data}}(x,y) \log D(x \mid y)\; dx\; dy + \int_x \int_y p_G(x,y) \log(1 - D(x \mid y))\; dx\; dy$$

Using the sum rule of integration,

$$V(G,D) = \int_x \int_y p_{\text{data}}(x,y) \log D(x \mid y) + p_G(x,y) \log(1 - D(x \mid y))\; dx\; dy$$

For clarity, we label the above equation as follows:

$$h = D(x \mid y),\; a = p_{\text{data}}(x,y),\; b = p_G(x,y)$$

Since the sample x given y is sampled over all the possible values, the integrals can be safely ignored for the remainder of the proof. Therefore $V(G,D)$ becomes,

$$f(h) = a \log h + b \log(1 - h)$$

The objective is to find the best value for $D(x \mid y)$ to maximise the value function $V(G,D)$. The maximum of $V(G,D)$ can be found as follows: (i) differentiate $V(G,D)$ w.r.t $D(x \mid y)$ and equate to zero to find the critical points; (ii) perform the second derivative test for local extrema (i.e. h is maximum if $f''(h) < 0$) to find the maximum. Hence, we first differentiate $f(h)$ w.r.t to h,

$$f'(h) = \frac{a}{h} - \frac{b}{1-h}$$

Equating $f'(h)$ to zero to find the critical points,

$$\frac{a}{h} - \frac{b}{1-h} = 0$$

Using calculus,

$$h = \frac{a}{a+b}$$

if $a+b \neq 0$. Finding $f''(h)$ to identify whether h is the maximum using second derivative test,

$$f''(\frac{a}{a+b}) = -\frac{a}{a/(a+b)^2} - \frac{b}{(1-(a/a+b))^2}$$

Since $f''(h) < 0$, h is the maximum. Hence, for the fixed generator G the optimal discriminator $D^*_G(x \mid y) = h$. After substituting a and b labels in h with their corresponding values, $D^*_G(x \mid y)$ can be written as,

$$D^*_G(x \mid y) = \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)}$$

, concluding the proof. □

Secondly, using Lemma 1 and below Theorem 1, the following is theorised: the CGAN implemented for the methodology can be trained to estimate the optimal generator which can perfectly replicate the real failure data distribution.

Theorem 1. *The training criterion $C(G) = \min_G \max_D V(G,D)$ achieves a unique global minimum if and only if the generated failure data distribution p_G is equal to the real failure data distribution p_{data} (i.e. $p_G = p_{data}$).*

Proof. Assuming $p_G = p_{\text{data}}$ and using Lemma 1, the optimal discriminator $D^*_G(x \mid y)$ is,

$$D^*_G(x \mid y) = \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_{\text{data}}(x,y)} = \frac{1}{2}$$

Consider the integral form of the value function $V(G,D)$ introduced in Lemma 1,

$$V(G,D) = \int_x \int_y p_{\text{data}}(x,y) \log D(x \mid y) + p_G(x,y) \log(1 - D(x \mid y))\, dx\, dy$$

When $D(x \mid y) = D^*_G(x \mid y) = 1/2$, $C(G)$ is,

$$C(G) = \int_x \int_y p_{\text{data}}(x,y) \log \frac{1}{2} + p_G(x,y) \log(1 - \frac{1}{2})\, dx\, dy$$

As per the assumption made in the beginning of the proof, $p_G = p_{\text{data}}$ when $D^*_G(x \mid y) = 1/2$. Hence,

$$C(G) = \int_x \int_y p_{\text{data}}(x,y) \log \frac{1}{2} + p_G(x,y) \log(1 - \frac{1}{2})\, dx\, dy = -2\log 2$$

This means $-2\log 2$ is a candidate for the global minimum of the training criterion $C(G)$. However, we still need to prove that this is the only global minimum of $C(G)$ and it is achieved only when the generated failure data distribution p_G is equal to the real failure data distribution p_{data}. Hence, we first drop the assumption $p_G = p_{\text{data}}$. Again consider the integral form of the value function $V(G,D)$ introduced in Lemma 1,

$$V(G,D) = \int_x \int_y p_{\text{data}}(x,y) \log D(x \mid y) + p_G(x,y) \log(1 - D(x \mid y))\, dx\, dy$$

When $D(x \mid y) = D^*_G(x \mid y)$, $V(G,D)$ is,

$$V(G,D^*_G) = \int_x \int_y p_{\text{data}}(x,y) \log D^*_G(x \mid y) + p_G(x,y) \log(1 - D^*_G(x \mid y))\, dx\, dy$$

For the training criterion $C(G) = V(G,D^*_G)$ and since $D^*_G(x \mid y)$ is as per the equation proved in Lemma 1, $C(G)$ is,

$$C(G) = \int_x \int_y p_{\text{data}}(x,y) \log \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)} + p_G(x,y) \log(1 - \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)})\, dx\, dy$$

After using calculus to rearrange the second part of the integral,

$$C(G) = \int_x \int_y p_{\text{data}}(x,y) \log \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)} + p_G(x,y) \log \frac{p_G(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)}\, dx\, dy$$

Since we know $-2\log 2$ is a candidate for the global minimum, we integrate this value into $C(G)$ by adding and subtracting $\log 2$ and multiplying by the joint probability densities. This do not change the equality of the equation since ultimately we are adding 0 to the equation. Therefore,

$$\begin{aligned} C(G) = \int_x \int_y & (\log 2 - \log 2) p_{\text{data}}(x,y) + p_{\text{data}}(x,y) \log \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)} \\ & + (\log 2 - \log 2) p_G(x,y) + p_G(x,y) \log \frac{p_G(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)}\, dx\, dy \end{aligned}$$

Using calculus $C(G)$ can be rearranged as,

$$\begin{aligned} C(G) = -\log 2 \int_x \int_y p_G(x,y) + p_{\text{data}}(x,y)\, dx\, dy + \int_x \int_y p_{\text{data}}(x,y)(\log 2 + log \frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)}) \\ + p_G(x,y)(\log 2 + \log \frac{p_G(x,y)}{p_{\text{data}}(x,y) + p_G(x,y)})\, dx\, dy \end{aligned}$$

The definition of probability densities states that integrating two probability distributions over their domain is equal to 1. Hence,

$$-\log 2 \int_x \int_y p_G(x,y) + p_{\text{data}}(x,y)\, dx\, dy = -\log 2(1+1) = -2\log 2$$

Moreover, using the definition of the logarithm,

$$\log 2 + log\frac{p_{\text{data}}(x,y)}{p_{\text{data}}(x,y)+p_G(x,y)} = log\frac{p_{\text{data}}(x,y)}{(p_{\text{data}}(x,y)+p_G(x,y))/2}$$

Similarly,

$$\log 2 + \log\frac{p_G(x,y)}{p_{\text{data}}(x,y)+p_G(x,y)} = \log\frac{p_G(x,y)}{(p_{\text{data}}(x,y)+p_G(x,y))/2}$$

Substituting above equalities into $C(G)$,

$$\begin{aligned} C(G) = -2log2 \int_x \int_y p_{\text{data}}(x,y) log\frac{p_{\text{data}}(x,y)}{(p_{\text{data}}(x,y)+p_G(x,y))/2} \\ +p_G(x,y) log\frac{p_G(x,y)}{(p_{\text{data}}(x,y)+p_G(x,y))/2}\, dx\, dy \end{aligned}$$

Using the sum rule of integration $C(G)$ can be written as,

$$\begin{aligned} C(G) = -2log2 \int_x \int_y p_{\text{data}}(x,y) log(\frac{p_{\text{data}}(x,y)}{(p_{\text{data}}(x,y)+p_G(x,y))/2})\, dx\, dy \\ + \int_x \int_y p_G(x,y) log(\frac{p_G(x,y)}{(p_{\text{data}}(x,y)+p_G(x,y))/2})\, dx\, dy \end{aligned}$$

Using the Kullback-Leibler divergence (KL),

$$C(G) = -2log2 + KL(p_{\text{data}}|\frac{p_{\text{data}}+p_G}{2}) + KL(p_G|\frac{p_{\text{data}}+p_G}{2})$$

Using the Jenson-Shannon divergence (JSD),

$$C(G) = -2log2 + 2.JSD(p_{\text{data}}||p_{\text{G}})$$

Rearranging $C(G)$ using calculus,

$$C(G) - (-2log2) = 2.JSD(p_{\text{data}}||p_{\text{G}})$$

JSD between two distributions is non-negative. Hence,

$$C(G) - (-2log2) \geq 0$$

Therefore, we have a unique global minimum for the training criterion $C(G)$ when JSD is equal to 0,

$$C(G)_{min} = -2log2$$

This means the training criterion $C(G)$ can achieve a unique global minimum of $-2\log 2$ when the optimal discriminator $D^*_G(x \mid y) = 1/2$. Since $D^*_G(x \mid y) = 1/2$ only when $p_G = p_{\text{data}}$, the global minimum is achieved only when the generated failure data distribution p_G is equal to the real failure data distribution p_{data}, concluding the proof. □

In order to show that the training algorithm (i.e. Algorithm 1) can be used to converge the generated failure data distribution p_G to the real-failure data distribution p_{data}, and hence reach the unique global minimum provided in Theorem 1, the equilibrium point of the two-player minimax game is considered. During the two-player minimax game, the discriminator D tries to maximise the value function $V(G,D)$ for a given generator G whilst G tries to minimise it for the optimal D. Thus, the objective is to reach a saddle point as illustrated in Fig. 3.9. This saddle point is the equilibrium point of the minimax game. Moreover, according to Theorem 1, we know that at the equilibrium point the training criterion $C(G)$ achieves the unique global minimum and at that point, we have the optimal discriminator $D^*_G = 1/2$, and a generative model that is perfectly replicating the real failure data distribution. Hence, if we are able to show that the value function $V(G,D)$ can reach the saddle point through gradient descent, we have shown that Algorithm 1 converges the generated failure data distribution to the real-failure data distribution (i.e. $p_G = p_{\text{data}}$).

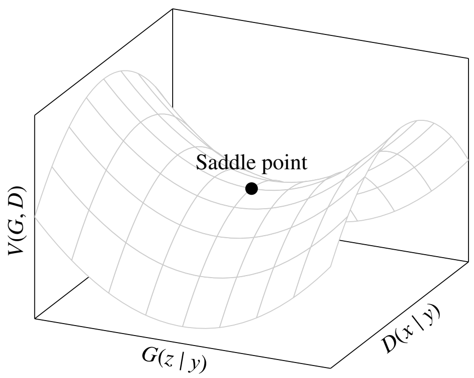


Fig. 3.9 Plot depicting the saddle point of the value function $V(G,D)$ of minimax game. The objective of training the proposed CGAN architecture is to reach the saddle point and hence achieve the unique global minimum of the training criterion $C(G) = \min_G \max_D V(G,D)$.

In order to reach the saddle point, gradient descent is applied to discriminator D for each fixed point of p_G to get the optimum D for p_G as shown in the inner loop of Algorithm 1. Then keeping D fixed, gradient descent is applied to generator G to get closer to the saddle point (see the outer loop of Algorithm 1). Since partial derivatives of p_G at optimum D points include the saddle point, given enough capacity (i.e. computation power and training time) to discriminator and generator DNNs, we will eventually reach the saddle point using the training algorithm. Thus, Algorithm 1 converges the generated failure data distribution p_G to the real failure data distribution p_{data}, and hence reaches the unique global minimum provided in Theorem 1.

Step 4: Model overfitting assessment

Even though the convergence evaluation presented in previous step has shown that the generator has converged the generated failure data distribution to the real failure data distribution, this result can be biased if the generator is overfitting to real failure data samples in the original train set (i.e. train failure data samples). The maximum mean discrepancy (MMD) is proven to be a suitable technique for assessing model overfitting in CGANs [91]. It measures the distance between distributions to identify whether two data samples are generated from different distributions [107].

In the methodology, MMD is used to quantify the mean discrepancy between real failure data and generated failure data. If the generator is overfitting to the train failure data samples, the MMD between generated failure data samples and train failure data samples should be significantly lower than generated failure data samples and real failure data samples in the test set (i.e. test failure data samples). Thus, the objective of this evaluation is to statistically identify whether the following null hypothesis can be rejected: *the generator is not overfitting to train failure data*. If the null hypothesis can be retained, the MMD between generated failure data samples and test failure data samples is at most as large as the MMD between generated failure data samples and train failure data samples. In other words, the objective is to identify whether the following statement holds: generated failure data samples do not look more similar to the train failure data samples than they do to the test failure data samples, hence the generator is not overfitting to train failure data. If the null hypothesis cannot be retained, the convergence evaluation is affected by generator overfitting to train failure data samples. In this case, one needs to retune the hyperparameters of the CGAN and repeat the convergence evaluation and overfitting assessment.

3.3.4 Phase 3: Plausible failure data generation

In this phase, the trained generator is used as the generative model to produce plausible failure data samples, and thus augment the original train set so that an increased number of failure data samples is available for prognostics modelling. The steps outlined in Fig. 3.10 are followed in this phase.

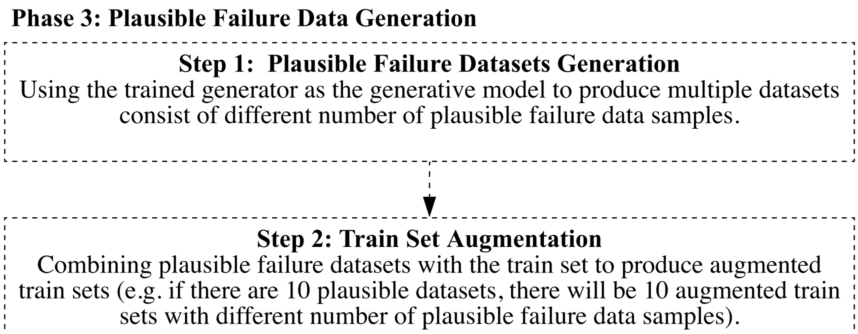


Fig. 3.10 Flowchart outlining the steps of plausible failure data generation phase.

Step 1: Plausible failure datasets generation

As shown in Fig. 3.11, a new noise vector which is sampled from a standard multivariate normal distribution (i.e. $Z \sim N(0,1)$, where 0 is the mean and 1 is the variance) and the vector of auxiliary information on failure mechanisms are used as inputs to the generative model to produce a dataset consists of new failure data samples. More specifically, given the joint distribution of noise and auxiliary information $Pr(z,y)$ as the input, the trained generator $G^*(Z \mid Y)$ produces a dataset X_G consists of plausible failure data samples.

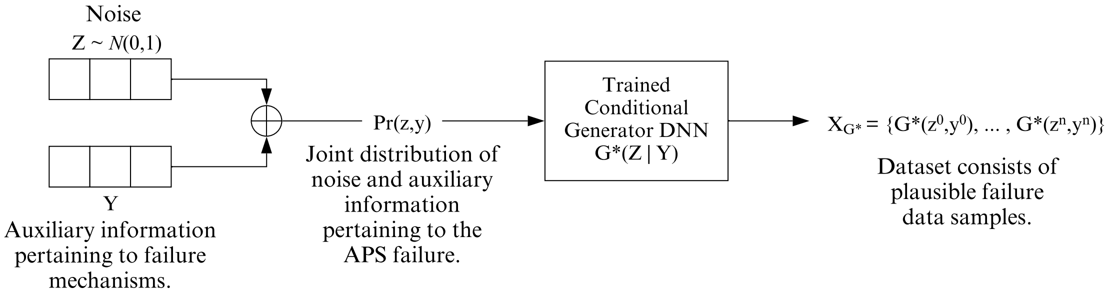


Fig. 3.11 Diagram depicting the process for producing plausible failure datasets using the estimated generative model.

The aforementioned process for producing plausible failure datasets is employed to produce multiple datasets so that each dataset consists of more plausible failure data samples than its predecessor (e.g. $X_{G_{100}}$ has 100 samples, $X_{G_{200}}$ 200 samples, $X_{G_{300}}$ has 300 samples and so on). A new plausible failure dataset is generated until the normalised Shannon entropy of the corresponding augmented train set of the last generated plausible failure dataset is equal to 1, that is, until the augmented train set is no longer imbalanced. However, this does not mean that the dataset that produces the best prognostics performance is, for example, the dataset with normalised Shannon entropy of 1. This is since having too much failure data could cause the prognostics model to overfit. Hence, even though multiple plausible failure datasets are generated until the normalised Shannon entropy is equal to 1 in this step, the best dataset for prognostics modelling is identified by observing the prognostics performance when evaluated on the test set.

Step 2: Train set augmentation

The approach used in the methodology to augment the original train set can be described using Fig. 3.12. As discussed under the auxiliary information conversion step in Phase 1, prior to splitting prognostics modelling dataset into the stratified train (i.e. original train set), validation and test sets, the auxiliary information vector is added as a new feature to the dataset (multiple auxiliary information features are added if multiple types of auxiliary information are used). Hence, the original train set already includes auxiliary information features. The plausible failure datasets produced in the previous step also includes the same auxiliary information features in addition to the original features. Hence, as shown in the figure, samples from each plausible failure dataset can be added to a copy of the original train set to produce augmented train sets. The augmented train sets are used in the next phase for prognostics modelling and the set that produced the best prognostics performance is chosen for the effective implementation of PdM.

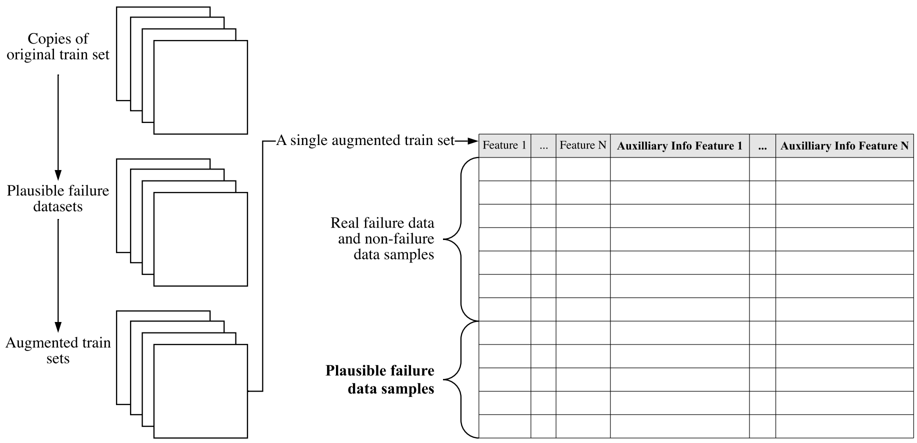


Fig. 3.12 Diagram depicting an augmented train set which includes auxiliary information features and plausible failure data samples in addition to the original features and real failure data samples. Multiple augmented train sets are produced by combining each plausible failure dataset with a copy of the original train set.

3.3.5 Phase 4: Prognostics modelling

The prognostics problem involves predicting the TTF of equipment or the probability that equipment operates without failure (i.e. the reliability of equipment) up to some future time (e.g. until the next inspection time or the end of current mission window) [1]. Even though the TTF prediction is the widely used approach, in scenarios in which the failure is catastrophic (e.g. aircraft equipment maintenance and nuclear power plant equipment maintenance), it would be more desirable to predict the reliability of equipment up to some future time [1].

The methodology for prognostics under the conditions of limited failure data availability provides two prognostics modelling approaches to address the prognostics problem: (i) binary classification-based approach for predicting the reliability of equipment up to some future time; (ii) multi-class classification-based approach for predicting the TTF of equipment. This phase involves using these approaches to estimate prognostics models for each augmented train set produced in the last phase, and then identifying the best model by evaluating the prognostics performance and model uncertainty on the test set (also see Fig. 3.13).

Step 1: Prognostics modelling approach establishment

In this step, data labelling strategies and problem formulations of the aforementioned prognostics modelling approaches are presented.

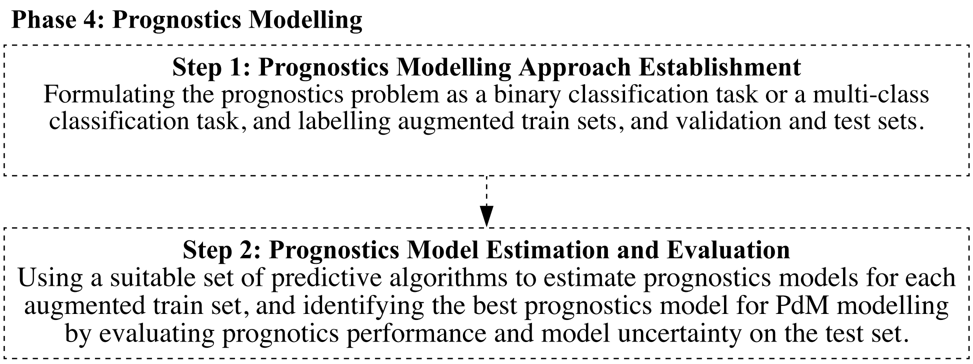


Fig. 3.13 Flowchart outlining the steps of prognostics modelling phase.

(I) Binary classification-based prognostics modelling

The objective of this approach is to predict the failure probability of equipment within a future time window (i.e. prediction horizon) using a probabilistic binary classifier and then calculate the reliability of equipment within the prediction horizon (i.e. 1 (100%) minus the failure probability).

In this approach, the data labelling strategy outlined in Fig. 3.14 is used to label the run-to-failure data samples in augmented train sets, and validation and test sets as 0 (for samples pertain to the normal condition of equipment) and 1 (for samples pertain to an impending failure). The length of prediction horizon W needs to be determined by balancing maintenance objectives (e.g. the minimum lead time required to plan and schedule maintenance tasks, deploy maintenance engineers and perform maintenance) and the estimated length of degradation period of the equipment for the failure mode that needs predicting. This is determined with the help of a domain expert.

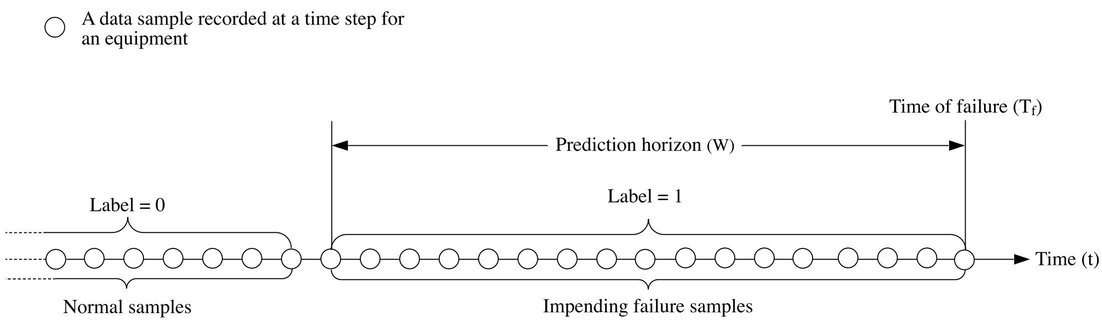


Fig. 3.14 Diagram depicting the data labelling strategy used for binary classification-based prognostics modelling.

Once the data are labelled, a probabilistic binary classifier is trained separately on each labelled augmented train sets. This produces multiple trained classifiers which are then separately applied to the labelled validation and test sets for hyperparameter tuning and testing respectively. The classifiers should now be able to identify new data samples related to the equipment condition as samples pertain to the normal condition or samples pertain to an impending failure over the next W units of time.

More specifically, the classifiers produce two probabilities indicating how much they believe a sample belongs to label 0 (i.e. label of samples pertain to the normal condition of equipment) and label 1 (i.e. label of samples pertain to an impending failure). If the optimal label (i.e. the label with highest classification probability) is 1, its classification probability is considered as the probability of failure of the equipment over the next W units of time. Thus, the reliability of equipment over the next W units of time is, $1(100\%)-$ probability of failure.

Formally, the objective of binary classification-based prognostics modelling is to first estimate a probabilistic binary classifier using historical run-to-failure data in an augmented train set $x \in X$, which is already labelled using the finite set $y \in Y,\ where\ Y = \{0, 1\}$. Then use the classifier to assign conditional probabilities to all the class labels in Y for a given sample x' related to the current condition of the equipment, such that all the probabilities sum to one. Finally, identify the optimal label y^* which is the label with highest $Pr(Y \mid X)$ for x'. That is, $y^* = \arg\max_y Pr(Y = y \mid x')$. If y^* is label 1, $Pr(Y = y \mid x')$ is the probability of failure occurring before time t (denoted $F(t)$). The reliability of the equipment until t is therefore the probability of failure occurring after t, that is, $1(100\%) - F(t)$.

(II) Multi-class classification-based prognostics modelling

Multi-class classification involves using a classification algorithm (e.g. random forest, kNN and SVM) to classify samples into three or more classes [108]. Even though it does not consider temporal patterns in condition monitoring data compared to the regression approach, it has been highly successful for TTF prediction in the prognostics domain, especially when the temporal patterns are difficult to extract [109, 110]. The objective of multi-class classification-based prognostics modelling approach is to predict the TTF of equipment by employing multi-class classification, and thus answer the following question: "what is the probability that equipment will fail in the next nL units of time, where n is the number of pre-failure time windows with a fixed length L?".

In this approach, run-to-failure data in augmented train sets, and validation and test sets are labelled using the strategy shown in Fig. 3.15. For each run-to-failure trajectory in a dataset, the time series segment pertains to an impending failure is further segmented into n number of pre-failure time windows with a fixed length L. For example, if the number of time windows is 5 and each of them has a fixed length of 1 day, the segments are 1 day before the failure, 2 days before the failure, 3 days before the failure, 4 days before the failure and 5 days before the failure. Then all the data samples are labelled with the corresponding pre-failure time window identify. For example, data samples belong to 1 day before the failure segment is labelled with 1, data samples belong to 2 days before the failure segment is labelled with 2, data samples belong to 3 days before the failure segment is labelled with 3 and so on. The normal segment of time series, which includes data samples pertain to the normal condition of the equipment is labelled with the reserved label 0. It must be noted that the labels do not need to form ordinal classes. One can also use nominal classes for labelling: data samples belong to 1 day before the failure segment is labelled with B, data samples belong to 2 days before the failure segment is labelled with A, data samples belong to 3 days before the failure segment is labelled with C and so on.

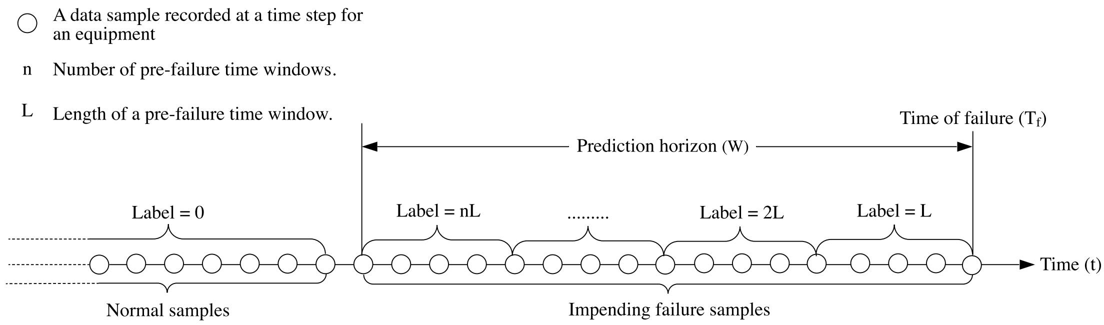


Fig. 3.15 Diagram depicting the data labelling strategy used for multi-class classification-based prognostics modelling.

A probabilistic multi-class classifier is trained separately on each labelled augmented train sets. Similar to the previous approach, this produces multiple trained classifiers which are then separately applied to the labelled validation and test sets for hyperparameter tuning and testing respectively. The classifiers should now be able to identify new data samples related to the equipment condition as samples pertain to the normal condition or samples pertain to one of the pre-failure time windows. More specifically, the classifiers produce a set of probabilities indicating how much it believes a sample belongs to the normal label (i.e. label 0) and pre-failure time window labels (i.e. labels 1 to n). If the optimal label (i.e. the label with the highest classification probability) is a pre-failure time window label, then its identity is the TTF failure of the equipment.

Formally, the objective of multi-class classification-based prognostics modelling is to first estimate a probabilistic multi-class classifier using historical run-to-failure data in an augmented train set $x \in X$, which is already labelled using the finite set $y \in Y,\ where\ Y = \{0, 1 \text{ to } n \text{ pre-failure time window labels}\}$. Then use the classifier to assign conditional probabilities to all the class labels in Y for a given sample x' related to the current condition of the equipment, such that all the probabilities sum to one. The optimal label y^* is the label with the highest $Pr(Y \mid X)$ for x'. That is, $y^* = \arg\max_y Pr(Y = y \mid x')$. If y^* is one of the pre-failure time window labels, the identity of y^* is the TTF of the equipment. Since the TTF prediction already predicts the failure, $F(t)$ is given by $Pr(y^* \mid x')$.

Step 2: Prognostics model estimation and evaluation

Once a prognostics modelling approach is established, predictive algorithms that are suitable for the approach need to be employed to estimate prognostics models. To this end, this methodology recommends the following algorithms since they have been widely used in the prognostics domain for binary classification and multi-class classification-based prognostics: random forest (RF), decision tree (DT), k-nearest neighbour (kNN), support vector machine (SVM), adaptive boosting (AdaBoost), gradient boosting machine (GBM), multilayer perceptron (MLP) and Naive Bayes (NB) (see Sec. 2.1.2 in the literature review chapter and [56]). The objective of this step is to estimate prognostics models for each augmented train set produced in Phase 3 using predictive algorithms, and then identify the best model by evaluating prognostics performance and model uncertainty.

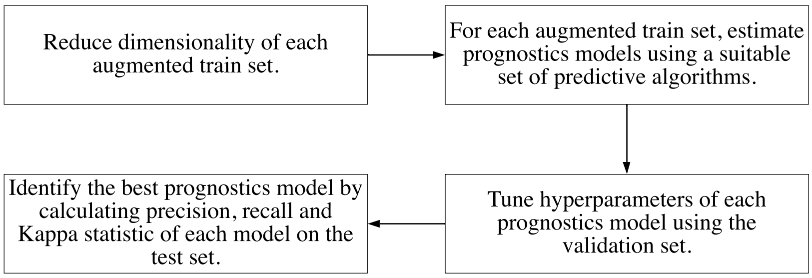


Fig. 3.16 Diagram depicting the process followed for estimating prognostics models and identifying the best prognostics model.

As shown in Fig. 3.16, for each augmented train set, a set of prognostics models are estimated and their hyperparameters are tuned using the validation set. The prognostics performance is measured on the test set using precision and recall. The standard evaluation metrics such as correlation coefficient, accuracy and error rate are not suitable for evaluating prognostics models when failure data are limited since they will be biased to the majority class (i.e. non-failure data class) regardless of the minority class (i.e. failure data class) leads to the poor performance [111]. Precision and recall, however, are not affected by the majority class hence suitable for measuring predictive performance when the data are limited [111]. Precision is the fraction of correctly predicted failures among all the predicted instances that include actual failures and false alarms. Recall is the fraction of correctly predicted failures among all the actual failures. This means higher the precision lower the number of false alarms and higher the recall lower the number of undetected failures. Formally, the precision and recall are given by Eq. 3.7 and Eq. 3.8 respectively.

$$\text{Precision} = \frac{\text{Number of predicted failures}}{\text{Number of predicted failures} + \text{Number of false alarms}} \tag{3.7}$$

$$\text{Recall} = \frac{\text{Number of predicted failures}}{\text{Number of predicted failures} + \text{Number of undetected failures}} \tag{3.8}$$

Similar to any other predictive model, the prognostics performance of a prognostics model can be affected by statistical fluke (i.e. random chance) [112]. Hence, it is important to evaluate the model uncertainty of prognostics models in addition to evaluating their prognostics performance. Kappa statistic can be used as a statistical method for identifying whether a classification-based predictive model simply guesses at random, and thus quantify model uncertainty [112]. It is always less than or equal to 1, and values of 0 or less indicate a poor model that guesses at random and conversely, 1 indicates a model that does not guess at random. A widely accepted schema for the Kappa statistic is shown in Table 3.3 [113]. In this methodology, when evaluating prognostics models, the value of the Kappa statistic is observed to identify whether the prognostics performance is affected by statistical fluke. If the prognostics performance is not affected by statistical fluke (i.e. Kappa statistic is in substantial or almost perfect agreement with the null hypothesis in Table 3.3), the precision and recall of the model are used to quantify the prognostics performance.

Table 3.3 Schema for the Cohen's Kappa statistic [113].

Kappa statistic range	Strength of agreement with H_0[a]
Less than 0	Poor (i.e. due to random chance)
0 to 0.2	Slight
0.21 to 0.4	Fair
0.41 to 0.6	Moderate
0.61 to 0.8	Substantial
0.81 to 1	Almost perfect

[a] The null hypothesis (H_0) is, *the classifier performance is not due to random chance.*

Once the prognostics performance and model uncertainty for all the prognostics models are quantified, the model that produced the highest prognostics performance (i.e. highest precision and recall) and the lowest model uncertainty (i.e. highest Kappa statistic) is selected for producing prognostics predictions under the conditions of limited failure data availability.

3.4 Analysis of methodology

The theoretical results provided in Sec. 3.3.3 proved that the conditional generative model estimation using auxiliary information pertaining to failure mechanisms can produce a generative model that is capable of replicating the real failure data distribution of the failure mode that needs predicting. More specifically, Lemma 1 proved that when the real failure data distribution and generated failure data distribution are given as the input to the CGAN implemented for the methodology, the optimal discriminator is able to identify the real failure data fraction. Then Theorem 1 proved that the CGAN can be trained (in a two-player minimax game with the optimal discriminator) to estimate the optimal generator which can replicate the real failure data distribution. Nevertheless, it is necessary to further understand the behaviour and working mechanism of conditional generative model estimation from a theoretical perspective since it is the integral part of the methodology.

The objective of this section is to conduct a theoretical experiment using sensitivity analysis to quantitatively identify how the behaviour and working mechanism of the model used for conditional generative model estimation using auxiliary information pertaining to failure mechanisms (i.e. the CGAN model given in Eq. 3.9) changes depending on its three input parameters: real failure data X, auxiliary information Y and noise Z.

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{\text{data}(X,Y)}}[\log D(x \mid y)] + \mathbb{E}_{z\sim p_{\text{noise}(Z,Y)}}[\log(1 - D(G(z \mid y)))] \qquad (3.9)$$

3.4.1 Theoretical models of system and failure

In this experiment, Matlab and Simulink are used to model and simulate a system that has developed a fault. MathWork's physics-based theoretical model of the vehicle transmission system with a gear tooth fault (i.e. worn and dented gear tooth) is used to represent the system [114]. Using a theoretical model of a system in this experiment prevents the results and analysis to be contaminated due to the high stochasticity and complexity of the real-world degradation process and uncertainty associated with the expert knowledge of failure mechanisms in real-world equipment [2, 12].

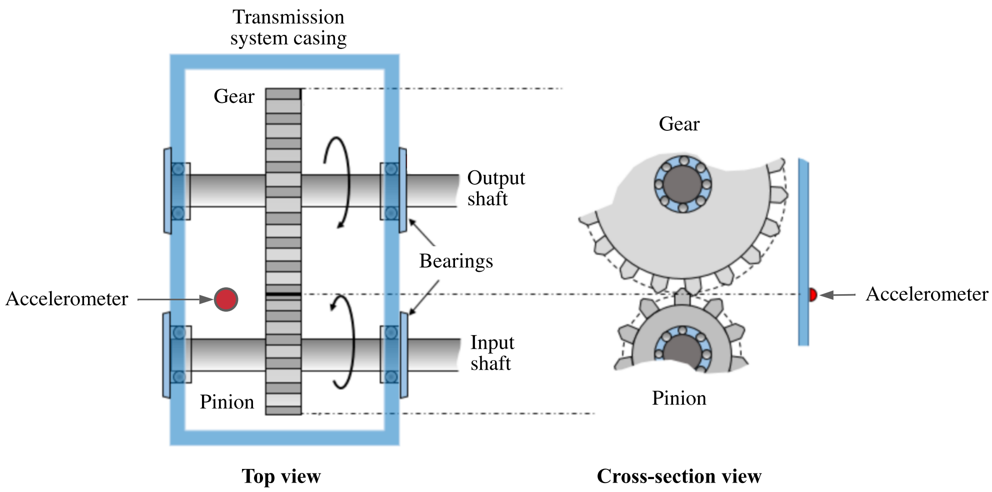


Fig. 3.17 Diagram depicting an idealised real-world system representation of the theoretical model of vehicle transmission system used in the experiment. *Adapted from [115].*

As shown in Fig. 3.17, the transmission system model consists of a 13-tooth pinion meshing with a 35-tooth gear. The pinion and the gear are connected to the input and output shafts respectively. The shafts are supported by roller bearings on the transmission system casing. An accelerometer is placed on the casing for measuring casing vibration which is caused by pinion and gear tooth meshing. Since the casing is modelled as a mass-spring-damper system, it translates the shaft angular displacement to a linear displacement on the casing. This allows the vibration generated by gear tooth meshing during shaft rotation to be measured from the casing.

The gear tooth fault is modelled by injecting a disturbance force at a fixed position (around 0 radians) in the rotation of the input shaft. The magnitude of the disturbance force, that is, the severity of gear tooth fault is controlled by a model variable (denoted F_μ). $F_\mu = 0$ implies the absence of gear tooth fault, and $F_\mu > 0$ implies the presence of gear tooth fault. When the fault is present, a 2kHz $\times$ F_μ vibration is triggered to simulate the resulting change of vibration caused by the impact of meshing between the pinion and faulty gear tooth. The rationale behind this is that when there is a faulty gear tooth in a real-world transmission system (see Fig. 3.18), the fault (i.e. wear and dent) on the tooth surface causes high-frequency oscillations over the duration of meshing between the pinion and faulty gear tooth.

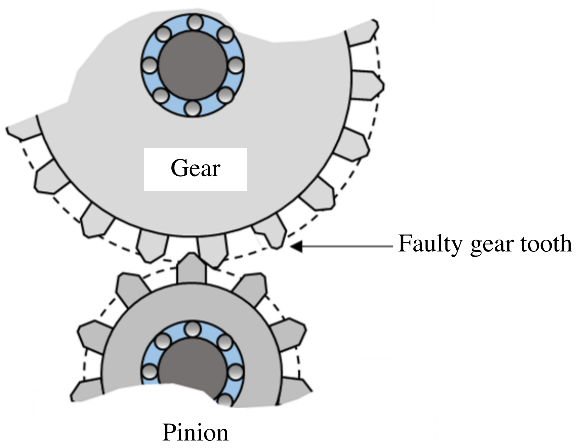


Fig. 3.18 Diagram depicting the meshing of a pinion and a gear that has a faulty tooth (i.e. worn and dented tooth). *Adapted from [115].*

The high-frequency oscillations will cause vibration signals to have a different signature to the vibration signals generated during the meshing between a pinion and a gear that do have the gear tooth fault. This allows developing a data-driven prognostics model that is capable of extracting the gear tooth fault signature from vibration data, and hence predicting the failure of transmission system due to gear tooth wear or dent.

3.4.2 Simulation of condition monitoring data and auxiliary information

The methodology uses sound engineering knowledge of failure mechanisms as auxiliary information to control and direct the failure data generation process. Hence, in order to perform sensitivity analysis on the CGAN model, it is necessary to have auxiliary information pertaining to the transmission system failure in addition to condition monitoring data.

It can be noticed from the theoretical model of vehicle transmission system discussed in the previous section, the change in magnitude of the disturbance force F_μ has a direct effect on the vibration signal measured from the casing since the system is modelled as a mass-spring-damper system. This relationship between the change in F_μ and the change in vibration signal can be used to perform a simulation that considers the impact of gear tooth fault severity (i.e. the change in F_μ) on the transmission system failure. More specifically, in order to have auxiliary information pertaining to failure mechanisms in this theoretical setting, the simulation of the transmission system and its failure can be performed to simulate condition monitoring data considering how different levels of gear tooth fault severity impact differently on the transmission system failure. This allows using the following auxiliary information for conditional generative model estimation:

Auxiliary information. Different levels of gear tooth fault severity impact differently on the transmission system failure.

Using the Simulink environment, the transmission system and gear tooth fault models are integrated into a simulation that simulates vibration data pertaining to the normal and failure conditions. 700 simulation runs are executed to simulate 700 run-to-failure data trajectories and each trajectory is a time series consists of 200 time steps. The gear tooth fault model is triggered at the half point of each time series to simulate the failure condition. This means, the first 100 steps of each time series consists of data pertaining to the normal condition, and the last 100 steps consists of data pertaining to the failure condition.

More importantly, the gear tooth fault model increases the magnitude of disturbance force F_μ every $\approx 100/3$ time steps starting from the 101^{st} time step using the ranges outlined in Table 3.4. Thus, the impact of gear tooth fault severity on the transmission system failure is increased as the time steps increases. This means, the failure condition data are simulated considering how different levels of gear tooth fault severity impact differently on the transmission system failure.

Table 3.4 Different ranges of the disturbance force magnitude F_μ used to simulate the impact of different gear tooth fault severities on the transmission system failure.

Range of F_μ	Simulated time steps	Effect generated by simulation
0.5-1.5	101 to 134	Impact of low severity gear tooth fault on failure
2.5-3.6	135 to 167	Impact of medium severity gear tooth fault on failure
4.5-5.5	168 to 200	Impact of high severity gear tooth fault on failure

Once condition monitoring data are simulated, vibration signals are transformed into a feature set that is calculated using the statistical and spectral properties of the signals (see Table 3.5). Thus, a condition monitoring dataset with 18 features that represent data pertaining to the normal and failure condition (at different levels of gear tooth fault severity) of the transmission system is obtained at the end of the simulation.

Table 3.5 Statistical and spectral features of vibration signals used in the condition monitoring dataset.

Statistical features	Spectral features
Signal mean	Peak frequency of the time synchronous average
Signal median	High frequency power
Signal root mean square (RMS)	Peak frequency of spectral kurtosis
Signal variance	
Signal peak	
Signal peak-to-peak	
Signal skewness	
Signal kurtosis	
Signal crest factor	
Signal minimum absolute difference (MAD)	
Signal range cumulative sum	
Signal correlation dimension	
Signal approximate entropy	
Signal Lyapunov exponent	

3.4.3 Scenarios and evaluation method for analysis

Since the condition monitoring dataset consists of 70000 normal data samples and 70000 failure data samples, it is not an imbalanced dataset for prognostics modelling. In order to simulate the problem of limited failure data availability, time series segments pertaining to the failure condition are randomly removed from the dataset until the extent of limited failure data availability problem is high. The resulting dataset contains 70000 normal data samples and 1000 failure data samples, meaning the failure data class only covers 1.4% of the entire dataset.

Using the Shannon entropy-based method introduced in Sec. 3.2, the severity of the problem of limited failure data availability can be measured as follows: the number of positive samples C_1 and negative samples C_2 are 1000 and 70000 respectively. The number of classes K is 2. The normalised Shannon entropy H' of the dataset is therefore 0.1, which indicates a highly imbalanced dataset. Thus, the extent of the problem of limited failure data availability for prognostics modelling is high.

Now that an imbalanced condition monitoring dataset is available, sensitivity analysis can be performed on the real failure data input parameter X, auxiliary information pertaining to the failure mechanism input parameter Y and the noise input parameter Z to gain deeper insights into the behaviour and working mechanism of the CGAN model.

The conditional generative model estimation using auxiliary information pertaining to failure mechanisms aims to estimate a generative model that is capable of replicating the real failure data distribution. In order for the CGAN to estimate this generative model, it needs achieve convergence, that is, find an equilibrium for the two-player minimax game between the conditional generator $G(Z \mid Y)$ and conditional discriminator $D(X \mid Y)$. As discussed in Sec. 3.3.3, the convergence can be evaluated by measuring the statistical similarity between generated failure data samples and real failure data samples since at the equilibrium, the generator is expected to have learnt the real failure data distribution. The objective of sensitivity analysis is to evaluate the convergence of CGAN model under varying input parameters X, Y and Z, and thus identify how its behaviour and working mechanism changes depending on the input parameters.

Table 3.6 Description of the scenarios considered in sensitivity analysis.

Scenario name	Parameters	Variation
Base scenario	X is all the real failure data, Y is auxiliary information and is given, $Z \sim N(0,1)$	
Scenario 1		$X \rightarrow X/2$ (half of real failure data)
Scenario 2		$X \rightarrow X/4$ (quarter of real failure data)
Scenario 3		$X \rightarrow$ random noise with dimension equals to X
Scenario 4		Y is given $\rightarrow Y$ is not given (not conditioned)
Scenario 5		Y is auxiliary $\rightarrow Y$ is arbitrary
Scenario 6		Z is noise $\rightarrow Z$ is normal condition data
Scenario 7		Z is noise $\rightarrow Z$ is real failure data in X
Scenario 8		Z is noise $\rightarrow Z$ is real failure data not in X

Sensitivity analysis is performed on variations of input parameters to measure how the ability of CGAN model to achieve the convergence changes. To this end, scenarios outlined in Table. 3.6 are used. The base scenario uses all the 1000 failure data samples in the condition monitoring dataset for the X input, a vector that encodes the auxiliary information stated in Sec. 3.4.2 for the Y input, and a vector sampled from a standard multivariate normal distribution for the Z input. Scenario 1, 2 and 3 apply variations to the X input parameter for measuring how the quantity and availability of real failure data samples affect the convergence. Scenario 4 and 5 apply variations to the Y input parameter to measure how the availability and meaningfulness of auxiliary information pertaining to failure mechanisms affect the convergence. Scenarios 6, 7 and 8 apply variations to the Z input parameter to measure how the meaningfulness of latent space provided to the CGAN model affects the convergence (latent space in the base scenario is not meaningful since Z is noise).

The convergence of the CGAN model in each scenario is evaluated quantitatively using the K-S test and MMD and then compared to the convergence obtained for the base scenario. In order to use the K-S test and MMD, the features that are of the highest importance for the transmission system prognostics is identified using the random forest algorithm [116]. Fig. 3.19 shows the feature importance obtained for the features in the condition monitoring dataset used in this experiment. It can be observed that the signal approximate entropy and peak frequency (i.e. peak frequency of the time synchronous average) are the features with the highest importance, hence the convergence is evaluated using these two features.

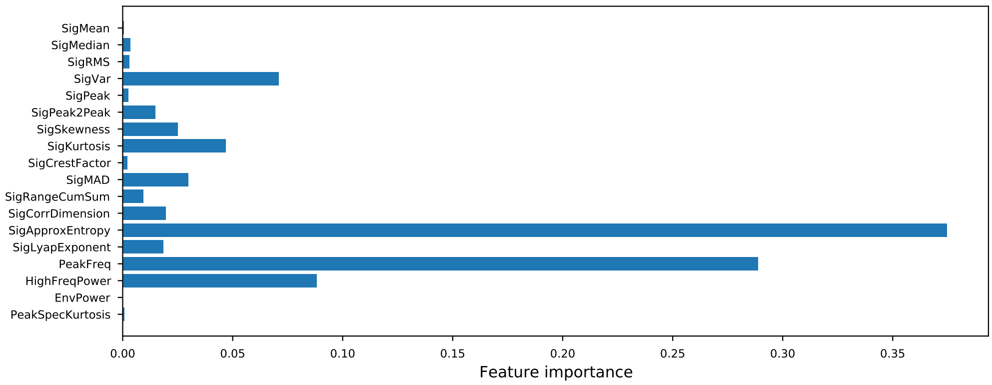


Fig. 3.19 Plot depicting the importance of features for transmission system prognostics.

3.4.4 Estimation of conditional generative model for failure data generation

Prior to executing the conditional generative model estimation for experiment scenarios, the auxiliary information needs to be converted into the vector form first. As discussed in Sec. 3.3.2, the methodology presents two approaches for the auxiliary information conversion: (i) constructing an auxiliary information vector from a continuous distribution; (ii) constructing an auxiliary information vector from class labels. The latter approach is used in this theoretical experiment to convert the auxiliary information on how the different levels of gear tooth fault severity impact differently on the transmission system failure.

For each run-to-failure trajectory in the condition monitoring dataset, the time series segment pertains to the failure condition is further segmented and labelled with the corresponding identity of the disturbance force magnitude F_μ range. For example, the time series segment simulated when F_μ is between 0.5 and 1.5 is labelled with 1, the time series segment simulated when F_μ is between 2.5 and 3.5 is labelled with 2, and the time series segment simulated when F_μ is between 4.5 and 5.5 is labelled with 3. Then this label column is added to the condition monitoring dataset as shown in Fig. 3.20. This means, the dataset now stores data that encode the auxiliary information. This column is also a vector representation of the auxiliary information: class labels that represent different ranges of F_μ which represent different levels of gear tooth fault severity in the simulation with natural numbers 1,2 and 3 are collectively is a vector of natural numbers $Y = \{y \in \mathbb{N} | 1 \leq y \leq 3\}$.

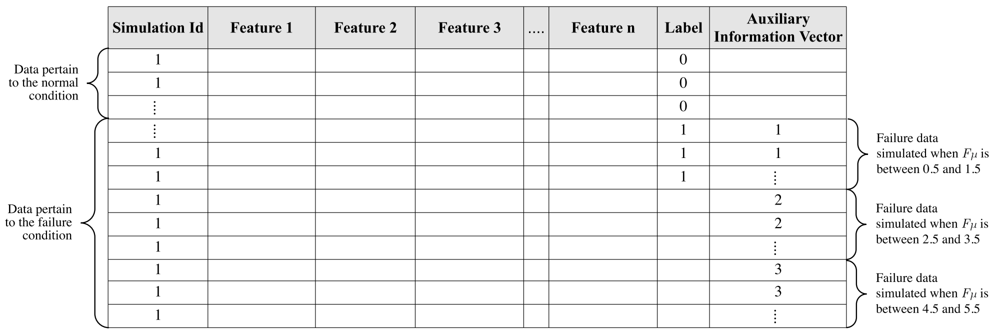


Fig. 3.20 Diagram depicting the dataset with the converted vector of auxiliary information.

The objective of executing the conditional generative model estimation in this experiment is to use the real failure data input parameter X, the vector of auxiliary information input parameter Y, and the vector of noise input parameter Z according to the scenarios outlined in Table. 3.6 as inputs to the CGAN to estimate a generative model that is capable of generating plausible transmission system failure data samples. To this end, the conditional generator $G(Z \mid Y)$ and the conditional discriminator $D(X \mid Y)$ needs to be constructed first. Then G needs to be trained in a two-player minimax game with the value function $V(G,D)$ (see Eq. 3.2) until D is unable to identify whether a given data sample x' is a real failure data sample or a generated failure data sample (i.e. until $D(x' \mid y) = 1/2$).

For Scenario 4 however, the generator and discriminator remain in the forms of $G(Z)$ and $D(X)$ respectively since the failure data generation process is not conditioned on auxiliary information in this scenario. For all the other scenarios, including the base scenario, the joint distributions $Pr(Z,Y)$ and $Pr(X,Y)$ are given as inputs to $G(Z \mid Y)$ and $D(X \mid Y)$ respectively. Since $Pr(Z,Y)$ is the input of G, it learns to produce data samples $G(z \mid y)$ which are conditioned on auxiliary information pertaining to the transmission system failure. Similarly, since $Pr(X,Y)$ is used as the input to D, it produces a probability $D(x' \mid y)$ indicating how much it believes the given data sample x' is sampled from the real failure data distribution that is conditioned on auxiliary information pertaining to the transmission system failure.

Table 3.7 Hyperparameters used to train a CGAN model for each experiment scenario (denoted S).

Hyperparameter	Value								
	Base	S1	S2	S3	S4	S5	S6	S7	S8
Optimiser	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	$2e-4$	$3e-4$	$2e-4$	$2e-4$	$1e-4$	$2e-4$	$2e-4$	$2e-4$	$2e-4$
Weight decay	$1e-6$	$3e-6$	$1e-6$	$1e-6$	$1e-6$	$1e-6$	$1e-6$	$1e-6$	$1e-6$
Exponential decay rate for first moment estimate	0.3	0.2	0.3	0.3	0.1	0.3	0.3	0.3	0.4
Exponential decay rate for second moment estimate	0.9	0.8	0.8	0.8	0.2	0.6	0.7	0.9	0.9
Number of batches	32	32	32	32	32	32	32	32	32
Number of epochs	50	45	43	60	65	45	50	50	50
Number of training iterations	1560	700	450	1870	2030	1410	1560	1560	1560

Adam=Adaptive Moment Estimation
Number of training iterations=number of batches per epoch $\times$ number of epochs, where the number of batches per epoch is, number of samples in X / number of batches

Once the generators and discriminators are constructed according to the variations of $X, Y,$ and Z model parameters defined in the experiment scenarios, a CGAN model is trained for each scenario separately using the minibatch stochastic gradient descent algorithm presented in Algorithm 1. The hyperparameters used to train a CGAN model for each scenario in this experiment are outlined in Table. 3.7.

To reiterate, the convergence evaluation of the methodology aims to statistically identify whether the following null hypothesis can be rejected: *the generated failure data distribution and real failure data distribution are statistically similar*. If this null hypothesis can be retained (i.e. p-value of the K-S test is higher than 0.05), the CGAN has achieved the convergence.

However, the convergence evaluation can be biased if the model is overfitting to the real failure data samples in X. Hence, the MMD evaluation is performed to identify whether the following null hypothesis can be rejected: *the model is not overfitting to the real failure data samples in X*. If this hypothesis can be retained (i.e. p-value of the MMD test is higher than 0.05), the CGAN is not overfitting to the real failure data samples in X. Hence, the convergence evaluation is not affected by overfitting.

The results obtained for convergence and overfitting evaluations performed for all the CGAN models trained in the previous section are summarised in Table 3.8. In the remainder of this section,

Table 3.8 Results of convergence and overfitting evaluations, that is, p-values of the K-S and MMD tests which range from 0 (no agreement with the null hypothesis) to 1 (absolute agreement with the null hypothesis) obtained for all the experiment scenarios.

Scenario name	Convergence		Overfitting	
	Signal approx. entropy	Peak frequency	Signal approx. entropy	Peak frequency
Base scenario	9.2×10^{-1}	9.3×10^{-1}	7.5×10^{-1}	7.3×10^{-1}
Scenario 1	4.4×10^{-1}	3.8×10^{-1}	7.4×10^{-1}	7.3×10^{-1}
Scenario 2	3.9×10^{-3}	1.0×10^{-3}	7.3×10^{-1}	7.3×10^{-1}
Scenario 3	3.0×10^{-7}	4.0×10^{-8}	9.3×10^{-1}	9.1×10^{-1}
Scenario 4	5.0×10^{-7}	3.0×10^{-7}	7.1×10^{-1}	7.3×10^{-1}
Scenario 5	7.0×10^{-8}	1.0×10^{-9}	7.9×10^{-1}	7.4×10^{-1}
Scenario 6	8.8×10^{-1}	9.1×10^{-1}	4.6×10^{-1}	4.3×10^{-1}
Scenario 7	9.8×10^{-1}	9.7×10^{-1}	3.0×10^{-6}	4.0×10^{-6}
Scenario 8	9.5×10^{-1}	9.7×10^{-1}	1.0×10^{-6}	2.0×10^{-6}

these results are compared to the base scenario to provide an analysis on how the input parameters X, Y and Z impact the ability of the CGAN of the methodology to achieve the convergence.

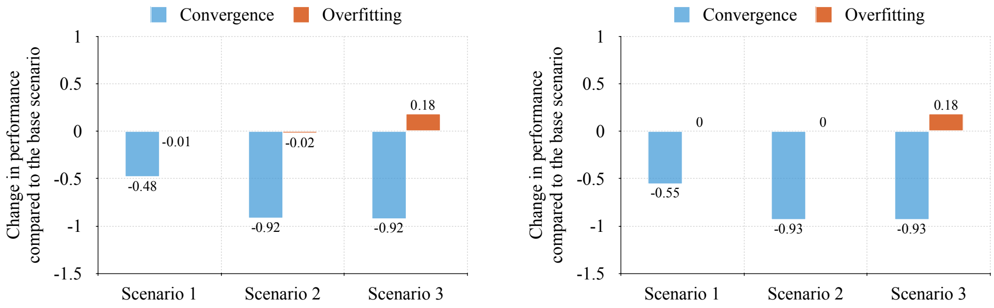
3.4.5 Sensitivity analysis on real failure data samples input parameter

The real failure data input parameter X provides condition monitoring data and/or event data pertaining to the failure mode to CGAN model. This is to allow the discriminator D in CGAN to learn to recognise real failure data samples better and thus provide useful feedback to the generator G to learn to generate fake failure data samples better. Formally, during the two-player minimax game with the value function $V(G,D)$, the X input parameter allows D to optimise the first part of its loss function shown in Eq. 3.10. Hence, the ability of D to recognise real failure data samples affects the ability of CGAN to achieve the convergence.

$$\max_{D} V(D) = \overbrace{\mathbb{E}_{x \sim p_{\text{data}(X,Y)}}[\log D(x \mid y)]}^{\text{Recognise real failure data samples better}} + \underbrace{\mathbb{E}_{z \sim p_{\text{noise}(Z,Y)}}[\log(1 - D(G(z \mid y)))]}_{\text{Recognise generated failure data samples better}} \tag{3.10}$$

As shown in Table 3.6, in order to perform sensitivity analysis on the X input parameter, variations considering the quantity and availability of real failure data samples are applied to the conditional generative model estimation. In Scenario 1 and 2, the quantity of X is halved and quartered respectively to measure the effect of the decreasing number of real failure data samples on the CGAN convergence. Then in Scenario 3, the extreme case in which there are no real failure data samples is considered.

The results obtained for Scenario 1,2 and 3 are shown in Fig. 3.21. The results show the change in convergence and overfitting performance obtained by these scenarios compared to the base scenario. It can be observed from the results obtained for Scenario 1 and 2, the amount of real failure data samples in the X input has a noticeable effect on the convergence of CGAN model. More specifically, when the



(a) Signal approximate entropy (b) Peak frequency

Fig. 3.21 Plots depicting the change in performance obtained for variations of the real failure data X input parameter compared to the base scenario.

amount of real failure data samples is reduced, the convergence performance is also reduced. Since the overfitting performance remains around 0.73 p-value for both features in Scenario 1 and 2 (see Table. 3.8), the convergence evaluations performed in the scenarios are not impacted by the CGAN model overfitting to real failure data in X. In Scenario 4, the overfitting performance is improved compared to the base scenario, and scenarios 1 and 2. The reason for this is that the use of random noise in X does not expose any real failure data to the CGAN model. Hence, the only information about the failure provided to the model is the auxiliary information pertaining to the failure in the Y input. However, similar to Scenario 2, this scenario has a low convergence performance (i.e. less than 0.05 p-value) compared to the base scenario. This also shows that the availability of a sufficient amount of real failure data samples is necessary for the CGAN to achieve the convergence. In other words, the following insight can be gained from this analysis:

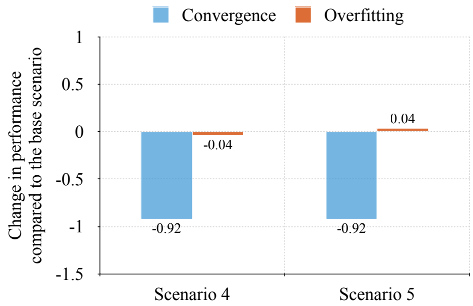
Insight. When the extent of limited failure data availability problem is approximately equal to or less than 0.1 normalised Shannon entropy, the convergence performance of the CGAN model starts to decrease.

3.4.6 Sensitivity analysis on auxiliary information input parameter

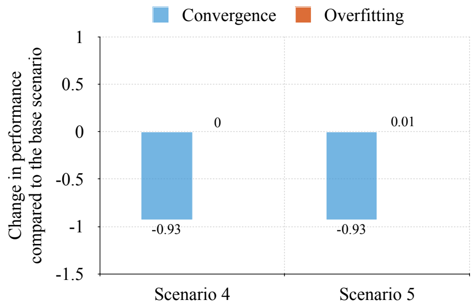
The auxiliary information pertaining to failure mechanisms input parameter Y provides sound engineering knowledge of failure mechanisms to the CGAN model to control and direct the failure generation process. More specifically, Y allows constructing the conditional generator $G(Z \mid Y)$ and discriminator $D(X \mid Y)$ of the CGAN model, and thus enables the model to condition the noise being added to the newly generated failure data samples. Formally, in the conditional generative model estimation, the loss function of D shown in Eq. 3.10 and the loss function of G shown in Eq. 3.11 are constructed by conditioning G and D on Y.

$$\min_G V(G) = \underbrace{\mathbb{E}_{z\sim p_{\text{noise}(Z,Y)}}[\log(1 - D(G(z \mid y)))]}_{\text{Optimise G to fool the discriminator the most}} \tag{3.11}$$

Scenario 4 and 5 shown in Table 3.6 apply variations considering the availability and meaningfulness of the auxiliary information pertaining to failure mechanisms to perform sensitivity analysis on the Y input parameter. In Scenario 4, the generative model estimation is executed using the unconditioned generator $G(Z)$ and unconditioned discriminator $D(X)$. The effect of the meaningfulness of Y input is measured in Scenario 5 by providing an arbitrary vector of information (i.e. information not pertaining to the failure mode that needs predicting) as the Y input the CGAN model. More specifically, a vector consists of labels with random numbers between 0 and 10 is used.



(a) Signal approximate entropy



(b) Peak frequency

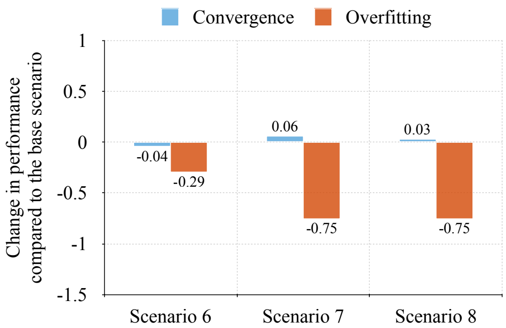
Fig. 3.22 Plots depicting the change in performance obtained for variations of the auxiliary information Y input parameter compared to the base scenario.

The convergence and overfitting results obtained by Scenario 4 and 5 are shown in Fig. 3.22. It can be observed that the variations applied to the Y input parameter do not have a noticeable effect on the overfitting performance. Moreover, since this performance in both scenarios is high (i.e. around 0.71 p-value as shown in Table 3.8), the convergence evaluations are not impacted by overfitting. From the convergence results, it can be observed that the both scenarios have achieved a poor convergence. The reason for this in Scenario 4 is that not conditioning the generator and discriminator on auxiliary information makes it difficult for the generator to learn the real failure data distribution [70]. This causes the model to not converge since an equilibrium for the two-player minimax game is difficult to be found. Since Scenario 5 has obtained a poor convergence performance, the relationship between auxiliary information and the failure mode is also important for the CGAN model convergence. The following insight follows from this analysis:

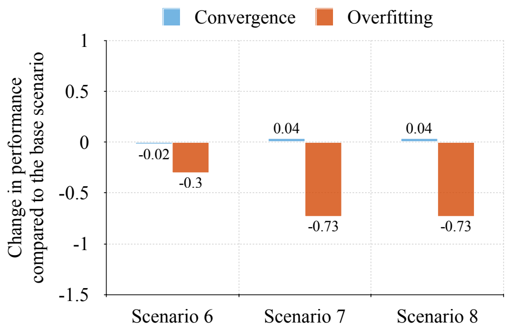
Insight. The integration of auxiliary information pertaining to failure mechanisms into the conditional generative model estimation process is crucial to estimate a generative model that is capable of generating plausible failure data samples.

3.4.7 Sensitivity analysis on noise input parameter

The noise input parameter Z provides the latent space (i.e. set of data samples) required for the CGAN to sample data that will be converted into plausible failure data samples $G(z,y)$ by the generator G in order to fool the discriminator. As shown in Table 3.6, to perform sensitivity analysis on the Z input parameter, variations considering the meaningfulness of the latent space are applied to the conditional generative model estimation. In Scenario 6, 7 and 8 the latent space is changed to normal condition data, failure data in X and failure data not in X respectively to vary its meaningfulness compared to the base scenario in which the latent space is not meaningful (since Z is noise sampled from a standard multivariate normal distribution). In other words, data samples in the latent space of Scenario 6,7 and 8 are not random compared to the random samples in the base scenario.



(a) Signal approximate entropy



(b) Peak frequency

Fig. 3.23 Plots depicting the change in performance obtained for variations of the noise Z input parameter compared to the base scenario.

The results obtained for Scenario 6,7 and 8 are shown in Fig. 3.23. It can be observed from all the scenarios, the overfitting performance has degraded compared to the base scenario, particularly in Scenario 7 and 8. This means the CGAN model is overfitting to the real failure data samples in X when the latent space is meaningful. The considerable loss in overfitting performance in Scenario 7 and 8 compared to the base scenario and Scenario 6 is since the latent space now contains data pertaining to the failure. The convergence results in all the scenarios are close to the performance obtained by the base scenario. However, the convergence evaluation in Scenario 7 and 8 are affected by overfitting, hence in these scenarios, the CGAN model is not capable of producing a generative model capable generating plausible failure data sample that maximise the prognostics and PdM performance.

Even though the overfitting performance in Scenario 6 is lower than the base scenario, the CGAN model still produced an overfitting performances of 0.46 and 0.43 *p-value* for both features as outlined in Table 3.8. The reason for this is that the theoretical models of the system and failure studied in this experiment provide a well-defined separation between the data pertaining to normal and failure conditions of the system. Thus, the normal condition data provided by the Z input to the CGAN are not contaminated with the failure condition data. In practice however, a well-defined separation between the normal condition and failure condition data of the equipment is rarely found due to the stochasticity and complexity of real-world degradation processes [2, 12]. The following insight can be gained from the analysis of Z input parameter of the CGAN model:

Insight. When the noise input parameter has random samples, the CGAN is able to achieve the convergence without overfitting to the real failure data.

3.5 Conclusion

In this chapter, the first research question of this thesis was addressed by presenting a methodology for developing prognostics models when failure data are limited. A detailed description of the methodology and theoretical results of its integral parts are provided. The theoretical analysis performed in this chapter generated insights into the behaviour and working mechanisms of the methodology. In the next chapter, case studies are provided to demonstrate how the methodology can be applied in practice.

Chapter 4

Methodology implementation

This chapter presents three case studies to show how the methodology can be implemented in practice using the three kinds of auxiliary information: effect of environmental conditions on failure mechanisms, effect of harsh use of equipment on failure mechanisms and similarity between equipment that has failed under a single failure mode. The case studies are developed using real-world industrial scenarios involving British Telecom (BT) residential broadband line prognostics and Scania heavy-truck air processing system and turbocharger prognostics (also see Table 4.1). In order to address the second research question of this thesis, this chapter provides analyses to quantify the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques, and to provide insights into the key factors influencing the effectiveness of methodology.

Table 4.1 Outline of case studies used in this thesis.

Equipment	Failure mode	Measured signals	Prognostics task	Challenge
Electrical joints and junctions in residential broadband lines	Broadband connection drops and broadband service disconnection due to corrosion and electrical shorts	Electrical signals measured from broadband lines (e.g. attenuation, power, rate, signal-to-noise ratio, loss of framing and cell-delineation)	Predict the TTF of residential broadband lines	Sparse population of broadband users causes failure data to be limited in some areas in UK
Air purge valve in heavy-truck air processing systems	Leakage of compressed air due to air purge valve surface crack	Air pressure, air consumption, air dryer generation time, air dryer heating time, outside ambient temperature, and speed, mileage and engine load of trucks	Predict the probability of heavy-truck air processing system failures	High reliability and overprotective maintenance regimes cause air processing system failures to be rare
Compressor wheel in heavy-truck turbochargers	Lack of compressed air for the internal combustion engine due to compressor wheel low-cycle fatigue	Engine load, engine revolutions per minute, and turbocharger thrust load, turbine speed, atmospheric air pressure, boost air pressure and ambient temperature	Predict the TTF of heavy-truck turbochargers	Overprotective maintenance regimes cause turbocharger failures to be rare

4.1 Case study 1: Effect of environmental conditions as auxiliary information

The information on the effect of environmental conditions on failure mechanisms refers to the information related to failure mechanisms that are initiated by the deviation from specified environmental conditions under which the equipment has to operate. In the methodology, this auxiliary information kind is used to condition the plausible failure data generation process on monotonic trends in data captured for natural and induced environmental factors.

In this section, a case study of the methodology using the effect of environmental conditions on failure mechanisms auxiliary information kind is provided. More specifically, the case study uses the effect of increasing rainfall on the corrosion and electrical shorts failure mechanisms of electronic components used in BT residential broadband lines as auxiliary information. This auxiliary information is used to generate plausible broadband line failure data to address the limited failure data availability for prognostics problem to enable the effective implementation of PdM for BT residential broadband lines operating in areas in the UK that do not have a dense population of broadband service users.

4.1.1 BT residential broadband line prognostics under the conditions of limited failure data availability

BT is the largest provider of telecommunications services in the UK. Its motivation for the effective implementation of PdM is to increase the uptime of telecommunications services for its customers whilst preventing additional costs due to over maintenance and false alarms. In the following, a description of BT residential broadband lines and their failure, and an introduction to the prognostics modelling dataset are provided. Then the problem statement for BT residential broadband line prognostics under the conditions of limited failure data availability is stated.

Description of the system

Broadband lines provide a signalling method for transporting multiple signals through coaxial cables, twisted pair and optical fibre transmission mediums. One of their main applications in the telecommunications industry is the high-speed broadband internet [117]. Broadband lines provide the broadband internet service in two forms: Asymmetric Digital Subscriber Line (ADSL) and Very High Bitrate Digital Subscriber Line (VDSL) [118]. BT ADSLs and VDSLs provide broadband service to approximately 20 million residential and 6 million business customers in the UK [119].

Although the number of customers who use broadband service delivered entirely over optical fibre is increasing, many residential broadband lines continue to be served in part by metallic paths (i.e. paired wires) [117]. As shown in Fig. 4.1, broadband lines start as large optical cables from the multi-service edge routers. Then they connect to one of the 5000 BT exchanges in the UK in which the large optical cables connect to paired wires [119]. Paired wires pass through underground and

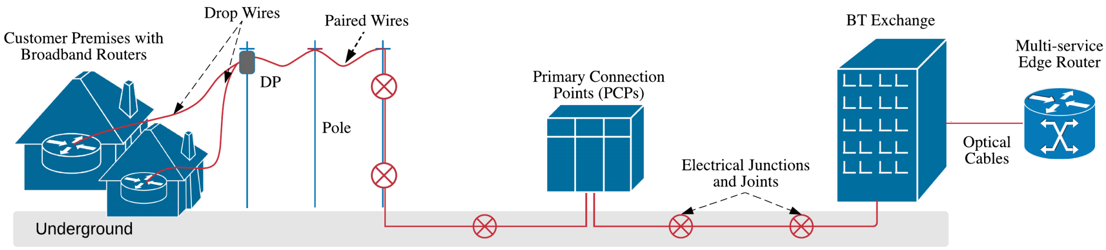


Fig. 4.1 Diagram depicting a schematic of a BT residential broadband line.

overhead electrical junctions and joints, and typically end at distribution points (DP). DPs are boxes consist of electronic circuits and located at the top of telegraph poles. They subsequently connect broadband service to one or more residential customer broadband routers using drop wires. The part of the broadband lines that uses paired wires is the system studied in this case study. More specifically, the electronic system of electrical junctions, joints and DPs.

System description. The electronic system of electrical junctions, joints and DPs in BT residential ADSLs and VDSLs.

Description of failure

Water ingress triggered by the driving rain (i.e. rain propelled by prevailing winds) is one of the main causes of failures in electronic components inside electrical junctions, joints and DPs in broadband lines [120]. It initiates the corrosion and electrical shorts failure mechanisms in these electronic components (also see Fig. 4.2) [120, 121]. These failure mechanisms degrade the electrical capability of broadband lines which consequently leads to the degradation of broadband service [120]. Historically, water ingress into electronic components inside electrical junctions, joints and DPs affected the broadband service performance, in terms of attenuation (i.e. degradation of signal), crosstalk (i.e. undesired coupling between signal paths) and noise (i.e. electromagnetic interference) [119]. For the residential broadband customer, this results in dropping connection or the complete disconnection of broadband service [121]. Thus, the failure of BT residential broadband lines is defined as follows:

Failure definition. Water ingress (failure cause) initiates corrosion and electrical shorts (failure mechanisms) in electronic components inside electrical junctions, joints and DPs in BT residential broadband lines. Corrosion and electrical shorts degrade the electrical capability of broadband lines which consequently leads to the broadband connection drops and broadband service disconnection (failure modes). This results in poor broadband service for the customer (failure). Hereinafter, this failure definition is referred to as *the broadband line failure*.

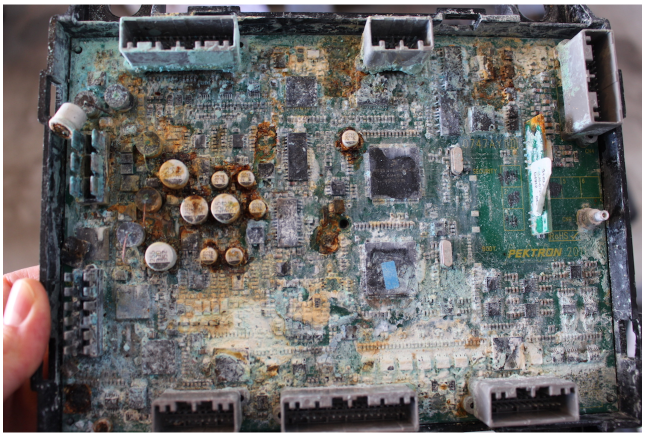


Fig. 4.2 Image depicting a broadband line electronic component that is affected by corrosion and electrical shorts due to water ingress. Water ingress causes these electronic components to become virtual galvanic cells with two metals forming the anode and cathode within the electrolyte. This combined with the effect of oxidisation leads to rapid corrosion of aluminium parts of electronic components and leads to electrical shorts [119]. *This image is provided by BT.*

Prognostics modelling dataset

The dataset used for BT residential broadband line prognostics modelling contains condition monitoring data collected from 475 ADSLs and 2150 VDSLs operating in Plymouth. The data are time series measured from June to September 2019 and have a sampling interval of 2 hours. Table 4.2 outlines the dataset features and their relevance for developing a prognostics model to predict the broadband line failure. Table 4.3 provides a descriptive analysis of these features. In Sec. 4.1.2, a detailed description of how these features are used to generate plausible failure data is provided.

In addition to the condition monitoring data, historical failure records which contain information recorded for the broadband connection drops and broadband service disconnection failure modes are used for prognostics modelling. Each failure record contains the timestamp of the broadband line failure and the issue discovered by the maintenance engineers: connection drop or service disconnection.

Limited failure data availability for BT residential broadband line prognostics

The Office of Communications (Ofcom) requires BT to provide the same quality of broadband service to customers throughout the UK including in the areas that do not have a dense population of broadband service users (e.g. Devon and Cumbria) [122]. However, this poses a challenge to the effective implementation of PdM for residential broadband lines in the UK since the low number of users leads to the limited broadband line failure data availability. The availability of a sufficient amount of failure data for each area in the UK is crucial for predicting the failure studied in this case study since the age, type and length of paired wires and the effect of weather conditions on broadband lines are different for different areas in the UK [119, 121]. Since limited failure data availability

Table 4.2 Features in the BT residential broadband line prognostics modelling dataset.

Feature	Description	Relevance for prognostics modelling
Attenuation	Degradation of signal over the distance between customer router and exchange.	Caused by degradation of electrical capability of lines due to water ingress.
Rate	Number of data packets transferred over a fixed amount of time (throughput).	Affected by broadband connection drops and broadband service disconnection.
Power	Electrical power transmitted from the exchange to the customer router.	Represents the performance of electrical joints, junctions and DPs in lines.
Signal-to-noise ratio	Relationship between signal and noise strength (electromagnetic interference).	Caused by the noise produced by degrading electronic components.
Reboots	Number of times the customer router has rebooted itself.	Represents the number of times the router was unable to connect to or receive the broadband service.
Receive and transmit blocks	Number of successfully received and transmitted data blocks.	Affected by broadband connection drops and broadband service disconnection.
Cyclic redundancy checks	Errors occur when part of a data packet is corrupt and requires retransmission.	Represents the impact of broadband connection drops on attenuation and rate.
Forward error corrections	Number of errors corrected.	Represents the stability and robustness of broadband service.
Header error corrections	Number of errors discarded.	Represents the stability and robustness of broadband service.
Loss of framing	Number of events occur when four consecutive frames do not contain a valid frame word.	Represents the stability and robustness of broadband service.
Cell-delineation errors	Number of events occurs when seven consecutive cells are invalid.	Represents the stability and robustness of broadband service.
Link retrains	Number of times exchange tried to establish a connection with customer router.	Caused by broadband service disconnections.
Error seconds	Percentage of time the customer router output errors.	Represents the severity of broadband connection drops and broadband service disconnections.
Uptime	Percentage of time broadband service is available to the customer.	Represents the uptime of broadband service provided from exchange.
Showtime	Percentage of time service is available yet the customer is not able to connect.	Represents the downtime of broadband service to the customer.

causes prognostics predictions to be associated with high uncertainty, a PdM policy that incorporates these predictions is ineffective (i.e. introduces costs due to under maintenance and over maintenance of residential broadband lines and false alarms) for BT.

Table 4.3 Descriptive analysis of BT residential broadband line dataset features.

Feature	Min		Max		Mean		Standard deviation	
	ADSL	VDSL	ADSL	VDSL	ADSL	VDSL	ADSL	VDSL
Attenuation	0	4.4	64.4	55.3	32.43	18.41	14.58	6.28
Rate	0	0	73978	80000	11331.08	50430.87	18203.91	20859.82
Power	0	0	213	145	149.34	61.81	67.29	49.97
Signal-to-noise ratio	0	1.4	38.9	35.2	6.46	7.56	3.72	3.99
Reboots	0	0	1	1	0.003	0.003	0.046	0.052
Receive and transmit blocks	0	0	18900743	0	3200886.6	0	39511010	0
Cyclic redundancy checks	0	0	100000	100000	738.69	35.93	6834.14	1505.85
Forward error corrections	0	0	100000	100000	26988.53	14316.27	39957.63	31966.98
Header error corrections	0	0	100000	100000	1689.94	36.85	10325.13	1610.65
Loss of framing	0	0	167	41	0.028	0.0093	1.31	0.57
Cell-delineation errors	0	0	863	33	0.71	0.132	9.62	1.29
Link retrains	0	0	100	22	0.077	0.032	1.14	0.55
Error seconds	0	0	8289	5892	116.63	1.6	276.38	58.88
Uptime	70	98	15766598	25744925	608042.52	767382.1	908592.68	1296576.55
Showtime	0	27	10929089	12659270	464271.53	584645.66	592692.42	869803.47

Due to technical limitations around data acquisition, the prognostics modelling dataset available for this case study is collected from Plymouth which has a dense population of residential broadband service users. Nevertheless, the following reasons make this dataset suitable for evaluating the methodology: (i) the dataset provides a good representation of residential ADSLs and VDSLs that are affected by water ingress, hence it can be used to evaluate the methodology using the effect of environmental conditions on failure mechanisms auxiliary information kind and to model the failure studied in this case study; (ii) more importantly, the limited failure data availability problem can be induced by removing failure data from the dataset as shown in Fig. 4.3. This allows evaluating the methodology when applied to areas that have a limited amount of failure data samples for residential broadband line prognostics. The following problem statement of BT residential broadband line prognostics under the conditions of limited failure data availability follows from this analysis and aligns with the two research questions of this thesis:

Problem statement. Address the problem of limited failure data availability for BT residential broadband line prognostics, and quantify the impact of methodology on BT residential broadband line prognostics under the conditions of limited failure data availability compared to the existing techniques.

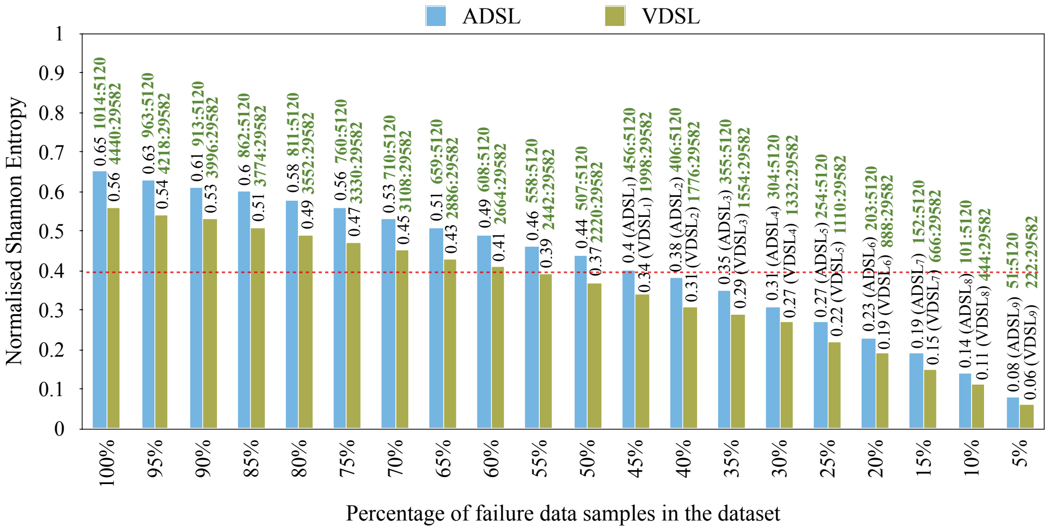


Fig. 4.3 Plot depicting how the limited failure data availability problem is induced for BT residential broadband line prognostics. The failure data are incrementally removed from the ADSL and VDSL datasets until the normalised Shannon entropy is below 0.1. The resulting datasets that have a high extent of limited failure data availability problem (denoted $ADSL_1$ to $ADSL_6$ and $VDSL_1$ to $VDSL_6$) are selected for applying and evaluating the methodology. The data imbalance ratio between positive and negative samples is depicted in green color.

4.1.2 Methodology application and results

In this section, a description of how the methodology is implemented using the effect of environmental conditions on failure mechanisms auxiliary information kind to generate plausible broadband line failure data, and thus address the problem of limited failure data availability for BT residential broadband line prognostics is provided. The $ADSL_1$ and $VDSL_1$ datasets which has a high and the expected extent of limited failure data availability problem in areas with a low density of BT residential broadband service users are used for methodology application. In the analysis performed at the end of this case study, which involves quantifying the impact of methodology compared to the existing techniques and identifying the key factors influencing the effectiveness of methodology, all the datasets denoted in Fig. 4.3 are used.

The BT residential broadband line prognostics problem satisfies the prerequisite of the methodology due to the following: the effect of increasing rainfall on the broadband line failure can be used as auxiliary information since water ingress initiates the corrosion and electrical shorts failure mechanisms of electronic components inside electrical junctions, joints and DPs, and thus degrades the electrical capability of broadband lines [120, 121]. The problem satisfies the assumption due to the following: the failure mechanisms that need predicting are corrosion and electrical shorts and they cause equipment to fail under degradation, hence the failure is not random [93].

In the remainder of this section, the description of the methodology provided in the previous chapter is used to form a discussion on how to implement the methodology using the effect of environmental conditions on failure mechanisms auxiliary information kind.

Phase 1: Auxiliary information processing

In this phase, suitable pieces of auxiliary information pertaining to the effect of increasing rainfall on the broadband line failure is identified, validated and converted into the vector form which allows constructing the conditional generator and discriminator of the CGAN.

(I) Auxiliary information identification

In order to identify potentially suitable auxiliary information, a model of the broadband line failure process initiated by the water ingress failure cause is constructed (see Fig. 4.4). Then expert knowledge pertaining to different stages in the broadband line failure process is obtained from maintenance engineers and applied research scientists at BT.

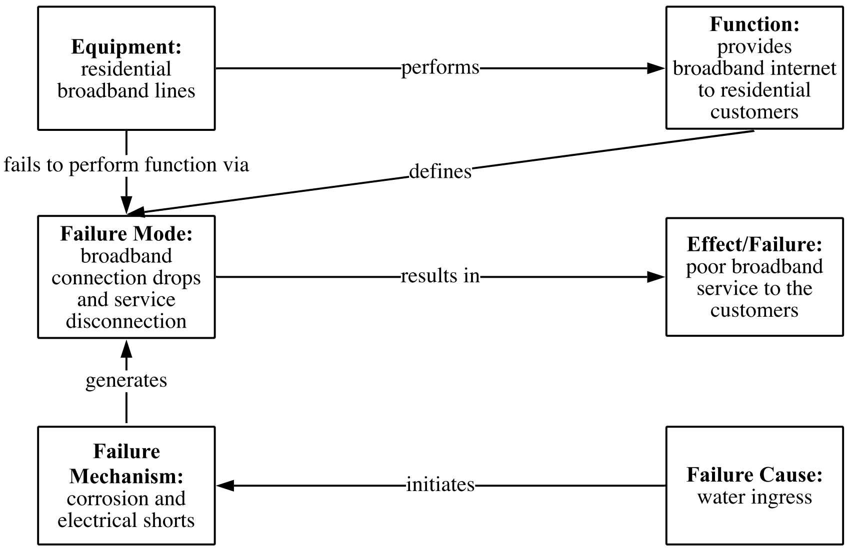


Fig. 4.4 Diagram depicting a model of the broadband line failure process which is based on the broadband line failure definition.

The following effects of rainfall on the broadband line failure are identified as potentially suitable auxiliary information from the expert knowledge: *historically, residential broadband line failures in the south-west of the UK, for example in Plymouth, is observed to be increasing as the rainfall increases; historically, an increase in residential broadband line failures is observed in the south-west of the UK when the driving rain is easterly*. The rationale for this auxiliary information is the following observations and experiences of BT maintenance engineers and applied research scientists:

- As shown in Fig. 4.5, UK rain is usually coming from the west from the Atlantic Ocean. Since Plymouth is in the south-west of the UK, it gets rain for the majority of the year. Whilst the number of broadband line failures remains low during the clear and clouds rainfall levels, when the rainfall increases to light rain and heavy rain, water ingress is observed to be taking place since there is an increase in broadband line failures.

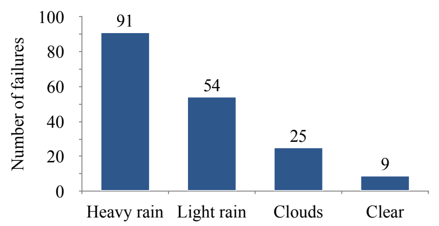
- Anecdotally, engineering practice favoured placing overhead junctions, joints and DPs on the east side of telegraph poles to get cover from the poles when there is driving rain. This is since prevailing winds and consequently, driving rain is usually westerly or south-westerly. However, when the direction of rain changes due to change in wind direction caused by phenomenon such as storms and sudden change in atmospheric pressure, water ingress is observed to be taking place since there is an increase in broadband line failures.

Fig. 4.5 Digram depicting rain and wind patterns in Plymouth which is located in the south-west of the UK. Plymouth usually gets rain from the west from the Atlantic Ocean. However, the change in wind direction due to phenomenon such as storms and sudden change in atmospheric pressure can cause the rain to come from the east.

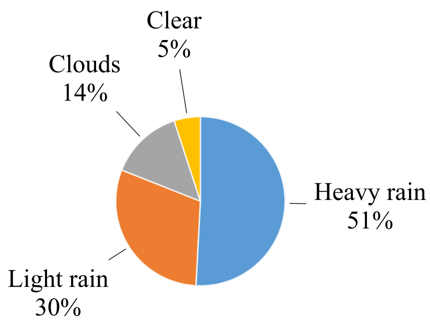
(II) Auxiliary information validation

A preliminary analysis is first performed to compare different rainfall levels to the number of broadband lines failures occurred for ADSLs and VDSLs in the prognostics modelling dataset. To this end, historical failure records provided by BT and historical weather reports obtained from the OpenWeatherMap API are used [101]. The former provides the date and time of failures and the latter provides rainfall levels and direction of the wind when it was raining.

Fig. 4.6 summarises the number of failures for different rainfall levels. It can be observed that the number of broadband line failures in Plymouth increases as the rainfall increases. Fig. 4.7 summarises the number of broadband line failures for different wind directions when it was raining. It can be observed that the south and south-west wind directions have a higher number of failures compared to the other wind directions. This is since the UK rain is usually westerly or south-westerly. However, there is no evidence to suggest that when the driving rain is easterly, the number of broadband line failures in Plymouth is increasing.

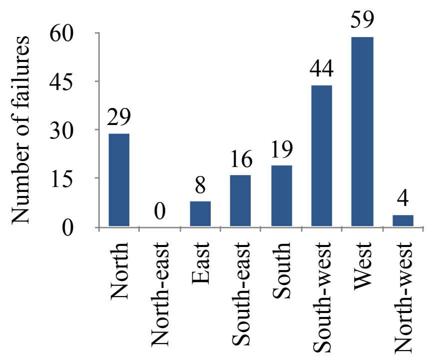


(a) Number of failures under different rainfall levels.

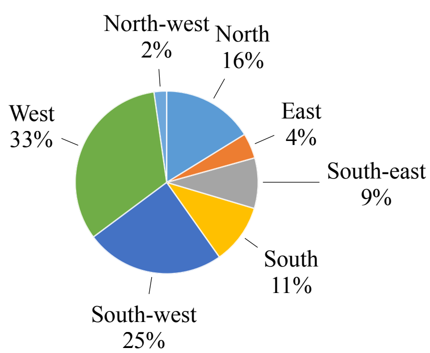


(b) Percentage of failures under different rainfall levels.

Fig. 4.6 Plots summarising the impact of rainfall levels on broadband line failures in Plymouth from June to September 2019.



(a) Number of failures under different wind directions whilst raining.



(b) Percentage of failures under different wind directions whilst raining.

Fig. 4.7 Plots summarising the impact of prevailing winds whilst raining on broadband line failures in Plymouth from June to September 2019.

In order to conduct a more robust experiment, the auxiliary information is validated using Welch's t-test. To this end, the following two null hypotheses are constructed: (i) *the failure rate of residential broadband lines is increasing when the rainfall is increasing*; (ii) *the failure rate of residential broadband lines is increasing when it is raining and when prevailing winds are easterly.* The rationale for these hypotheses are as follows: according to the auxiliary information identified from expert knowledge, the majority of failed broadband lines in the prognostics modelling dataset must have failed when the rainfall was increased and/or when the driving rain was easterly.

The objective of Welch's t-test is to identify whether the corresponding null hypothesis can be rejected. For the first hypothesis, p-values of 0.81 and 0.78 are obtained for ADSLs and VDSLs respectively. This means, there is weak evidence against the null hypothesis, thus it is retained. For the second test, p-values of 0.03 and 0.02 are obtained respectively for ADSLs and VDSLs. This means, there is strong evidence against the null hypothesis, thus it needs to be rejected. Therefore, the increase in residential broadband line failure rate when the rainfall is increasing is identified as a valid piece of auxiliary information. However, there is no strong evidence to support the increase in residential broadband line failure rate when the driving rain is easterly. Hence, the following piece of auxiliary information which is based on the effect of increasing rainfall on the corrosion and electrical shorts failure mechanisms of electronic components inside electrical junctions, joints and DPs can be used in the methodology for generating plausible broadband line failure data:

Auxiliary information. When the rainfall increases, the failure rate of BT residential broadband lines in Plymouth increases.

(III) Auxiliary information conversion

Constructing an auxiliary information vector from a continuous distribution approach is used in this case study to convert the aforementioned validated auxiliary information. As discussed in Sec. 3.3.2, this approach focuses on monotonic trends (i.e. trends that are either increasing or decreasing) in data captured for natural and induced environmental factors.

In order to convert the validated auxiliary information, it is converted into the following abstract form first: *some variable X increases*. Then the abstracted information is converted into the statistical form by representing it as some continuous variable C. The continuous variable C is a distribution between some values y_0 and y_1. This distribution is represented as a vector Y containing some values $\{y \in Y \mid y_0 < y < y_1, \text{ and y increases}\}$. Finally, Y is added as a new column to the prognostics modelling dataset as shown in Fig. 4.8. Hence, the dataset now stores data that encodes the validated auxiliary information in addition to the broadband line condition monitoring data.

Run-to-failure data of a broadband line

Line Id	Line Mode (ADSL/VDSL)	Feature 1	...	Feature n	Auxiliary Information Vector
1					0
1					0
⋮					⋮
1					0
2					1
2					1
⋮					⋮
2					1
⋮					⋮
⋮					m
⋮					m
⋮					⋮
n					m

Numerical values in the vector increase as the number of failures increases

Fig. 4.8 Diagram depicting the BT residential broadband line prognostics modelling dataset with the converted vector of auxiliary information pertaining to the effect of environmental conditions on failure mechanisms.

Phase 2: Conditional generative model estimation

In this phase, the vector of auxiliary information and the limited amount of broadband line failure data samples available are integrated to the CGAN to estimate a generative model that is capable of generating plausible broadband line failure data samples.

(I) Data preprocessing and structuring

The $ADSL_1$ and $VDSL_1$ datasets are preprocessed using the process outlined in Fig. 3.5 discussed in Sec. 3.3.3. The resulting ADSL train and test sets and VDSL train and test sets are then structured according to Fig. 3.6 to produce the below datasets. These are used in the remainder of this section to estimate and evaluate a generative model that is capable of generating plausible broadband line failure data samples.

(a) Stratified ADSL and VDSL train sets: the stratified ADSL train set and stratified VDSL train set which also include the vector of auxiliary information encoding the effect of increasing rainfall on the broadband line failure in addition to condition monitoring data.
(b) Stratified ADSL and VDSL validation sets: the stratified ADSL validation set and stratified VDSL validation set which also include the vector of auxiliary information encoding the effect of increasing rainfall on the broadband line failure.
(c) Stratified ADSL and VDSL test sets: the stratified ADSL test set and stratified VDSL test set which also include the vector of auxiliary information encoding the effect of increasing rainfall on the broadband line failure.
(d) ADSL and VDSL train failure data subsets: an ADSL subset and a VDSL subset that only contain failure data samples in ADSL train set and VDSL train set respectively.
(e) ADSL and VDSL train normal data subsets: an ADSL subset and a VDSL subset that only contain normal data samples in ADSL train set and VDSL train set respectively.

(II) Model construction and training

In order to estimate a generative model that is capable of generating plausible broadband line failure data, the methodology requires using noise Z, the real failure data samples X, and a vector of auxiliary information pertaining to failure mechanisms Y as the input to the CGAN model. In this case study, Z is noise sampled from a standard multivariate normal distribution, X is the train failure data subset, and Y is the vector of auxiliary information that encodes the effect of increasing rainfall on corrosion and electrical shorts failure mechanisms of electronic components inside electrical junctions, joints and DPs in BT residential broadband lines. Conditioning the generator and discriminator on auxiliary information pertaining to the broadband line failure allows controlling and directing the failure data generation process. Thus, the estimated generative model is able to generate new and plausible broadband line failure data samples that conform to the following condition: *newly generated broadband line failure data samples should collectively have the monotonic trend represented by the* *vector of auxiliary information which encodes the effect of increasing rainfall on the broadband line failure*.

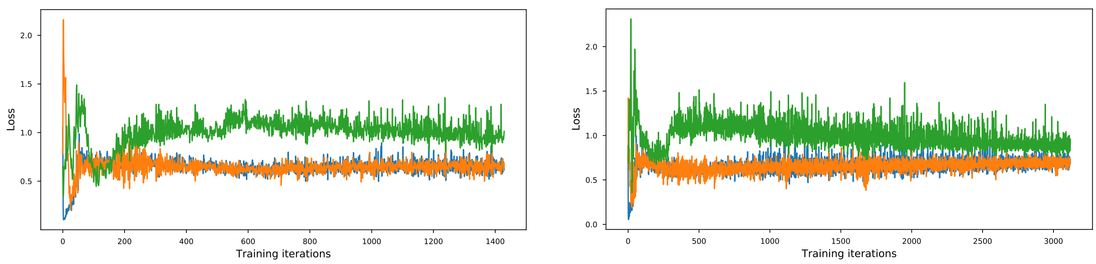
The conditional generator and discriminator of the CGAN are constructed using X, Y and Z for ADSL and VDSL datasets as per the description provided in Sec. 3.3.3. Then the ADSL and VDSL CGANs are trained using Algorithm 1. The hyperparameters used to train the CGANs are outlined in Table 4.4.

Table 4.4 Hyperparameters used to train ADSL and VDSL CGANs for estimating the two generative models that are capable of generating plausible ADSL and VDSL failure data.

Hyperparameter	Value	
	ADSL	VDSL
Optimiser	Adam	Adam
Learning rate	$2e-4$	$2e-4$
Weight decay	$1e-6$	$1e-5$
Exponential decay rate for first moment estimate	0.3	0.3
Exponential decay rate for second moment estimate	0.9	0.9
Number of batches	32	32
Number of epochs	150	72
Number of training iterations	1430	3120

(III) Model convergence evaluation

As shown in Fig. 4.9, the learning curves of ADSL and VDSL CGANs trained in the previous step provide a visual indication of whether the CGANs have achieved the convergence. The following can be observed from the learning curves: (i) discriminator loss on real and generated broadband line failure data samples is stable around 1/2 which indicates that the discriminator cannot discriminate between real and generated broadband line failure data samples; (ii) generator loss has become stable as the number of training iterations increases which indicates the generator is consistent; (iii) variance of generator and discriminator loss is modest. As discussed in Sec. 3.3.3, these properties are expected from learning curves of a CGAN that has achieved an equilibrium between its generator and discriminator, that is, the convergence of CGAN.



(a) ADSL (b) VDSL

Fig. 4.9 Generator (G) and discriminator (D) learning curves of ADSL and VDSL CGANs. D loss on real samples, D loss on generated samples, G loss are depicted with blue, orange, green curves respectively.

Since the analysis of learning curves is only a qualitative evaluation, the K-S test is performed to conduct a more robust and quantitative evaluation. Table 4.5 outlines the K-S test results obtained for a set of features in the prognostics modelling dataset. It can be observed that all the p-values produced by the K-S test are above 0.05 which indicates weak evidence against the null hypothesis. Thus, the null hypothesis cannot be rejected and therefore the CGAN has achieved the convergence. More specifically, the trained generative models, $G^*_{\text{ADSL}}(Z \mid Y)$ and $G^*_{\text{VDSL}}(Z \mid Y)$ can replicate the real broadband line failure data distribution, and hence capable of generating plausible broadband line failure data samples for ADSL and VDSL prognostics modelling.

Table 4.5 K-S test results obtained for a set of features when statistically comparing the real broadband line failure data distribution to the generated broadband line failure data distribution.

Feature	K-S test p-value	
	ADSL	VDSL
Downstream attenuation	0.84	0.88
Upstream attenuation	0.83	0.89
Downstream noise	0.79	0.71
Upstream noise	0.81	0.87
Downstream rate	0.83	0.88
Upstream rate	0.81	0.83
Downstream power	0.54	0.67
Upstream power	0.67	0.74
Error seconds	0.78	0.89

(IV) Model overfitting assessment

The MMD test outlined in Sec. 3.3.3 is employed to evaluate whether the generator is overfitting. The p-values observed for this statistical test are 0.32 (ADSL) and 0.42 (VDSL), which show that there is weak evidence against the null hypothesis. This means the null hypothesis cannot be rejected, thus the generative models, $G^*_{\text{ADSL}}(Z \mid Y)$ and $G^*_{\text{VDSL}}(Z \mid Y)$ are not overfitting to train failure data. Hence, the convergence performance of the estimated generative models is not affected by overfitting.

Phase 3: Plausible failure data generation

In this phase, the generative models which are estimated by conditioning the failure data generation process on the effect of increasing rainfall on the broadband line failure are used to generate plausible ADSL and VDSL failure data samples. These data samples are then used to augment the ADSL and VDSL train sets produced in the data preprocessing and structuring step in Phase 2, so that an increased number of broadband line failure data samples is available for prognostics modelling.

(I) Plausible failure datasets generation

As per the process outlined in Fig. 3.11, a new noise vector which is sampled from a standard multivariate normal distribution and the vector of auxiliary information which encodes the effect of increasing rainfall on the broadband line failure are used as inputs to the generative models to produce datasets consist of plausible ADSL and VDSL failure data samples. More specifically, given the joint distribution of noise and auxiliary information $Pr(Z,Y)$ as the input, the trained ADSL generator $G^*_{\mathrm{ADSL}}(Z \mid Y)$ and VDSL generator $G^*_{\mathrm{VDSL}}(Z \mid Y)$ produce datasets $X^{\mathrm{ADSL}}_{G^*}$ and $X^{\mathrm{VDSL}}_{G^*}$ consist of plausible ADSL and VDSL failure data samples.

Multiple datasets are produced so that each dataset consists of more plausible ADSL and VDSL failure data samples than its predecessor (e.g. $X^{\mathrm{ADSL}}_{G^*_{100}}$ has 100 samples, $X^{\mathrm{ADSL}}_{G^*_{200}}$ 200 samples, $X^{\mathrm{ADSL}}_{G^*_{300}}$ has 300 samples and so on). A new plausible failure dataset is generated until the normalised Shannon entropy of the corresponding augmented train set of the last generated plausible failure dataset is equal to 1, that is, until the augmented train set is no longer imbalanced.

(II) Train set augmentation

In order to produce augmented train sets, samples from each plausible broadband line failure dataset generated in the previous step are added to a copy of the original train set (also see Fig. 3.12). The augmented train sets are used in the next phase for broadband line prognostics modelling, and the set that produced the best prognostics performance is selected for quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques.

Phase 4: Prognostics modelling

In this phase, ADSL and VDSL prognostics models for BT residential broadband line failure prediction are estimated using the augmented train sets and evaluated on the original ADSL and VDSL test sets.

(I) Prognostics modelling approach establishment

The multi-class classification-based prognostics approach introduced in Sec. 3.3.5 is selected for this case study for prognostics modelling. This is due to the following: (i) prognostics modelling dataset contains time series data, hence predicting the TTF using this approach is feasible; (ii) since the failure of broadband lines is not catastrophic compared to, for example the failure of aircraft and nuclear power plant equipment, the TTF prediction is more desirable than the prediction of the reliability of equipment up to some future time.

The ADSL and VDSL datasets are labelled using the data labelling strategy of the multi-class classification-based prognostics modelling approach (see Fig. 3.15). Thus, the datasets now include a new label column indicating to which pre-failure time window (i.e. 1,2,3,4 or 5 days before failure) each data sample belongs to. The normal segment of time series, which includes data samples pertain to the normal condition of the broadband lines is labelled with the reserved label 0.

(II) Prognostics model estimation and evaluation

In the previous phase, multiple augmented train sets are produced and each of these has a different number of plausible failure data samples. These augmented train sets, and the unchanged validation and test sets are used in the process recommended by the methodology (see Fig. 3.16) to estimate prognostics models for ADSLs and VDSLs. The augmented train sets that produced the best prognostics performance are the sets with 4250 (ADSL) and 6000 plausible failure data samples (VDSL). In the following, prognostics performance obtained when trained on the augmented train sets and increase in prognostics performance provided by the methodology compared to the performance on original dataset are discussed:

(a) Prognostics performance on augmented dataset

Table 4.6 outlines the Kappa statistic values obtained for ADSL and VDSL prognostics using the algorithms discussed in Sec. 3.3.5. It can be observed that the RF-based prognostics models have obtained the highest Kappa statistic for both ADSL and VDSL. Moreover, these models produce prognostics predictions associated with minimal uncertainty since their Kappa statistic values are in substantial or almost perfect agreement with the null hypothesis of the Kappa statistic (see Table 3.3).

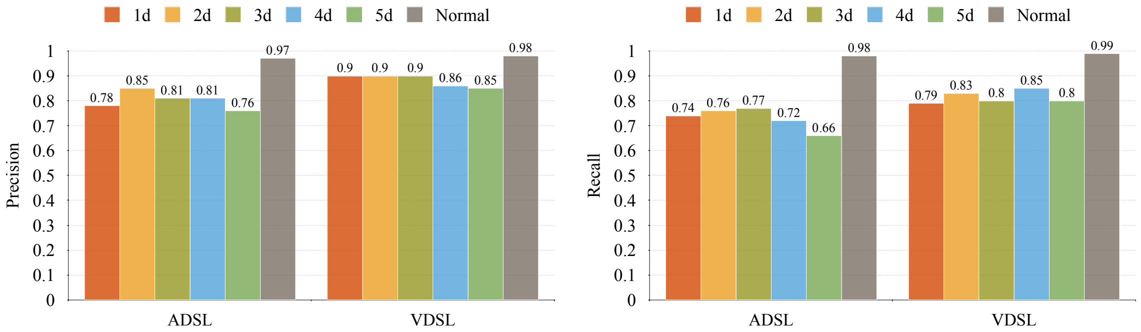
Table 4.6 Kappa statistic values obtained by ADSL and VDSL prognostics models when trained on the augmented train sets and evaluated on the test set.

Model	Kappa statistic	
	ADSL	VDSL
RF	0.79	0.86
GBM	0.76	0.7
DT	0.6	0.75
kNN	0.6	0.73
SVM	0.45	0.29
AdaBoost	0.12	0.01
MLP	0.1	0.1
NB	0.01	0.01

The prognostics performance obtained by RF-based ADSL and VDSL prognostics models is shown in Fig. 4.10 and their confusion matrixes are shown in Fig. 4.11. It can be observed that the prognostics models have obtained a high recall (i.e. low number of undetected failures) and a high precision (i.e. low number of false alarms) for BT residential broadband line prognostics under the conditions of limited failure data availability.

(b) Prognostics performance compared to original dataset

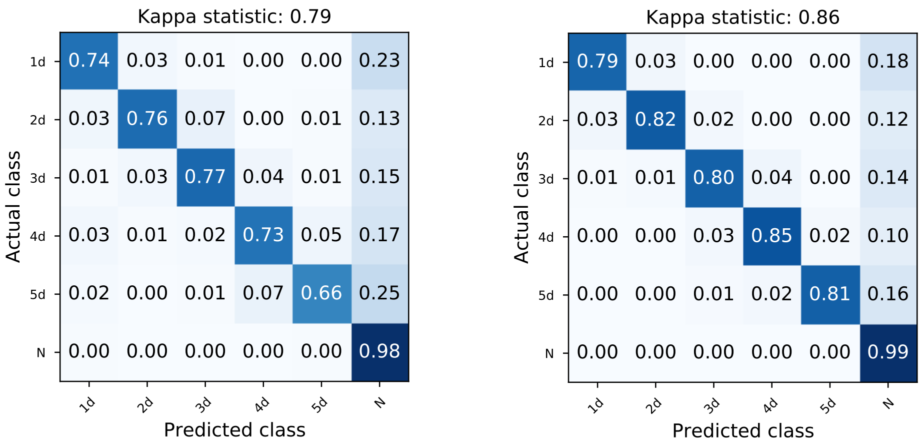
In this evaluation, the prognostics performance provided by the methodology is compared to the prognostics performance obtained when the models are trained on the following: (i) original dataset which has the limited failure data availability problem (i.e. the dataset with 30% of real failure data shown in Fig. 4.3); (ii) original dataset which do not have the limited failure data availability problem (i.e. the dataset with 100% of real failure data shown in Fig. 4.3). In this case study, the latter dataset is available for comparative evaluation since Plymouth has a dense population of residential broadband service users, and the limited failure data availability problem is induced to represent an area in the UK that do not have a dense population of broadband service users as discussed in Sec. 4.1.1.



(a) Precision (higher the precision lower the number of false alarms)

(b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.10 Prognostics performance obtained by the RF-based ADSL and VDSL prognostics models when trained on the augmented train sets and evaluated on the original test set. The failure data classes represent the pre-failure time windows: 1 day before failure (1d), 2 days before failure (2d) and so on. The normal class contains data pertain to the normal condition of BT residential broadband lines.

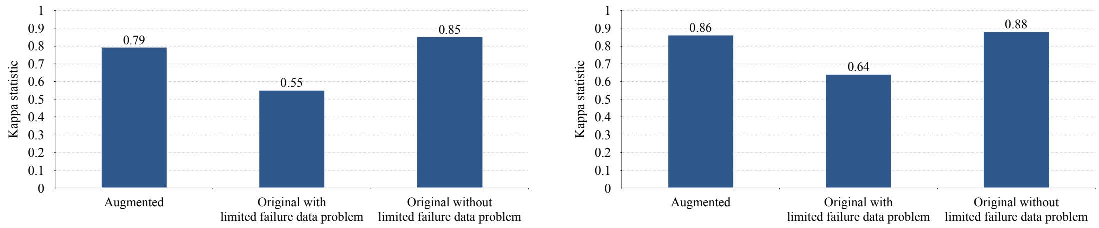


(a) ADSL

(b) VDSL

Fig. 4.11 Confusion matrixes obtained by the RF-based ADSL and VDSL prognostics models when trained on the augmented train sets and evaluated on the original test set. The normal class is denoted with N.

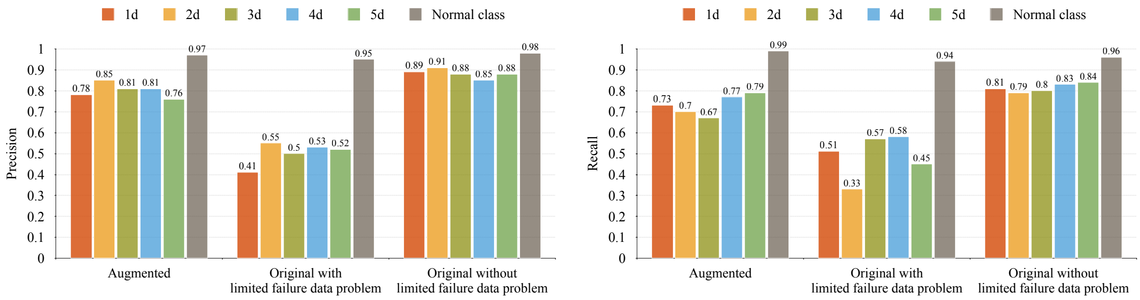
Fig. 4.12 provides a comparison of the uncertainty associated with prognostics predictions when models are trained the aforementioned datasets. It can be observed that the methodology provides a 44% reduction in uncertainty (i.e. 44% increase in Kappa statistic) for ADSL and 34% reduction in uncertainty for VDSL (i.e. 34% increase in Kappa statistic) compared to the original ADSL and VDSL datasets with the limited failure data availability problem. The reason for this is that the augmented train sets contain plausible failure data samples (in addition to the real failure data samples) which are generated by the methodology by conditioning failure data generation on the effect of increasing rainfall on the broadband line failure. When compared to the original train sets that do not have the limited failure data availability problem, the methodology achieves a close performance in reducing the uncertainty.



(a) Kappa of ADSL RF-based model (b) Kappa of VDSL RF-based model

Fig. 4.12 Comparison of uncertainty associated with prognostics predictions when the models are trained on augmented ADSL and VDSL train sets and the original ADSL and VDSL train sets. The original ADSL and VDSL test sets are used to evaluate prognostics models.

Fig. 4.13 and Fig. 4.14 show the prognostics performance provided by the methodology compared to the original train set. It can be observed that the augmented train sets outperform the original ADSL and VDSL train sets with the limited failure data availability problem by a large margin. More specifically, the increase in recall (i.e. the decrease in undetected failures) and the increase in precision (i.e. the decrease in false alarms) outlined in Table 4.7 are provided by the methodology. Fig. 4.13 and Fig. 4.14 show that compared to the original train set which do not have the limited failure data availability problem, the prognostics performance provided by the methodology is as expected is low, however, more importantly, the different in prognostics performance is marginal.

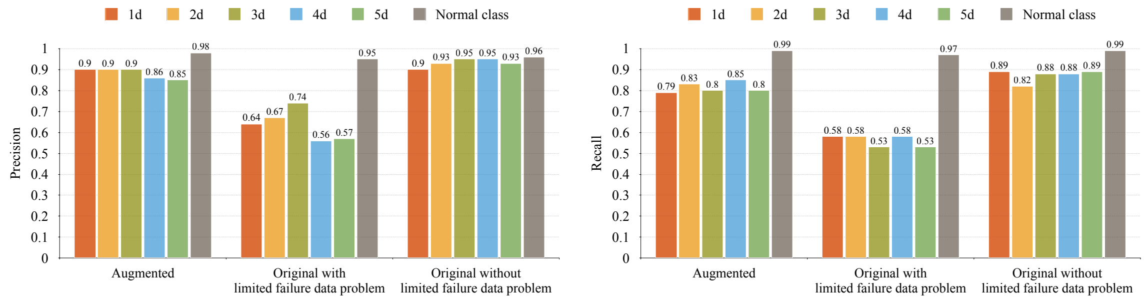


(a) Precision (higher the precision lower the number of false alarms) (b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.13 Comparison of prognostics performance obtained by the RF-based ADSL prognostics model when trained on the augmented ADSL train set and the original ADSL train set. The original ADSL test set is used to evaluate prognostics models.

Validation of prognostics performance

As stated previously, all the results discussed in this thesis are obtained using test sets that are not previously seen by the prognostics models. Nevertheless, it is worth further validating the prognostics model performance and its consistency by collecting another test set from the industrial scenario. To



(a) Precision (higher the precision lower the number of false alarms)

(b) Recall (higher the recall lower the number of undetected failures)

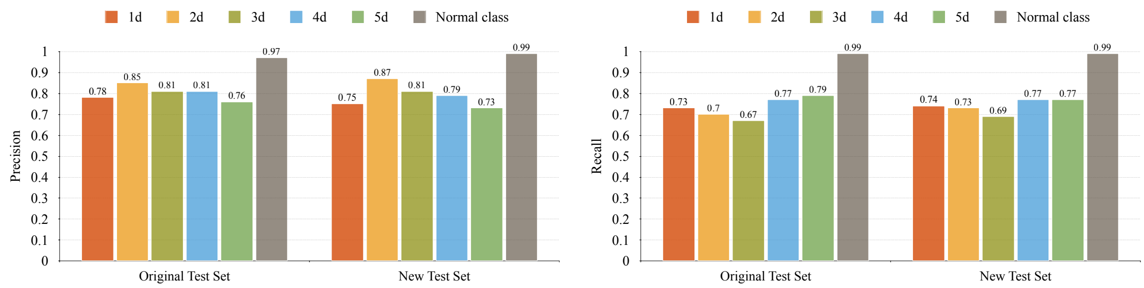
Fig. 4.14 Comparison of prognostics performance obtained by the RF-based VDSL prognostics model when trained on the augmented VDSL train set and the original VDSL train set. The original VDSL test set is used to evaluate prognostics models.

Table 4.7 Increase in prognostics performance provided by methodology compared to original dataset. Higher the precision lower the number of false alarms, and higher the recall lower the number of undetected failures.

	Increase in precision					Increase in recall				
	1d	2d	3d	4d	5d	1d	2d	3d	4d	5d
ADSL	90%	55%	62%	53%	46%	43%	112%	18%	33%	76%
VDSL	41%	34%	22%	54%	49%	36%	43%	51%	47%	51%

this end, after the models were developed and evaluated on the original test set, another test set is collected from BT ADSLs and VDSLs operating in Plymouth.

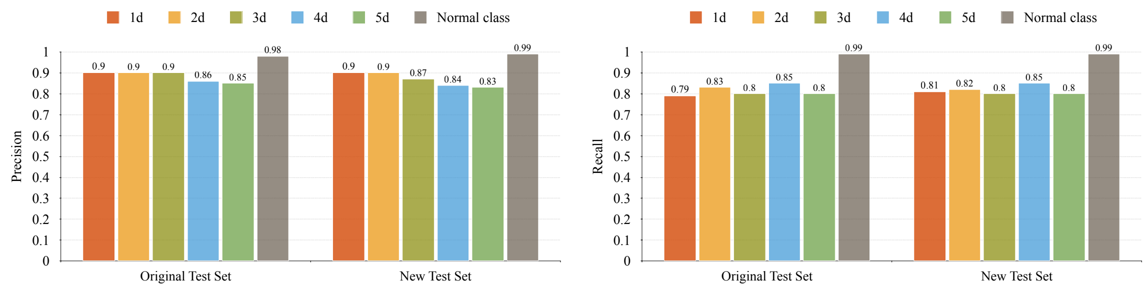
Fig. 4.15 and Fig. 4.16 provide a comparison of prognostics performance obtained for the original and newly collected test sets. It can be observed that the prognostics performance remains consistent across both test sets, which demonstrates the validity of the prognostics model and the consistency of prognostics performance.



(a) Precision

(b) Recall

Fig. 4.15 Comparison of prognostics performance obtained by the RF-based ADSL prognostics model when evaluated on the original test set and new test set.



(a) Precision (b) Recall

Fig. 4.16 Comparison of prognostics performance obtained by the RF-based VDSL prognostics model when evaluated on the original test set and new test set.

Summary of methodology application using the effect of environmental conditions

The problem statement of this case study stated in Sec. 4.1.1, which aligns with the two research questions of this thesis involves addressing the problem of limited failure data availability for prognostics and quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques.

In this case study, the methodology which addresses the first research question of this thesis is applied to generate plausible failure data samples using the effect of environmental conditions on failure mechanisms auxiliary information kind, and thus address the problem of limited failure data availability for prognostics in practice. It is shown that the methodology is capable of addressing this problem by producing a prognostics model that is capable of estimating accurate prognostics predictions with minimal uncertainty. Thus, this case study has shown that the application of the methodology using the effect of environmental conditions on failure mechanisms auxiliary information kind addresses the problem of limited failure data availability for prognostics in practice.

4.1.3 Analysis of prognostics impact and key factors influencing effectiveness

As discussed in the literature review chapter, the existing techniques used to address the problem of limited failure data availability for prognostics include random oversampling (ROS), synthetic minority oversampling (SMOTE), adaptive synthesis (ADASYN), random undersampling (RUS) and nearest neighbour undersampling (NearMiss). In this section, insights are produced using the BT residential broadband line prognostics case study to achieve the following: to quantify the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques when it is applied using the effect of environmental conditions on failure mechanisms auxiliary information kind; to identify the key factors influencing the effectiveness of methodology when it is applied using the effect of environmental conditions on failure mechanisms auxiliary information kind.

Impact on prognostics under the conditions of limited failure data availability

In the following, the prognostics performance provided by the methodology when applied using the effect of environmental conditions on failure mechanisms auxiliary information kind is compared to the prognostics performance provided by the existing techniques. The same prognostics algorithm (i.e. in this case, the RF algorithm as it produced the best prognostics performance), train and test sets are used across all the techniques including the methodology to allow a comparative evaluation.

Fig. 4.17 provides a comparison of the uncertainty associated with prognostics predictions produced by the methodology and existing techniques. It can be observed that the methodology provides a significant reduction in uncertainty compared to the existing techniques. Moreover, whilst the methodology provides a significant reduction in uncertainty compared to the original dataset which has the limited failure data availability problem, the reduction in uncertainty provided by existing techniques is marginal compared to the original dataset. Hence, the methodology outperforms existing techniques by a large margin in reducing the uncertainty associated with prognostics predictions. More specifically, the uncertainty reductions outlined in Table 4.8 are provided by the methodology compared to the existing techniques.

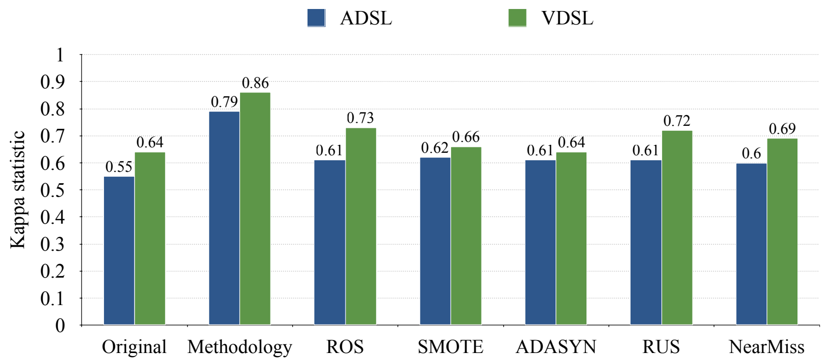
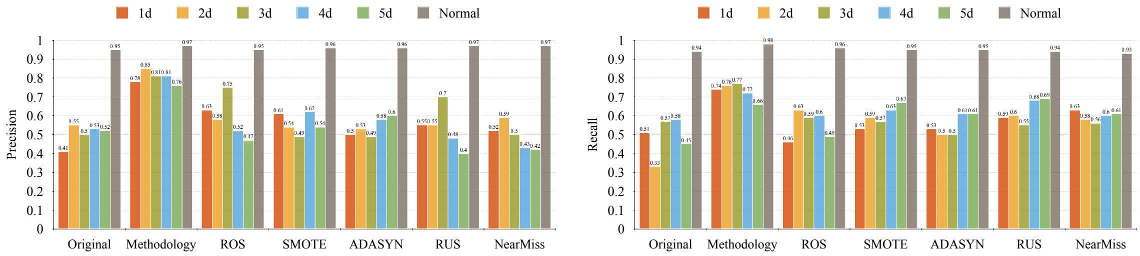


Fig. 4.17 Comparison of uncertainty associated with prognostics predictions produced by the methodology and existing techniques. The uncertainty produced on the original dataset is denoted as Original.

Table 4.8 Reduction in uncertainty (i.e. increase in Kappa statistic) provided by the methodology compared to the existing techniques.

	Reduction in uncertainty				
	ROS	SMOTE	ADASYN	RUS	NearMiss
ADSL	30%	27%	30%	30%	32%
VDSL	18%	30%	34%	19%	25%

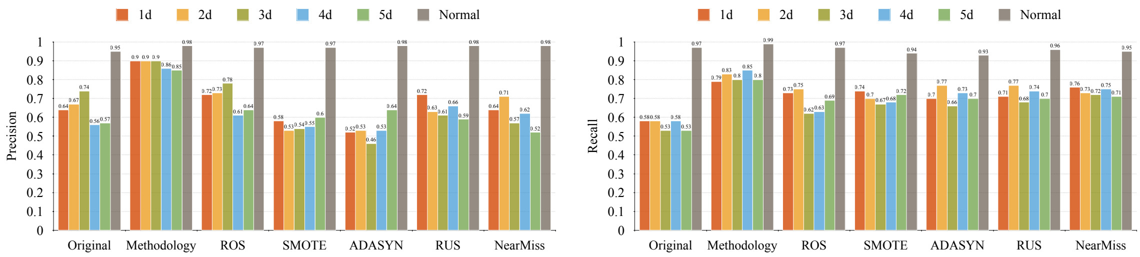
In Fig. 4.18 and Fig. 4.19, the prognostics performance provided by the methodology is compared to the existing techniques. It can be observed that the methodology allows predicting a higher number of failures (i.e. has a higher recall) and causes a lower number of false alarms (i.e. has a higher precision) compared to all the existing techniques. More specifically, the increase in prognostics performance outlined in Table 4.9 and Table 4.10 is provided by the methodology compared to the existing techniques.



(a) Precision (higher the precision lower the number of false alarms)

(b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.18 Comparison of ADSL prognostics performance produced by the methodology and existing techniques.



(a) Precision (higher the precision lower the number of false alarms)

(b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.19 Comparison of VDSL prognostics performance produced by the methodology and existing techniques.

Table 4.9 Decrease in the number of false alarms (i.e. increase in precision) and decrease in the number of undetected failures (i.e. increase in recall) provided by the methodology for ADSL prognostics compared to the existing techniques.

	Increase in precision					Increase in recall				
	1d	2d	3d	4d	5d	1d	2d	3d	4d	5d
ROS	24%	47%	8%	56%	62%	61%	21%	31%	20%	35%
SMOTE	28%	57%	65%	31%	41%	40%	29%	35%	14%	-1%
ADASYN	56%	60%	65%	40%	27%	40%	52%	54%	18%	8%
RUS	42%	55%	16%	69%	90%	25%	27%	40%	6%	-4%
NearMiss	50%	44%	62%	88%	81%	17%	31%	38%	20%	8%

The reason for the significant increase in prognostics performance is since the methodology is capable of addressing the problem of limited failure data availability by generating new and plausible failure data samples. In contrast, as discussed in the literature review chapter, the existing techniques either duplicate existing failure data or randomly generate data (oversampling techniques) or do not increase the number of failure data samples available for prognostics modelling (undersampling

Table 4.10 Decrease in the number of false alarms (i.e. increase in precision) and decrease in the number of undetected failures (i.e. increase in recall) provided by the methodology for VDSL prognostics compared to the existing techniques.

	Increase in precision					Increase in recall				
	1d	2d	3d	4d	5d	1d	2d	3d	4d	5d
ROS	25%	23%	15%	41%	33%	8%	11%	29%	35%	16%
SMOTE	55%	70%	67%	56%	42%	7%	19%	19%	25%	11%
ADASYN	73%	70%	96%	62%	33%	13%	8%	21%	16%	14%
RUS	25%	43%	48%	30%	44%	11%	8%	18%	15%	14%
NearMiss	41%	27%	58%	39%	63%	4%	14%	11%	13%	13%

techniques). Hence, they do not address the fundamental problem of limited failure data availability for prognostics. The following insight can be gained from this analysis:

Insight. When the methodology is applied using the effect of environmental conditions on failure mechanisms auxiliary information kind for prognostics under the conditions of limited failure data availability, it produces effective prognostics predictions and outperforms existing techniques used in the literature by a large margin.

Key factors influencing the effectiveness of methodology

The analysis performed from a theoretical perspective in Chapter 3 showed that the availability of a sufficient amount of real failure data samples and the availability of auxiliary information pertaining to failure mechanisms are the key factors influencing the convergence performance of the CGAN developed for estimating the plausible failure data generation model. In this section, an analysis is performed from a practical perspective to identify the influence of these factors on the effectiveness of methodology in practice (i.e. the influence on the prognostics performance) when it is applied using the effect of environmental conditions on failure mechanisms auxiliary information kind.

The analysis uses the BT residential broadband line prognostics case study and scenarios under consideration are outlined in Table 4.11. The base scenario contains real failure data X, auxiliary information Y and noise Z used for impact analysis in the previous section. Scenarios 1 to 5 apply variations to the availability of real failure data so that the extent of the limited broadband line failure data availability problem is increased as shown in Fig. 4.3. Scenario 6 and 7 apply variations to the availability of auxiliary information pertaining to the effect of environmental conditions on failure mechanisms. The noise Z remains fixed since it is always available in practice. The scenarios are applied separately for ADSL and VDSL prognostics modelling.

(I) Effect of real failure data availability on prognostics performance

The effect on prognostics performance when scenarios 1 to 5 are applied to ADSL and VDSL prognostics modelling is shown in Fig. 4.20 and Fig. 4.21. It can be observed that the amount of real failure data samples has a noticeable effect on the prognostics performance produced by the methodology when the normalised Shannon entropy is approximately equal to or less than 0.1 (see

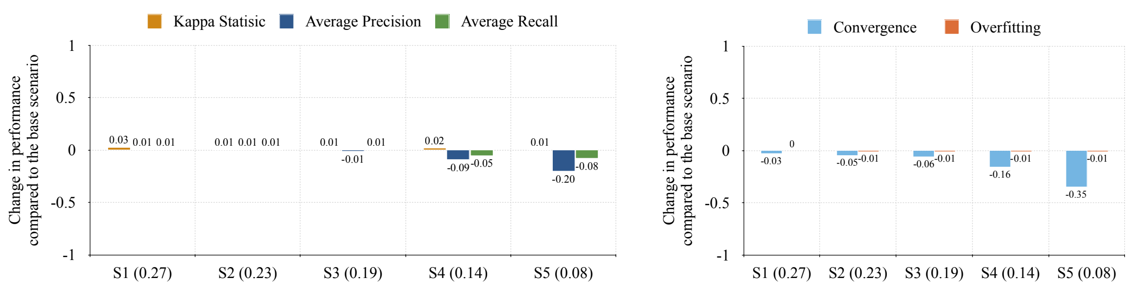
Table 4.11 Description of scenarios considered in the analysis.

Scenario name	Parameters	Variation
Base scenario	X is failure data in $ADSL_1$ or $VDSL_1$, Y is the vector encoding the effect of increasing rainfall on corrosion and electrical shorts failure mechanisms and is given, $Z \sim N(0,1)$ and is fixed	
Scenario 1		$X \rightarrow$ failure data in $ADSL_2$ or $VDSL_2$
Scenario 2		$X \rightarrow$ failure data in $ADSL_3$ or $VDSL_3$
Scenario 3		$X \rightarrow$ failure data in $ADSL_4$ or $VDSL_4$
Scenario 4		$X \rightarrow$ failure data in $ADSL_5$ or $VDSL_5$
Scenario 5		$X \rightarrow$ failure data in $ADSL_6$ or $VDSL_6$
Scenario 6		Y is given $\rightarrow Y$ is not given (not conditioned)
Scenario 7		Y is auxiliary $\rightarrow Y$ is arbitrary

Scenario 4 and 5). The reason for this can be observed from the convergence performance which shows that the convergence of the CGAN model degrades when the extent of limited failure data availability is approximately equal to or less than 0.1 normalised Shannon entropy, which is also observed in the theoretical analysis performed in Sec. 3.4.5.

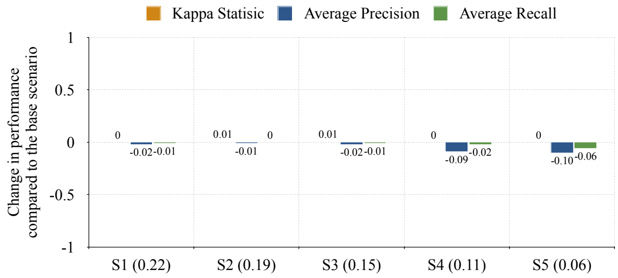
Fig. 4.22 and Fig. 4.23 show the effect of real failure data availability on the prognostics performance produced by the methodology compared to the prognostics performance produced by the existing techniques. It can be observed that whilst the methodology outperforms existing techniques in achieving a low uncertainty for prognostics predictions in all the scenarios, its precision and recall starts to degrade when the extent of limited failure data availability is approximately equal to or less than 0.1 normalised Shannon entropy. The following insight can be gained from this analysis:

Insight. In practice, when the availability of real failure data samples is approximately equal to or less than 0.1 normalised Shannon entropy, the prognostics performance of the methodology starts to become less effective and eventually becomes comparable to the existing techniques as the availability of real failure data is reduced.

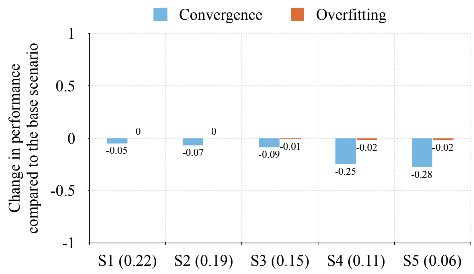


(a) Prognostics performance (b) Convergence performance

Fig. 4.20 ADSL prognostics and convergence performances produced by the methodology for scenarios (denoted S) 1 to 5 compared to the base scenario. The x-axis shows the scenario and corresponding extent of limited failure data availability (i.e. the normalised Shannon entropy).

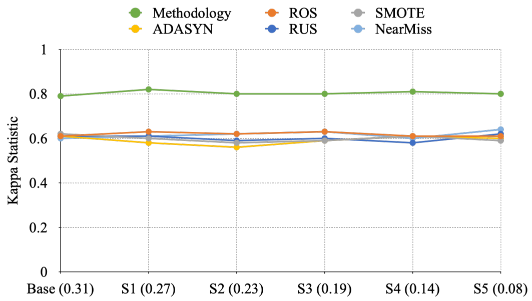


(a) Prognostics performance

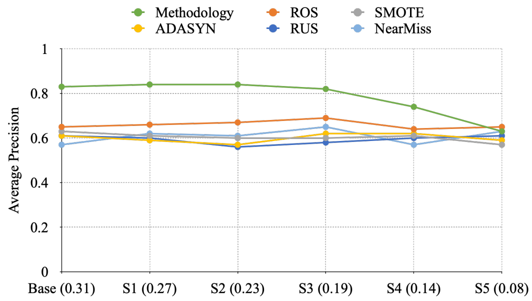


(b) Convergence performance

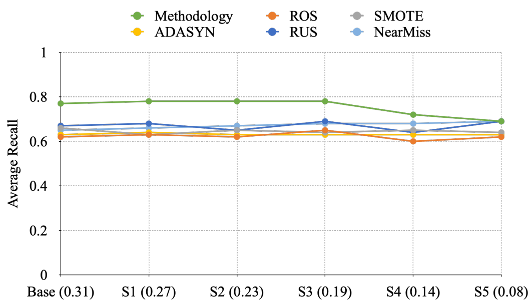
Fig. 4.21 VDSL prognostics and convergence performances produced by the methodology for scenarios 1 to 5 compared to the base scenario.



(a) Uncertainty



(b) Precision

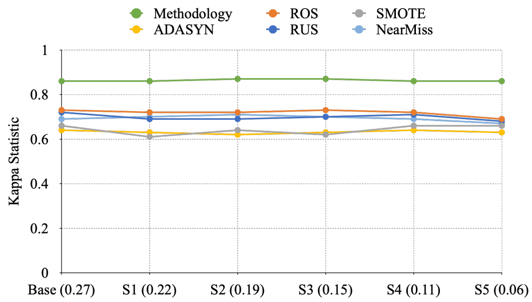


(c) Recall

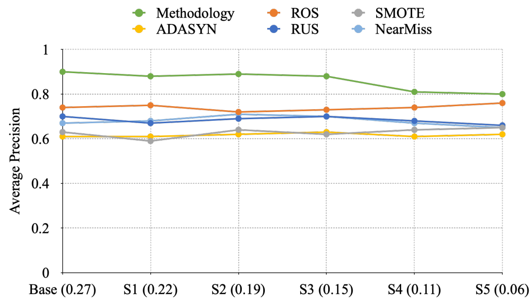
Fig. 4.22 Comparison of ADSL prognostics performance for scenarios 1 to 5 between the methodology and existing techniques. Higher the precision lower the number of false alarms, and higher the recall lower the number of undetected failures.

(II) Effect of auxiliary information availability on prognostics performance

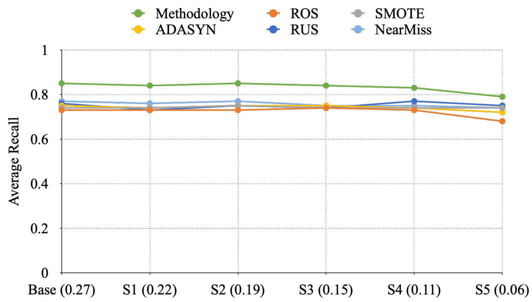
In the theoretical analysis, it is shown that the availability of auxiliary information pertaining to failure mechanisms is crucial for the methodology to estimate a generative model that is capable of generating plausible failure data samples (see Sec. 3.4.6). This can also be observed from a practical perspective as shown in Fig. 4.24 and Fig. 4.25. It can be observed that the prognostics performance produced by the methodology significantly reduces when auxiliary information pertaining to failure mechanisms is not integrated for controlling and directing the plausible failure data generation process.



(a) Uncertainty



(b) Precision

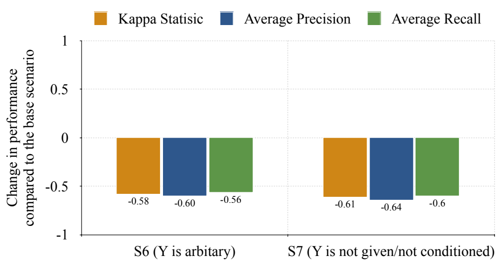


(c) Recall

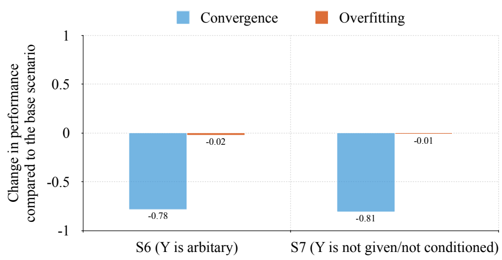
Fig. 4.23 Comparison of VDSL prognostics performance for scenarios 1 to 5 between the methodology and existing techniques. Higher the precision lower the number of false alarms, and higher the recall lower the number of undetected failures.

The effect of auxiliary information availability on prognostics performance produced by the methodology compared to the existing techniques can be observed from Fig. 4.26 and Fig. 4.27. It can be observed that the prognostics performance produced by the methodology for Scenario 6 and 7 is significantly low compared to the base scenario and all the existing techniques. The following insight can be gained from this analysis:

Insight. The integration of auxiliary information pertaining to failure mechanisms which allows controlling and directing the plausible failure data generation process is crucial to the methodology for producing effective prognostics under the conditions of limited failure data availability.

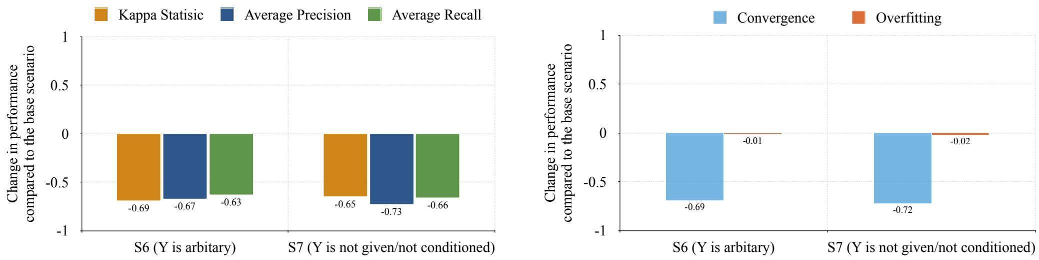


(a) Prognostics performance



(b) Convergence performance

Fig. 4.24 ADSL prognostics and convergence performances produced by the methodology for scenarios 6 and 7 compared to the base scenario.



(a) Prognostics performance (b) Convergence performance

Fig. 4.25 VDSL prognostics and convergence performances produced by the methodology for scenarios 6 and 7 compared to the base scenario.

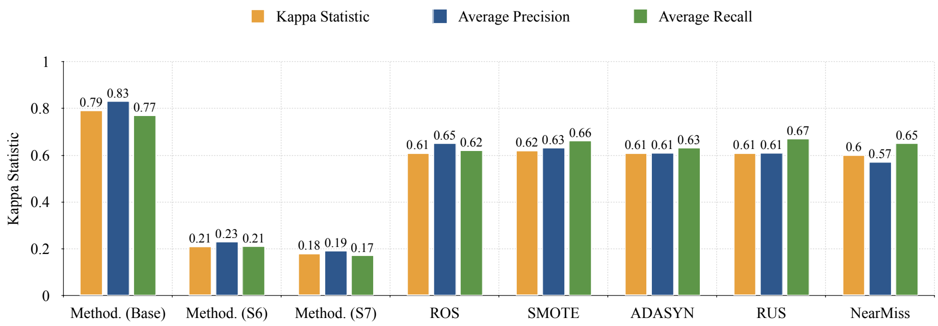


Fig. 4.26 ADSL prognostics performance produced by the methodology for scenarios 6 and 7 compared to the base scenario and existing techniques.

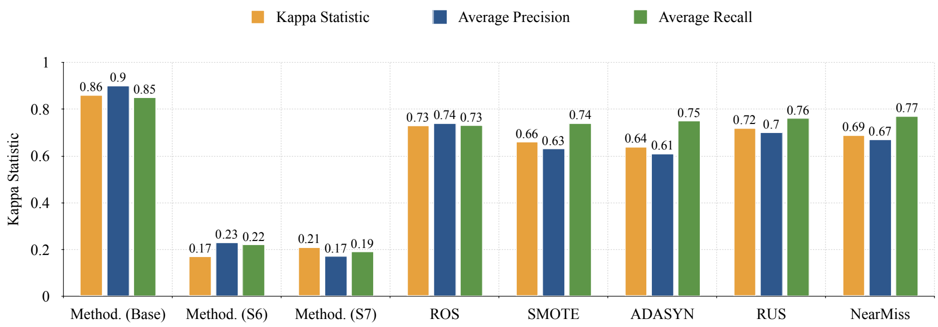


Fig. 4.27 VDSL prognostics performance produced by the methodology for scenarios 6 and 7 compared to the base scenario and existing techniques.

4.2 Case study 2: Similarity between failed equipment as auxiliary information

A relationship between the similarity between equipment that has failed and degradation patterns can often be established since there is an effect of equipment characteristics (e.g. equipment age, type, model, function and calculated metrics such as health indicators and natural groupings) on the failure process [71, 72]. The case study discussed in this section shows how to implement the methodology using the similarity between equipment that has failed under a single failure mode auxiliary information kind.

More specifically, this case study uses the similarity between Scania heavy-trucks that have failed due to the air processing system (APS) failure as auxiliary information. This auxiliary information is used to generate plausible APS failure data to address the limited failure data availability problem, which is caused by the rarity of the occurrence of APS failure to enable the effective implementation of PdM for Scania heavy-trucks.

4.2.1 Scania heavy-truck air processing system prognostics under the conditions of limited failure data availability

Scania has become one of the world's leading manufacturers of heavy-trucks and its motivation for the effective implementation of PdM is to increase the uptime of heavy-trucks whilst minimising maintenance costs for the organisation and customers by postponing maintenance until only when necessary (also see Fig. 4.28). In the following, a description of Scania heavy-truck APS and its failure, and an introduction to the prognostics modelling dataset are provided. Then the problem statement for Scania heavy-truck APS prognostics under the conditions of limited failure data availability is stated.

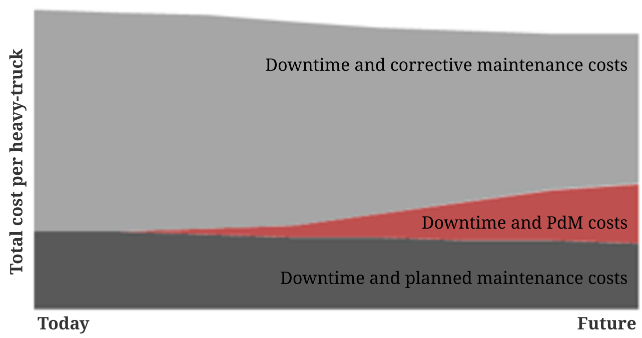
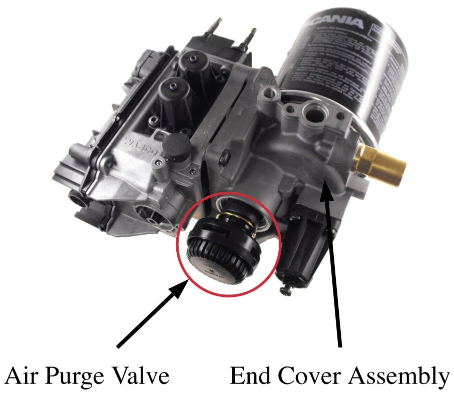


Fig. 4.28 PdM vision of Scania for heavy-truck maintenance. Scania aims to reduce downtime and corrective and planned maintenance costs by effectively implementing PdM. Thus, reduce the loss of production due to downtime and Total Cost of Ownership (TCO) for the customer and after-sales maintenance contract costs for Scania. *Adapted from material provided by Scania CV.*

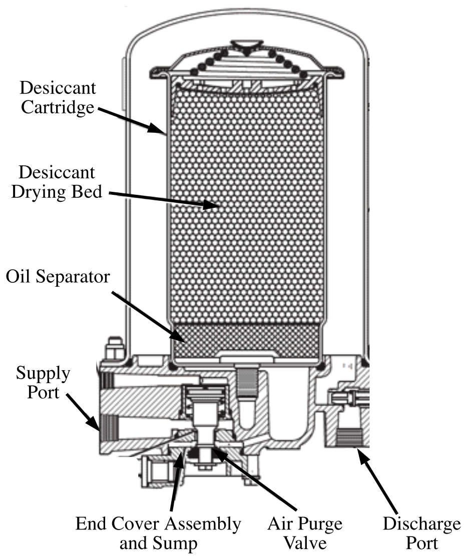
Description of the system

APS in a heavy-truck controls air pressure in the truck's pneumatic system and thus provides compressed air to the critical components such as air brakes, air suspension and gearbox. It contains an air dryer (see Fig. 4.29a) which removes water vapour from compressed air to prevent moisture and condensation from interfering with the critical components.

The air-drying process involves a charge cycle and a purge cycle. During the charge cycle, the compressed air enters from the supply port (see Fig. 4.29b) and travels through the desiccant cartridge whilst changing the direction of flow several times to reduce the temperature and deposit contaminants (e.g. water molecules, oil and solid contaminants) into the bottom of the cover sump. Then the purge cycle begins and expels contaminants deposited at the bottom of the cover sump. This is done using the air purge valve which is attached to the end cover assembly of the air dryer. At the end of each charge cycle, the APS opens the air purge valve to expel water, oil and solid contaminants from the system. Finally, the dry compressed air is served to the pneumatic system through the discharge port.



(a) Scania air dryer with an air purge valve attached to the end cover assembly.



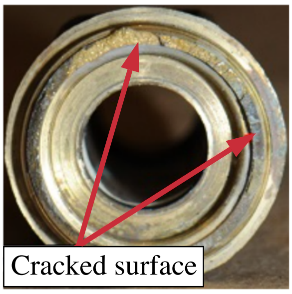
(b) Cut-away view of the Scania air dryer.

Fig. 4.29 Diagrams depicting Scania air dryer and its components. *Adapted from material provided by Scania CV.*

Description of failure

APS in a heavy-truck can fail due to air purge valve surface crack which causes compressed air to be leaked from the APS [123] (also see Fig. 4.30). This often results in the complete immobilisation of the truck since there is an insufficient amount of compressed air in the truck for performing its critical functions (i.e. air braking, air suspension and transmission) [88, 123]. Thus, the failure of heavy-truck APS is defined as follows:

Failure definition. Thermal shock and dust contamination (failure causes) initiate crack (failure mechanism) in the air purge valve surface. The surface crack eventually leads to the leakage of compressed air from the APS (failure mode) causing the truck to contain an insufficient amount of compressed air for performing its critical functions (failure). Hereinafter, this failure definition is referred to as *the APS failure*.



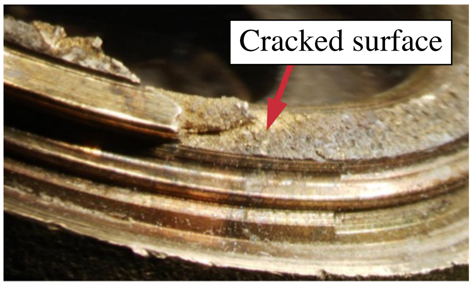


Fig. 4.30 Images depicting an air purge valve with a cracked surface. *Adapted from material provided by Scania CV.*

Prognostics modelling dataset

The dataset used for prognostics modelling is the Scania heavy-truck APS prognostics modelling dataset which was introduced in the 2016 Intelligent Data Analysis Challenge [124]. The dataset contains condition monitoring data collected from 76000 Scania heavy-trucks that perform logistics operations in 5 European markets [88]. The dataset is anonymised due to proprietary reasons, nevertheless, in their publication, the authors of the dataset mentioned that it contains data captured from sensors that measure, for example, air pressure, air consumption, air dryer generation time, air dryer heating time, outside ambient temperature, speed, mileage and engine load of the trucks [88].

Once the data are collected from the onboard condition monitoring system of a truck, they are converted into histogram variables. For example, in the case of outside ambient temperature (T), it is measured for a particular truck as time series data first. Then all the time series data samples for that truck are binned into 4 bins as follows: bin 1 contains data samples for $T < -20$, bin 2 contains data samples for $-20 \leq T < 0$, bin 3 contains data samples for $0 \leq T < 20$, and bin 4 contains data samples for $T > 20$. Thus, each feature in the dataset represents how often a sensor is measured within different intervals [88]. Fig. 4.31 shows the structure of APS prognostics modelling dataset. The label 1 in the *Label* column represents the trucks that have failed due to the APS failure.

Limited failure data availability for Scania air processing system prognostics

The effective implementation of PdM for Scania heavy-truck APS is challenging since the APS failure is rare [125, 126]. More specifically, Scania is committed and responsible for designing, manufacturing, testing and maintaining APS components according to standards that enforce high

Id	Sensor-1_Bin-1	...	Sensor-1_Bin-n	....	Sensor-2_Bin-1	....	Sensor-2_Bin-n	....	Label
Truck-1									0
Truck-2									0
Truck-3									0
Truck-4									0
Truck-5									0
Truck-6									0
Truck-7									0
Truck-8									0
Truck-9									0
⋮									0
⋮									1
Truck-n									1

Trucks failed due to the APS failure

Fig. 4.31 Diagram depicting the Scania heavy-truck APS prognostics modelling dataset structure.

reliability since their failures often lead to catastrophic consequences [127]. For example, air brakes and air suspension failures often lead to accidents and trucks to be stopped on road-sides disrupting vital supply chains such as medical and food supply chains [88, 127]. However, the high reliability of APS components and overprotective maintenance and replacement regimes pose a challenge to prognostics modelling since these components are rarely allowed to run to failure once degradation has been detected [86, 125]. In other words, these reasons reduce the number of degradation trajectories pertaining to APS failure modes, and hence cause failure data to be limited for developing prognostics models [6, 7]. Since limited failure data availability causes prognostics predictions to be associated with high uncertainty, a PdM policy that incorporates these predictions is ineffective (i.e. introduces costs due to under maintenance and over maintenance of APS components and false alarms) for Scania.

The Scania APS prognostics modelling dataset represents the limited failure data availability problem which is caused by the rarity of the occurrence of APS failure [125, 88]. The dataset is published in two parts: train set which contains 60000 data samples and test set which contains 16000 data samples [88]. Out of the 60000 train samples, only 1000 belong to the positive class (i.e. data samples pertaining to the APS failure). This imbalance ratio of 1000:59000 between positive and negative classes means that the positive class only covers 1.6% of the entire train set, whereas the negative class covers 98.4%. Thus, the APS prognostics modelling dataset is considered as a highly imbalanced dataset in the literature [125].

Using the Shannon entropy-based method introduced in Sec. 3.2, the extent of limited failure data availability problem for Scania APS prognostics is measured. The number of positive samples C_1 and negative samples C_2 are 1000 and 59000 respectively. The number of classes K is 2. The normalised Shannon entropy H' of the dataset is therefore 0.12, which indicates a highly imbalanced dataset. Thus, the extent of the problem of limited failure data availability for APS prognostics is high. The following problem statement of the Scania APS prognostics under the conditions of limited failure data availability problem follows from this analysis and aligns with the two research questions of this thesis:

Problem statement. Address the problem of limited failure data availability for Scania heavy-truck APS prognostics, and quantify the impact of methodology on Scania heavy-truck APS prognostics under the conditions of limited failure data availability compared to the existing techniques.

4.2.2 Methodology application and results

The Scania heavy-truck APS prognostics problem satisfies the prerequisite due to the following: similarity between trucks can be used as auxiliary information since the types of trucks and their purpose (e.g. transportation of goods, construction work and garbage collection) have an effect on degradation patterns of the APS failure [88]. The problem satisfies the assumption due to the following: the failure mechanism that needs predicting is crack and it causes equipment to fail under degradation, hence the failure is not random [2].

In the remainder of this section, a description of how the methodology is implemented using the similarity between equipment that has failed under a single failure mode auxiliary information kind to generate plausible APS failure data, and thus address the problem of limited failure data availability for Scania heavy-truck APS prognostics is provided.

Phase 1: Auxiliary information processing

In this phase, suitable pieces of auxiliary information pertaining to the similarity between heavy-trucks that have failed under the compressed air leakage failure mode which is generated by the air purge valve surface crack is identified, validated and converted into the vector form which allows constructing the conditional generator and discriminator of the CGAN.

(I) Auxiliary information identification

A model of the APS failure process is constructed to identify potentially suitable auxiliary information (see Fig. 4.32). Then expert knowledge pertaining to different stages in the APS failure process is obtained from APS component designers and maintenance engineers at Scania. The following auxiliary information is identified from this expert knowledge: *historically, the majority of heavy-trucks that have failed due to air purge valve surface crack perform short-distance haulage or long-distance haulage*; *historically, the type of haulage operation a truck performs affects the crack growth of its air purge valve surface differently*. The latter is due to the following two reasons:

- The impact of rapid change in temperature of air purge valve surface material which leads to the thermal shock failure cause is observed to be different for short-distance and long-distance haulage trucks - long-distance haulage trucks are affected more by the thermal shock than short-distance haulage trucks since the former perform haulage operations across multiple geographical regions in Europe, and therefore operate under frequently changing weather conditions.
- The impact of dust contamination failure cause on air purge valve surface crack is observed to be different for short-distance and long-distance haulage trucks - short-distance haulage trucks

are affected more by dust contamination than long-distance haulage trucks since the former perform haulage operations in construction sites and urban areas.

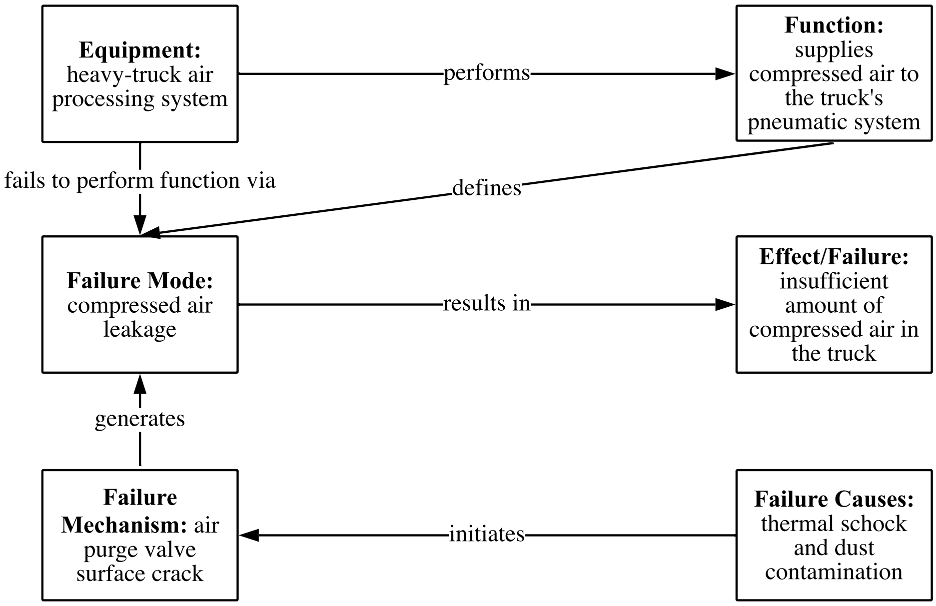


Fig. 4.32 Diagram depicting a model of APS failure process which is based on APS failure definition.

(II) Auxiliary information validation

Cluster analysis is used in this case study to validate the identified auxiliary information since degradation patterns in condition monitoring data should represent how the two types of haulage operations affect differently on the air purge valve surface crack. More specifically, according to the auxiliary information identified from expert knowledge, thermal shock and dust contamination failure causes are observed to impact differently on the air purge valve surface crack depending on the type of haulage operation a truck performs. Hence, condition monitoring data in the APS prognostics modelling dataset are expected to represent two clusters for the trucks that have failed due to the APS failure.

In order to validate the identified auxiliary information, a subset that only contains failure data samples in the APS prognostics modelling dataset is created first. Then clustering algorithms are used to identify the natural groupings in the newly created subset. The performance of clustering algorithms is evaluated using the Silhouette analysis, which allows studying the separation distance between the clusters produced by a clustering algorithm using a metric called the average silhouette score [99]. The worst value of the average silhouette score is -1, and the best value is 1. The negative values indicate that the samples are assigned to a wrong cluster, and positive values indicate the samples are properly clustered.

Table 4.12 summarises the average silhouette scores obtained by the clustering algorithms for a different number of clusters. Given the stochasticity in the real-world dataset, it is reasonable not to expect the values obtained for this metric to be closer to 1 [88]. It can be observed from the table that all the algorithms have obtained a highest average silhouette score when the number of clusters is 2 and as the number of clusters increases, the clustering performance of all the algorithms decreases. Moreover, the *k*-means algorithm has obtained the highest average silhouette score among all the algorithms. The average silhouette scores obtained by this algorithm are also depicted in Fig. 4.33. The 2 clusters of failed trucks detected by the algorithm can be observed visually from Fig. 4.34.

It can be concluded from the above analysis, in the APS prognostics modelling dataset, there are two groups of trucks that have failed due to the APS failure, and thus the auxiliary information identified from expert knowledge provided by the APS component designers and maintenance engineers is valid. This means, the following abstract piece of auxiliary information which is based on the similarity between Scania heavy-trucks that have failed under the compressed air leakage failure mode, which is generated by air purge valve surface crack can be used in the methodology for generating plausible APS failure data:

Auxiliary information. A Scania heavy-truck that has failed due to the APS failure belongs to one of two groups.

Table 4.12 Clustering performance of different clustering algorithms used for identifying the natural groupings of Scania heavy-trucks that have failed due to the APS failure.

Clustering algorithm	Number of clusters	Average Silhouette Score
k-means	2	0.2531
	3	0.18
	4	0.1748
	5	0.1517
Gaussian Mixture	2	0.2491
	3	0.1455
	4	0.1754
	5	0.1518
Agglomerative Hierarchical Clustering	2	0.2332
	3	0.1554
	4	0.1567
	5	0.1406
Spectral Clustering	2	0.0889
	3	0.0751
	4	-0.0176
	5	-0.0116

(III) Auxiliary information conversion

Constructing an vector of auxiliary information from class labels approach is used in this case study to convert the validated auxiliary information. As discussed in Sec. 3.3.2, this approach allows using the natural grouping in equipment that has failed to construct an auxiliary information vector.

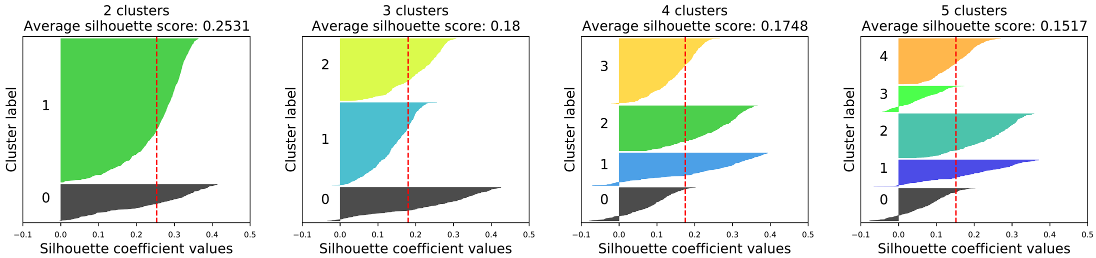


Fig. 4.33 Silhouette plots of natural groupings identified by the k-means algorithm for the failed trucks. The average silhouette score is depicted by the red dotted line.

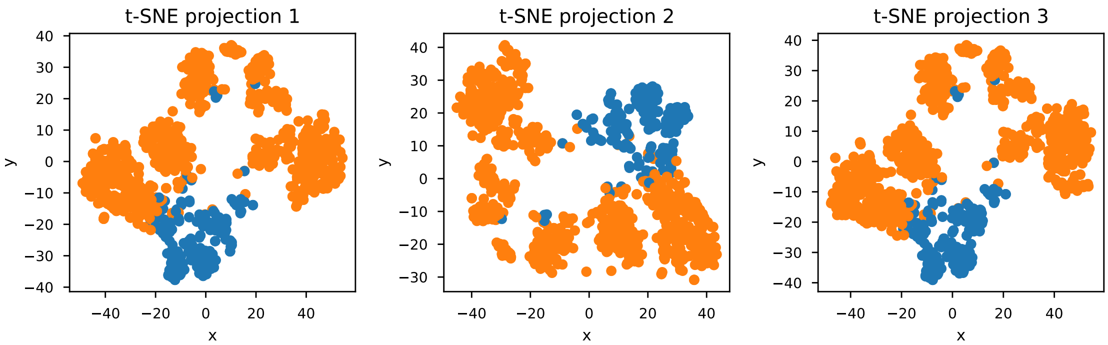


Fig. 4.34 t-SNE (t-distributed Stochastic Neighbour Embedding) projections of the clusters of failed trucks identified by the k-means algorithm. t-SNE algorithm is a tool for visualising high dimensional data in a 2-dimensional space [128]. In the figures, each sample in the failure data subset (i.e. each failed truck) is first projected into a 2-dimensional space as individual data points using the t-SNE algorithm. Then each data point is coloured based on its cluster which is detected by the k-means clustering algorithm.

In order to convert the validated auxiliary information, each failed truck in the APS prognostics modelling dataset is labelled using the cluster label predicted by the k-means clustering algorithm. Then a new label column is added to the APS prognostics modelling dataset as shown in Fig. 4.35. This means, the dataset now stores data that encode the validated auxiliary information. This column is also a vector representation of the auxiliary information: class labels that represent the two groups of failed trucks with natural numbers 0 and 1 are collectively is a vector of natural numbers $Y = \{y \in \mathbb{N} | 0 \leq y \leq 1\}$.

Phase 2: Conditional generative model estimation

In this phase, the vector of auxiliary information and the limited amount of APS failure data samples available are integrated to the CGAN to estimate a generative model that is capable of generating plausible APS failure data samples.

Id	Sensor-1_Bin-1	...	Sensor-1_Bin-n	Sensor-2_Bin-1	....	Sensor-2_Bin-n	....	Label	Equipment Similarity Information Vector
Truck-1								0	
Truck-2								0	
Truck-3								0	
Truck-4								0	
Truck-5								0	
Truck-6								0	
⋮								0	
⋮								1	0
⋮								1	0
⋮								1	0
⋮								1	1
⋮								1	1
Truck-n								1	1

Trucks failed due to the APS failure

Failed trucks belong to Group 0

Failed trucks belong to Group 1

Fig. 4.35 Diagram depicting the Scania heavy-truck APS prognostics modelling dataset with the converted vector of auxiliary information pertaining to the similarity between equipment that has failed under a single failure mode.

(I) Data preprocessing and structuring

Before using it to estimate the generative model, the APS prognostics modelling dataset is preprocessed as per the process shown in Fig. 3.5 discussed in Sec. 3.3.3. The Soft-Impute algorithm is used to impute missing data in the train set since it has been highly successful in the literature for imputing missing data in the APS prognostics modelling dataset [125]. Then the dataset is structured according to Fig. 3.6. At the end of this step, below five datasets are produced. These will be used in the remainder of this phase for estimating and evaluating a generative model that is capable of generating plausible APS failure data samples.

(a) Stratified train set: the stratified train split of the APS prognostics modelling dataset which also includes the vector of auxiliary information which encodes the similarity between heavy-trucks that have failed due to the APS failure.
(b) Stratified validation set: the stratified validation split of the APS prognostics modelling dataset which also includes the vector of auxiliary information which encodes the similarity between heavy-trucks that have failed due to the APS failure.
(c) Stratified test set: the stratified test split of the APS prognostics modelling dataset which also includes the vector of auxiliary information which encodes the similarity between heavy-trucks that have failed due to the APS failure.
(d) Train failure data subset: a subset of stratified train set that only contains APS failure data samples in the stratified train set.
(e) Train non-failure data subset: a subset of stratified train set that only contains non-failure data samples in the stratified train set.

(II) Model construction and training

As discussed in Sec. 3.3.3, the methodology integrates noise Z, the real failure data samples X, and a vector of auxiliary information pertaining to the failure mechanism Y to estimate generative models for plausible failure data generation. In this case study, Z is noise sampled from a vector of standard multivariate normal distribution, X is the train failure data subset, and Y is the vector of auxiliary information that encodes the similarity between Scania heavy-trucks that have failed due to the compressed air leakage failure mode generated by air purge valve surface crack. Conditioning the generator and discriminator on auxiliary information pertaining to the APS failure allows controlling and directing the APS failure data generation process. Thus, the estimated generative model is able to generate new APS failure data samples that conform to the following condition: *a newly generated APS failure data sample should belong to one of two groups*. Note that since the dataset contains histogram data, each data sample represents a heavy-truck as shown in Fig. 4.35. Hence, each generated failure data sample also represents a plausible failed truck.

The conditional generator and discriminator of the CGAN are constructed using X,Y and Z as per the description provided in Sec. 3.3.3. Then the CGAN is trained using the minibatch stochastic gradient descent algorithm presented in Algorithm 1. The hyperparameters used to train the CGAN in this case study are outlined in Table. 4.13.

Table 4.13 Hyperparameters used to train the CGAN for estimating a generative model that is capable of generating plausible heavy-truck APS failure data.

Hyperparameter	Value
Optimiser	Adam
Learning rate	$2e-4$
Weight decay	$1e-6$
Exponential decay rate for first moment estimate	0.3
Exponential decay rate for second moment estimate	0.9
Number of batches	32
Number of epochs	45
Number of training iterations	1400

(III) Model convergence evaluation

As shown in Fig. 4.36, the learning curves of CGAN trained in the previous step provide a visual indication of whether the convergence is achieved. The following can be observed from these learning curves: (i) discriminator loss on real and generated APS failure data samples is stable around 1/2 which indicates that the discriminator cannot discriminate between real and generated APS failure data samples; (ii) generator loss has become stable as the number of training iterations increases which indicates the generator is consistent; (iii) variance of generator and discriminator loss is modest. As discussed in Sec. 3.3.3, these properties are expected from learning curves of a CGAN that has achieved an equilibrium between its generator and discriminator.

Since the analysis of learning curves is only a qualitative evaluation, the K-S test is performed to conduct a more robust and quantitative evaluation. Table 4.14 outlines the K-S test results obtained for a set of features in the APS prognostics modelling dataset. It can be observed that all the p-values produced by the K-S test are above 0.05 which indicates weak evidence against the null hypothesis.

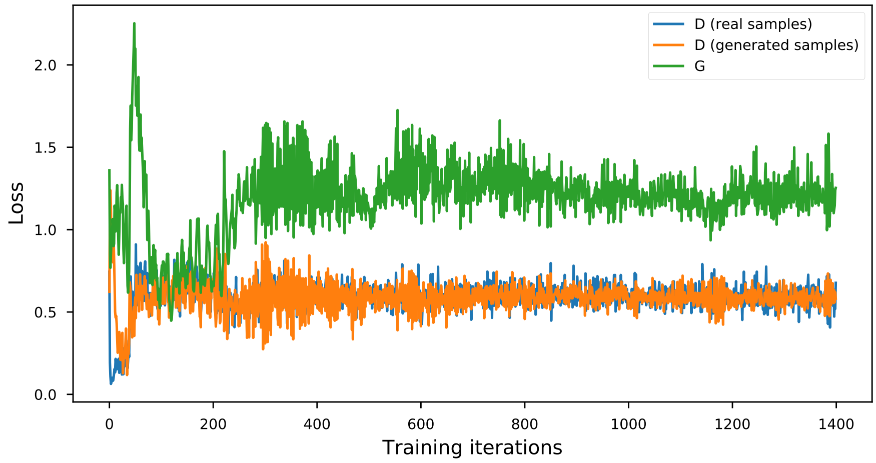


Fig. 4.36 The generator (G) and discriminator (D) learning curves of the CGAN trained for heavy-truck APS failure data generation.

Thus, the null hypothesis cannot be rejected and therefore the CGAN has achieved the convergence. More specifically, the generative model trained in the CGAN, $G^*(Z \mid Y)$ can replicate the real APS failure data distribution, and hence capable of generating plausible APS failure data samples.

Table 4.14 K-S test results obtained for a set of features when statistically comparing the real APS failure data distribution to the generated APS failure data distribution.

Feature	K-S test p-value
aa-000	0.97
ac-000	0.94
ag-004	0.93
ah-000	0.94
an-000	0.83
aq-000	0.98
ba-002	0.8
bd-000	0.93
bi-000	0.93
bs-000	0.93
cn-004	0.45
cv-000	0.67
cs-007	0.57
cz-000	0.46
du-000	0.86
dn-000	0.93
eb-000	0.94
ee-001	0.98
ee-005	0.86
ee-008	0.87

(IV) Model overfitting assessment

The MMD test outlined in Sec. 3.3.3 is employed to evaluate whether the generator is overfitting. The p-value observed for this statistical test is 0.39 (> 0.05) which indicates weak evidence against the null hypothesis. This means the null hypothesis cannot be rejected, thus the generative model, $G^*(Z \mid Y)$ is not overfitting to train APS failure data. The convergence performance of the estimated generative model is therefore not affected by overfitting.

Phase 3: Plausible failure data generation

The generative model estimated by conditioning the failure data generation process on the similarity between Scania heavy-trucks that have failed due to the APS failure is used to generate plausible APS failure data samples in this phase. These data samples are then used to augment the original train set produced in the data preprocessing and structuring step in the previous section, so that an increased number of APS failure data samples is available for prognostics modelling.

(I) Plausible failure datasets generation

As per the process outlined in Fig. 3.11, a new noise vector and the vector of auxiliary information pertaining to the APS failure are used as inputs to the generative model to produce a dataset consists of plausible APS failure data samples. More specifically, given the joint distribution of noise and auxiliary information $Pr(Z,Y)$ as the input, the trained generator $G^*(Z \mid Y)$ produces a dataset X_{G^*} consists of plausible APS failure data samples.

Multiple datasets are produced so that each dataset consists of more plausible APS failure data samples than its predecessor (e.g. $X_{G^*_{100}}$ has 100 samples, $X_{G^*_{200}}$ 200 samples, $X_{G^*_{300}}$ has 300 samples and so on). A new plausible APS failure dataset is generated until the normalised Shannon entropy of the corresponding augmented train set of the last generated plausible APS failure dataset is equal to 1, that is, until the augmented train set is no longer imbalanced.

(II) Train set augmentation

In order to produce augmented train sets, samples from each plausible APS failure dataset generated in the previous step are added to a copy of the original train set (also see Fig. 3.12). The augmented train sets are used in the next phase for APS prognostics modelling, and the set that produced the best prognostics performance is selected for quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques.

Phase 4: Prognostics modelling

In this phase, a suitable prognostics model for APS failure prediction is estimated using the augmented train sets and evaluated on the original test set.

(I) Prognostics modelling approach establishment

The binary classification-based approach introduced in Sec. 3.3.5 is selected for this case study for prognostics modelling. The reason for this is the prognostics modelling dataset is structured and labelled for the binary classification problem by Scania since the challenge is to predict whether a data sample captured from the onboard condition monitoring system of a heavy-truck is pertaining to an imminent APS failure (i.e. APS failure within next 1 day) or not [88]. The objective of Scania heavy-truck APS prognostics modelling is therefore to predict the APS failure probability within a prediction horizon of 1 day for the trucks in the test set using a probabilistic binary classifier, and then calculate the reliability of these trucks within the prediction horizon.

(II) Prognostics model estimation and evaluation

In the previous phase, multiple augmented train sets are produced and each of these has a different number of plausible failure data samples. These augmented train sets, and the unchanged validation and test sets are used in the process recommended by the methodology (see Fig. 3.16) to estimate a prognostics model for APS prognostics. The augmented train set that produced the best prognostics performance is the set containing 4000 plausible APS failure data samples. In other words, when the data imbalance ratio is 5000:59000 compared to 1000:59000 in the original train set. More formally, the normalised Shannon entropy of the augmented train set with 4000 plausible APS failure data samples and 1000 real APS failure data samples is 0.4 compared to the 0.12 in the original train set. Thus, the 4000 plausible APS failure data samples generated by the methodology reduced the extent of limited APS failure data availability problem.

In the following, prognostics performance obtained when trained on the augmented train set and the increase in prognostics performance provided by the methodology compared to the performance on original train set are discussed:

(a) Prognostics performance on augmented train set

Table 4.15 summarises the prognostics performance obtained by different prognostics models when trained on the augmented train set and evaluated on the test set. It can be observed that the GBM, RF and SVM-based prognostics models obtained a Kappa statistic higher than 0.81 which means, their prognostics predictions are associated with minimal uncertainty (also see Table 3.3).

Table 4.15 Prognostics performance obtained by models when trained on the augmented train set and evaluated on the test set.

Prognostics model	Kappa statistic	Precision		Recall	
		Failure data class	Non-failure data class	Failure data class	Non-failure data class
GBM	0.8814	0.947	0.9977	0.96	0.9965
RF	0.8637	0.9142	0.9958	0.824	0.9981
SVM	0.8612	0.9088	0.9956	0.824	0.998
AdaBoost	0.7476	0.7031	0.9961	0.84	0.9915
DT	0.7171	0.7448	0.9943	0.7627	0.9937
kNN	0.7077	0.8577	0.9907	0.6107	0.9976
MLP	0.6572	0.8565	0.9891	0.5413	0.9978
NB	0.3434	0.2278	0.9989	0.96	0.9219

Table 4.15 also outlines the precision and recall obtained by the prognostics models. All the models have obtained a high precision and recall for the non-failure data class since there is a large number of non-failure data samples available in the train set (i.e. 59000 data samples). The GBM-based prognostics model has obtained the highest precision and recall for the failure data class. This means it produces the lowest number of false alarms and undetected failures with minimal uncertainty. This can also be observed from the precision-recall curves of GBM, RF and SVM-based prognostics models shown in Fig. 4.37. The GBM-based model has the highest area under the curve, hence the highest precision and recall.

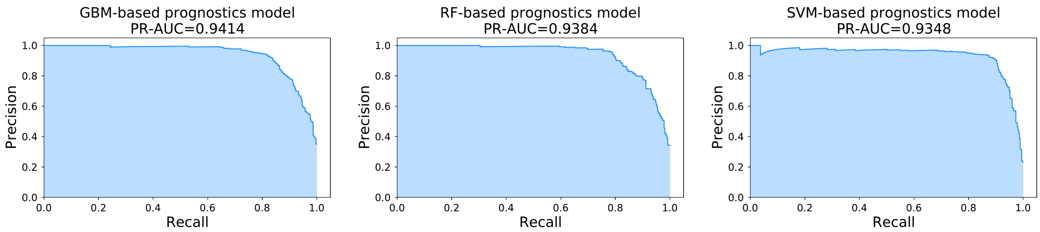


Fig. 4.37 Plots depicting precision-recall (PR) curves and the area under the curve (PR-AUC) of GBM, RF and SVM-based prognostics models.

(b) Prognostics performance compared to original train set

The impact of the methodology on Scania heavy-truck APS prognostics can be observed from Fig. 4.38. The figure compares the Kappa statistic, precision and recall obtained by GBM, RF and SVM-based prognostics models when trained on the augmented train set and original train set.

Compared to when trained on the original train set, the Kappa statistic and precision and recall of the failure data class are significantly higher when trained on the augmented train set. More specifically, the best model (i.e. GBM-based prognostics model) has obtained a 9% increase in Kappa statistic (i.e. reduced the uncertainty associated with prognostics predictions), 18% increase in precision and 32% increase in recall on the failure data class when trained on the augmented train set. This is since, in addition to the real failure data samples, the augmented train set includes plausible failure data samples which are generated by conditioning the failure data generation process on the similarity between equipment that has failed under a single failure mode auxiliary information kind.

The reliability-based confusion matrixes of the GBM-based prognostics model are shown in Fig. 4.39. It can be observed that when trained on the augmented train set, the model produces 33% more failure detections (i.e. true positives) and 67% less false alarms (i.e. false positives) compared to when trained on the original train set.

Validation of prognostics performance

In this section, the methodology is evaluated against the existing techniques proposed in the literature for Scania APS prognostics under the conditions of limited failure data availability. Since the IDA 2016 challenge test set (i.e. another test set collected from the Scania APS prognostics industrial

Fig. 4.38 Plots comparing prognostics performance obtained by GBM, RF and SVM-based prognostics models when trained on the augmented train set and original train set. The prognostics performance is evaluated on the test set.

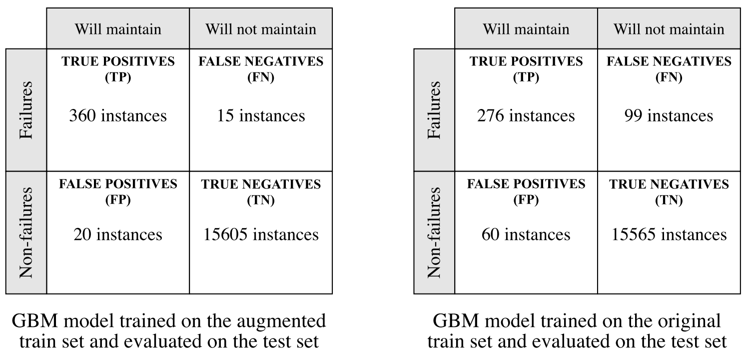


Fig. 4.39 Reliability-based confusion matrixes comparing the prognostics performance of GBM-based prognostics model when trained on the augmented train set and original train set. The prognostics performance is evaluated on the test set.

scenario) is used in this evaluation, this evaluation also serves the purpose of further validating the prognostics model performance and its consistency using extra data collected from the industrial scenario.

There are a few techniques proposed in the literature for addressing the problem of Scania APS prognostics under the conditions of limited failure data availability. Table 4.16 outlines the results obtained by the top three existing techniques. In the remainder of this section, the performance of the methodology is evaluated against these techniques.

When comparing the prognostics performance on Scania heavy-truck APS prognostics modelling dataset against the existing techniques proposed in the literature, the evaluation method provided in 2016 Intelligent Data Analysis challenge is commonly used [88, 124]. The prognostics problem is modelled as a binary classification task in which the challenge is to predict whether a truck faces an imminent APS failure (i.e. APS failure within next 1 day) [88]. The anonymised cost of an undetected

Table 4.16 Prognostics performance (T_{Cost} discussed below) obtained by the top three existing techniques for Scania heavy-truck APS prognostics [124, 88]. These results are the performance obtained when each technique was evaluated on the test set [124, 88].

Rank	T_{Cost} (€)	Undetected failures	False alarms	Technique reference	Performance reference
1	9920	9	542	[125]	[88, 124]
2	10900	12	490	[129]	[88, 124]
3	11430	12	543	[88]	[88, 124]

APS failure is €500 (denoted C_{FN}), and the anonymised cost of a false alarm is €10 (denoted C_{FP}) [88]. C_{FN} is due to the cost of an undetected APS failure that causes a breakdown of a truck. C_{FP} is due to the unnecessary checks that need to be done by a mechanic at the workshop due to false alarms. Let m be the number of undetected APS failures and n be the number of false alarms, the total cost of breakdowns and false alarms (denoted T_{Cost}) is given by the following:

$$T_{\text{Cost}} = mC_{FN} + nC_{FP}$$

The correct prediction of APS failures is given priority over false alarms since undetected failure results in a larger penalty (i.e. €500 compared to €10) [88]. Hence, as shown in Eq. 4.1, the APS prognostics problem is formulated as an optimisation problem [88]. Therefore in this evaluation, the objective of Scania heavy-truck APS prognostics under the conditions of limited failure data availability is to minimise T_{Cost} whilst optimising a binary classifier f to predict the optimal value pair for m and n with minimal uncertainty.

$$\begin{aligned} \underset{f}{\text{minimise}} \quad & T_{Cost} = mC_{FN} + nC_{FP}, \\ \text{subjected to} \quad & m \geq 0, \\ & n \geq 0, \\ & C_{FN} = 500, \\ & C_{FP} = 10. \end{aligned} \tag{4.1}$$

As discussed in Sec. 2.1.3 in the literature review chapter, cost-sensitive learning is a technique that allows taking the cost of prediction errors into account when developing prognostics models [81]. This technique is employed to integrate the optimisation problem in Eq. 4.1 into the GBM-based prognostics model. More specifically, instead of each data sample being either correctly or incorrectly classified, failure data and non-failure data classes are given the misclassifications costs of €500 and €10 respectively. Thus, instead of trying to optimise precision and recall, the GBM-based prognostics model now optimises precision and recall to minimise the total misclassification cost T_{Cost}.

The prognostics model is trained on the augmented train set that includes 4000 plausible APS failure data samples and evaluated on the test set. Fig. 4.40 shows the reliability-based confusion matrixes produced by the methodology and the best technique previously proposed in the literature (i.e.

the technique with T_{Cost} of €9920 in Table. 4.16). It can be observed that the methodology outperforms the best technique proposed in the previous literature by a large margin. More specifically, the T_{Cost} achieved by the methodology is €5550, and compared to the performance of the best technique in the existing literature, this is a 44% reduction of costs due to breakdowns and false alarms for Scania.

	Will maintain	Will not maintain
Failures	TRUE POSITIVES (TP) 369 instances	FALSE NEGATIVES (FN) 6 instances €3000 loss
Non-failures	FALSE POSITIVES (FP) 255 instances €2550 loss	TRUE NEGATIVES (TN) 15370 instances

Methodology

	Will maintain	Will not maintain
Failures	TRUE POSITIVES (TP) 366 instances	FALSE NEGATIVES (FN) 9 instances €4500 loss
Non-failures	FALSE POSITIVES (FP) 542 instances €5420 loss	TRUE NEGATIVES (TN) 15083 instances

Best solution proposed in the literature

Fig. 4.40 Reliability-based confusion matrixes summarising the prognostics performance obtained by the methodology compared to the best technique previously proposed in the literature. The methodology produced a Kappa statistic of 0.87, hence this performance is also associated with minimal uncertainty.

Summary of methodology application using the similarity between failed equipment

The problem statement of this case study stated in Sec. 4.2.1, which aligns with the two research questions of this thesis involves addressing the problem of limited failure data availability for prognostics and quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques.

In this case study, the methodology which addresses the first research question of this thesis is applied to generate plausible failure data samples using the similarity between equipment that has failed under a single failure mode auxiliary information kind, and thus address the problem of limited failure data availability for prognostics in practice. It is shown that the methodology is capable of addressing this problem by producing a prognostics model that is capable of estimating accurate prognostics predictions with minimal uncertainty. Moreover, the methodology has outperformed existing techniques used in the literature for addressing the problem of Scania APS prognostics under the conditions of limited failure data availability by a large margin. Thus, this case study has shown that the application of the methodology using the similarity between equipment that has failed under a single failure mode auxiliary information kind addresses the problem of limited failure data availability for prognostics in practice.

4.2.3 Analysis of prognostics impact and key factors influencing effectiveness

In this section, insights are produced using the Scania heavy-truck APS prognostics case study to quantify the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques, and to identify the key factors influencing the effectiveness of methodology when applied using the similarity between equipment that has failed under a single failure mode auxiliary information kind.

Impact on prognostics under the conditions of limited failure data availability

The GBM algorithm is used to model prognostics as it produced the best prognostics performance. Fig. 4.41 provides a comparison of the uncertainty associated with prognostics predictions produced by the methodology and existing techniques. It can be observed that the methodology provides a significant reduction in uncertainty compared to the existing techniques. Moreover, whilst the methodology reduces the uncertainty compared to the original dataset which has the limited failure data availability problem, existing techniques have increased the uncertainty compared to the original dataset. Hence, the methodology is successful in reducing the uncertainty associated with prognostics predictions compared to the existing techniques. More specifically, the uncertainty reductions outlined in Table 4.17 are provided by the methodology compared to the existing techniques.

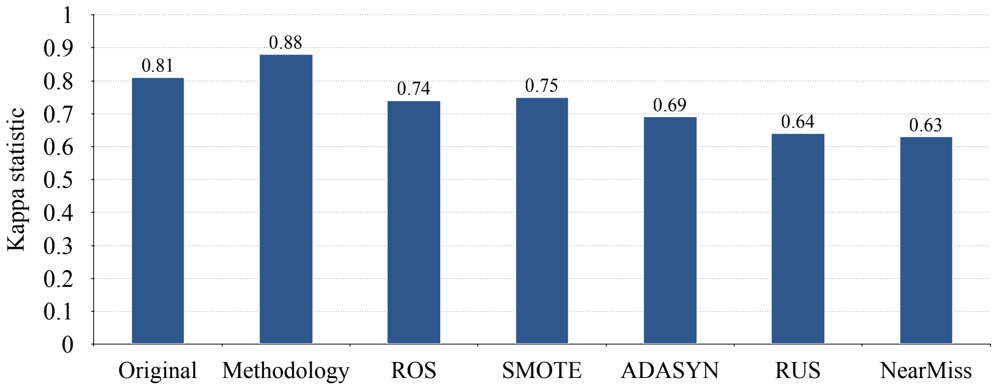


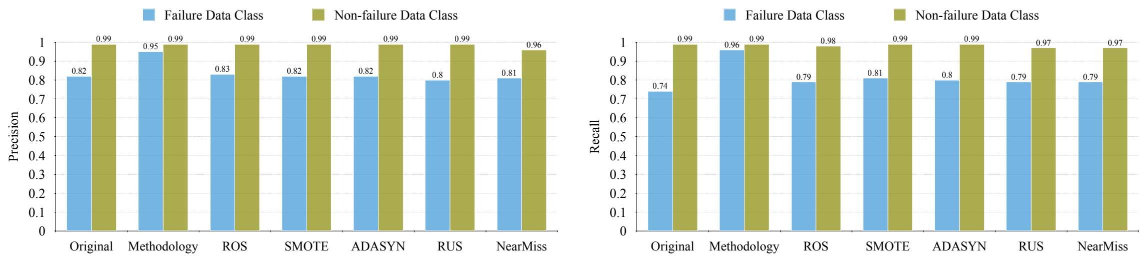
Fig. 4.41 Comparison of uncertainty associated with prognostics predictions produced by the methodology and existing techniques. The uncertainty produced on the original dataset is denoted as Original.

Table 4.17 Reduction in uncertainty (i.e. increase in Kappa statistic) provided by the methodology for APS prognostics compared to the existing techniques.

Reduction in uncertainty				
ROS	SMOTE	ADASYN	RUS	NearMiss
19%	17%	28%	38%	40%

In Fig. 4.42, the prognostics performance provided by the methodology is compared to the existing techniques. It can be observed that the methodology allows predicting a higher number of failures (i.e.

has a higher recall on failure data class) and causes a lower number of false alarms (i.e. has a higher precision on failure data class) compared to all the existing techniques. More specifically, the increase in prognostics performance outlined in Table 4.18 is provided by the methodology compared to the existing techniques.



(a) Precision (higher the precision lower the number of false alarms)

(b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.42 Comparison of prognostics performance produced by the methodology and existing techniques.

Table 4.18 Decrease in the number of false alarms (i.e. increase in precision) and decrease in the number of undetected failures (i.e. increase in recall) provided by the methodology for APS prognostics compared to the existing techniques.

	Failure data class	
	Precision	Recall
ROS	14%	22%
SMOTE	16%	19%
ADASYN	16%	20%
RUS	19%	22%
NearMiss	17%	22%

The reason for the significant increase in prognostics performance is since the methodology is capable of addressing the problem of limited failure data availability by generating new and plausible failure data samples. In contrast, as discussed in the literature review chapter, the existing techniques either duplicate existing failure data or randomly generate data (oversampling techniques) or do not increase the number of failure data samples available for prognostics modelling (undersampling techniques). Hence, they do not address the fundamental problem of limited failure data availability for prognostics. The following insight can be gained from this analysis:

Insight. When the methodology is applied using the similarity between equipment that has failed under a single failure mode auxiliary information kind for prognostics under the conditions of limited failure data availability, it produces effective prognostics predictions and outperforms existing techniques used in the literature by a large margin.

Key factors influencing the effectiveness of methodology

In this section, an analysis is performed from a practical perspective to identify the influence of real failure data availability and auxiliary information availability factors on the effectiveness of methodology in practice (i.e. the influence on the prognostics performance) when it is applied using the similarity between equipment that has failed under a single failure mode auxiliary information kind. The scenarios used in this analysis are similar to the ones used in the previous case study and are described in Table 4.19.

Table 4.19 Description of scenarios considered in the analysis.

Scenario name	Parameters	Variation
Base scenario	X is failure data in APS_1 (see Fig. 4.43), Y is the vector encoding the similarity between trucks that have failed due to APV surface crack and is given, $Z \sim N(0,1)$ and is fixed	
Scenario 1		$X \rightarrow$ failure data in APS_2
Scenario 2		$X \rightarrow$ failure data in APS_3
Scenario 3		$X \rightarrow$ failure data in APS_4
Scenario 4		$X \rightarrow$ failure data in APS_5
Scenario 5		$X \rightarrow$ failure data in APS_6
Scenario 6		Y is given $\rightarrow Y$ is not given (not conditioned)
Scenario 7		Y is auxiliary $\rightarrow Y$ is arbitrary

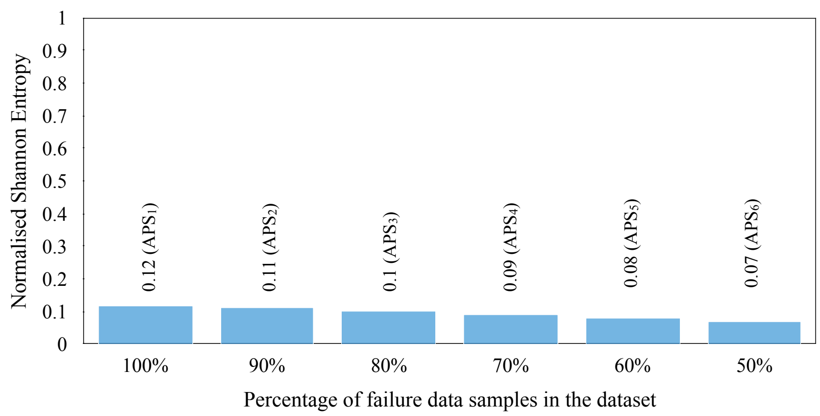


Fig. 4.43 Plot depicting how the extent of limited failure data availability problem is increased for analysing the effect of real failure data availability. The failure data are incrementally removed from the APS dataset and the resulting datasets are denoted APS_1, APS_2, APS_3 and so on.

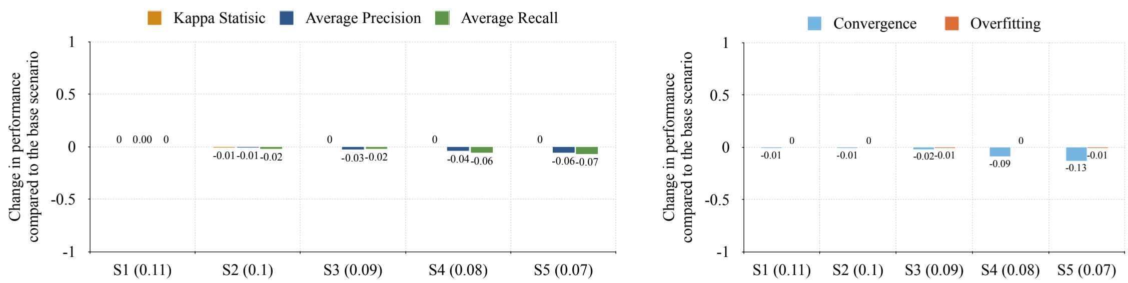
(I) Effect of real failure data availability on prognostics performance

The effect on prognostics performance when scenarios 1 to 5 are applied to APS prognostics modelling is shown in Fig. 4.44. It can be observed that the amount of real failure data samples has a noticeable effect on prognostics performance produced by the methodology when the normalised

Shannon entropy is less than 0.1 (see Scenario 4 and 5). The reason for this can be observed from the convergence performance which shows that the convergence of the CGAN model degrades when the extent of limited failure data availability is less than 0.1 normalised Shannon entropy, which is also observed in the theoretical analysis performed in Sec. 3.4.5.

Fig. 4.45 shows the effect of real failure data availability on the prognostics performance produced by the methodology compared to the existing techniques. It can be observed that whilst the methodology outperforms existing techniques in achieving a low uncertainty for prognostics predictions in all the scenarios, its precision and recall starts to degrade when the extent of limited failure data availability is less than 0.1 normalised Shannon entropy. The following insight which is also produced in the previous case study can be gained from this analysis:

Insight. In practice, when the availability of real failure data samples is approximately equal to or less than 0.1 normalised Shannon entropy, the prognostics performance of the methodology starts to become less effective and eventually becomes comparable to the existing techniques as the availability of real failure data is reduced.



(a) Prognostics performance

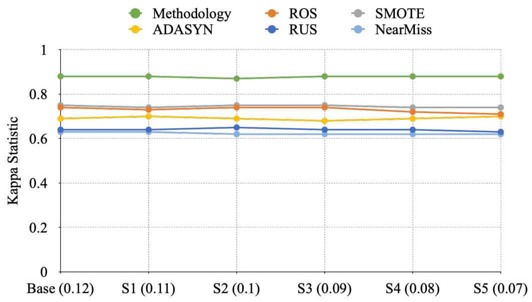
(b) Convergence performance

Fig. 4.44 APS convergence and prognostics performances produced by the methodology for scenarios (denoted S) 1 to 5 compared to the base scenario. The x-axis shows the scenario and corresponding extent of limited failure data availability (i.e. the normalised Shannon entropy).

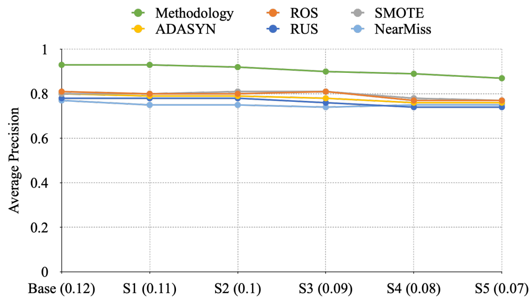
(II) Effect of auxiliary information availability on prognostics performance

It can be observed from Fig. 4.46, the prognostics performance produced by the methodology significantly reduces when auxiliary information pertaining to failure mechanisms is not integrated for controlling and directing the plausible failure data generation process. The effect of auxiliary information availability on prognostics performance compared to the existing techniques is shown in Fig. 4.47. It can be observed that the prognostics performance produced by the methodology for Scenario 6 and 7 is significantly low compared to the base scenario and all the existing techniques. The following insight which is also produced in the previous case study and theoretical analysis can be gained from this analysis:

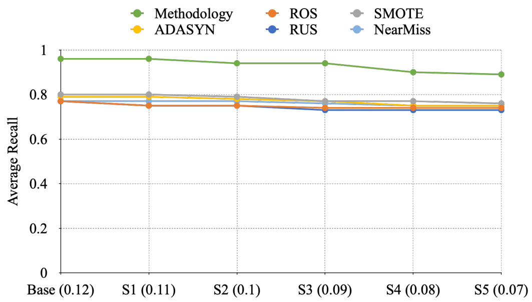
Insight. The integration of auxiliary information pertaining to failure mechanisms which allows controlling and directing the plausible failure data generation process is crucial to the methodology for producing effective prognostics under the conditions of limited failure data availability.



(a) Uncertainty

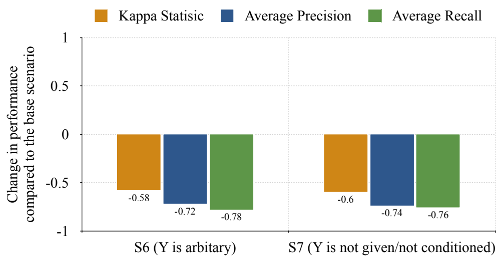


(b) Precision

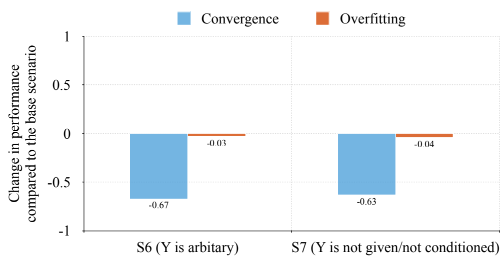


(c) Recall

Fig. 4.45 APS prognostics performance comparison for scenarios 1 to 5. Higher the precision lower the number of false alarms, and higher the recall lower the number of undetected failures.



(a) Prognostics performance



(b) Convergence performance

Fig. 4.46 APS convergence and prognostics performances produced by the methodology for scenarios 6 and 7 compared to the base scenario.

4.3 Case study 3: Effect of harsh use of equipment as auxiliary information

The harsh use of equipment, for example, the harsh acceleration and braking in vehicles due to aggressive driving behaviour and lead foot syndrome increase the wear-and-tear of equipment [97]. The methodology supports using the effect of harsh use of equipment on failure mechanisms as auxiliary information to condition the plausible failure data generation process.

The example discussed in this section uses the effect of harsh acceleration and braking on the low cycle fatigue (LCF) failure mechanism of compressor wheels used in the heavy-truck turbochargers.

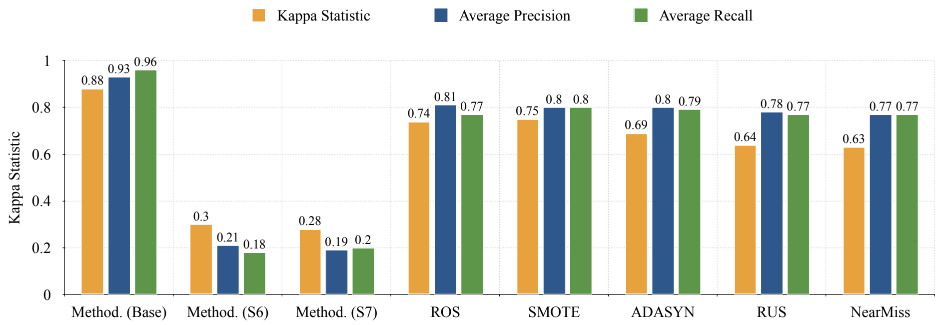


Fig. 4.47 APS prognostics performance produced by the methodology for scenarios 6 and 7 compared to the base scenario and existing techniques.

More specifically, the case study uses critical regions in the engine load matrix (i.e. a matrix comparing engine load and engine speed) which are caused by harsh acceleration and braking as auxiliary information to generate plausible failure data samples for the turbochargers used in Scania 6-cylinder diesel engines. Thus, the methodology addresses the limited turbocharger failure data availability problem for prognostics which is due to the following reasons to enable the effective implementation of PdM for Scania heavy-trucks: the rarity of Scania heavy-trucks being operated in the critical regions of the engine load matrix; and the high reliability of turbocharger components used in Scania heavy-trucks.

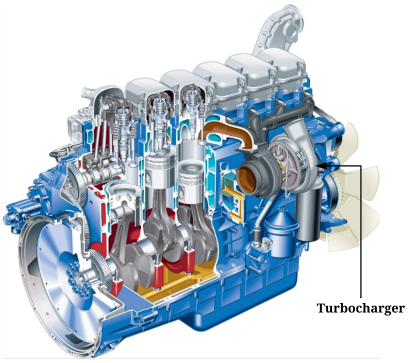
4.3.1 Scania heavy-truck turbocharger prognostics under the conditions of limited failure data availability

In this section, a description of the turbocharger used in Scania heavy-trucks and its failure, and an introduction to the prognostics modelling dataset are provided. Then the problem statement for Scania heavy-truck turbocharger prognostics under the conditions of limited failure data availability is stated.

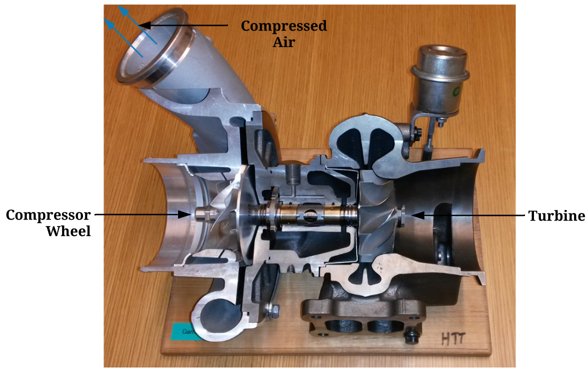
Description of the system

The function of a turbocharger is to provide extra compressed air into the internal combustion engine for increasing the power output and provide overall improvements in weight saving and fuel economy. Fig. 4.48a shows the Scania turbocharger compound attached to the 6-cylinder diesel engines in Scania heavy-trucks.

The two key sections of a turbocharger are the turbine and the compressor. The turbine provides the mechanical energy to drive the compressor by converting engine exhaust gas into mechanical energy (also see Fig. 4.48b). As the turbine turns the compressor wheel, the high-velocity spinning draws in air and compresses it. The compressor then converts the high-velocity, low-pressure air stream into a high-pressure, low-velocity air stream through the diffusion process. The compressed air is then pushed into the combustion chamber of the engine, allowing the engine to burn more fuel and thus produce more power.



(a) Scania 6-cylinder diesel engine with the turbocharger compound.



(b) Cut-away view of the Scania turbocharger.

Fig. 4.48 Diagrams depicting Scania turbocharger and its components. *Adapted from material provided by Scania CV.*

Description of failure

The harsh acceleration and braking cause frequent and abrupt changes in the operating condition of turbocharger components [98]. For the compressor wheel, this causes material fatigue and thus reduces the number of speed cycles it can withstand. In other words, the compressor wheel can fail due to LCF as shown in Fig. 4.49. The LCF failure of compressor wheel causes the compressor to draw less amount of air and hence the turbocharger provide an insufficient amount of compressed air to the combustion chamber of the engine. This leads to the complete immobilisation of heavy-trucks since the failure of turbocharger causes the engine to produce a much lower power output than required [130]. Thus, the failure of heavy-truck turbocharger is defined as follows:

Fig. 4.49 Image depicting a compressor wheel that has failed due to the low-cycle fatigue. The frequent and abrupt changes in turbocharger operating condition have weakened the material in the back face of the compressor wheel across several blades causing the material to separate from the wheel. The separated material has also caused significant damage to various other areas. Thus, the compressor wheel has failed well before its fatigue limit. *This image is provided by Scania CV.*

Failure definition. Frequent and abrupt changes in turbocharger operating condition (failure cause) initiate compressor wheel LCF (failure mechanism). The compressor wheel LCF eventually leads to an insufficient amount of compressed air to be produced by the turbocharger (failure mode) causing the internal combustion engine to produce a much lower power output than required (failure). Hereinafter, this failure definition is referred to as *the turbocharger failure*.

Prognostics modelling dataset

Scania heavy-truck turbocharger prognostics modelling dataset consists of condition monitoring data captured from 14440 Scania trucks currently operating in Europe. Similar to the APS dataset discussed in the previous case study, the turbocharger dataset also contains histogram variables. These variables are created by binning data captured from sensors that measure engine load, engine revolutions per minute (RPM), turbocharger thrust load, turbocharger turbine speed, atmospheric air pressure, boost air pressure and ambient temperature inside the turbocharger. Table 4.20 provides a description of these features and their relevance for modelling the turbocharger failure.

Limited failure data availability for Scania turbocharger predictive maintenance

The turbocharger failure in Scania heavy-trucks is rare due to two major reasons: (i) Scania heavy-trucks rarely operate in the critical regions of the engine load matrix as shown in Fig. 4.50. Hence, the operating condition of turbocharger components remains at optimal levels for the majority of the operational time; (ii) Scania turbocharger compressor wheels are highly reliable and thoroughly cycle tested prior to production release since its failure often leads to the complete immobilisation of heavy-trucks disrupting logistics and transportation missions. However, the rarity of Scania turbocharger failure reduces the number of degradation trajectories pertaining to the compressor wheel LCF, and hence causes failure data to be limited for developing prognostics models. Since limited failure data availability causes prognostics predictions to be associated with high uncertainty, a PdM policy that incorporates these predictions is ineffective (i.e. introduces costs due to under maintenance and over maintenance of turbocharger and false alarms) for Scania.

The limited turbocharger failure data availability problem is also evident in the prognostics modelling dataset used in this case study since the ratio between the number of failure data samples and normal data samples in the train set is 1500:110000, which means the failure data class only covers 1.36% of the entire set. The extent of the limited failure data availability problem for Scania turbocharger prognostics can be measured using the Shannon entropy-based method introduced in Sec. 3.2. The number of positive samples C_1 and negative samples C_2 are 1500 and 110000 respectively. The number of classes K is 2 (i.e. failure data class and normal data class). The normalised Shannon entropy H' of the dataset is therefore 0.1, which indicates a highly imbalanced dataset. Thus, the extent of the problem of limited failure data availability for prognostics is high. The following problem statement of the Scania turbocharger prognostics under the conditions of limited

Table 4.20 Features in Scania heavy-truck turbocharger prognostics modelling dataset.

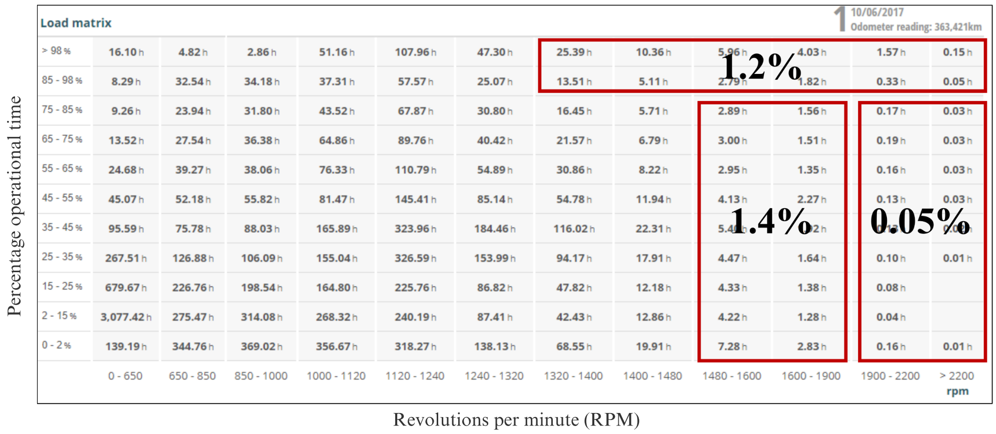
Feature	Description	Relevance for prognostics modelling
Engine load	Torque or brake horsepower produced by the engine.	Affected by the turbocharger failure and is a proxy for measuring the impact of turbocharger failure on the engine power output.
Engine RPM	Number of full rotations the crankshaft of the engine is making very minute.	Abrupt changes in RPM cause abrupt changes in turbocharger turbine rotational speed, which leads to high centrifugal stress in the turbocharger components at high temperature as well as high component temperatures in general.
Turbocharger thrust load	Pressure acting on compressor wheel and turbine wheel of the turbocharger.	Affected by frequent and abrupt changes in engine RPM.
Turbocharger turbine speed	The rotational speed of the turbocharger turbine wheel.	Affected by the compressor wheel LCF and is a proxy for measuring the compressor wheel speed.
Boost air pressure	Pressure of compressed air pushed by the turbocharger to the combustion chamber of the engine.	Affected by compressor wheel LCF as lower the number of compressor wheel speed cycles lower the performance of air intake and air compression.
Atmospheric air pressure	The pressure within the atmosphere of Earth.	Allows measuring the quality of compressed air produced by the turbocharger compared to the atmospheric air pressure at different altitudes.
Ambient temperature	The temperature of the air surrounding turbocharger components.	Abrupt changes in engine RPM and overall raise in temperature of components inside the turbocharger leads to compressor wheel material degradation, which causes compressor wheel LCF.

failure data availability problem follows from this analysis and aligns with the two research questions of this thesis:

Problem statement. Address the problem of limited failure data availability for Scania heavy-truck turbocharger prognostics, and quantify the impact of methodology on Scania heavy-truck turbocharger prognostics under the conditions of limited failure data availability compared to the existing techniques.

4.3.2 Methodology application and results

The Scania heavy-truck turbocharger prognostics problem satisfies the prerequisite of the methodology due to the following: the effect of harsh acceleration and braking on the compressor wheel LCF failure mechanism can be used as auxiliary information since they cause frequent and abrupt changes in the



Load matrix

1 10/06/2017 Odometer reading: 363,421km

Percentage operational time	0 - 650	650 - 850	850 - 1000	1000 - 1120	1120 - 1240	1240 - 1320	1320 - 1400	1400 - 1480	1480 - 1600	1600 - 1900	1900 - 2200	> 2200 rpm
> 98 %	16.10 h	4.82 h	2.86 h	51.16 h	107.96 h	47.30 h	25.39 h	10.36 h	5.96 h	4.03 h	1.57 h	0.15 h
85 - 98 %	8.29 h	32.54 h	34.18 h	37.31 h	57.57 h	25.07 h	13.51 h	5.11 h	2.79 h	1.82 h	0.33 h	0.05 h
75 - 85 %	9.26 h	23.94 h	31.80 h	43.52 h	67.87 h	30.80 h	16.45 h	5.71 h	2.89 h	1.56 h	0.17 h	0.03 h
65 - 75 %	13.52 h	27.54 h	36.38 h	64.86 h	89.76 h	40.42 h	21.57 h	6.79 h	3.00 h	1.51 h	0.19 h	0.03 h
55 - 65 %	24.68 h	39.27 h	38.06 h	76.33 h	110.79 h	54.89 h	30.86 h	8.22 h	2.95 h	1.35 h	0.16 h	0.03 h
45 - 55 %	45.07 h	52.18 h	55.82 h	81.47 h	145.41 h	85.14 h	54.78 h	11.94 h	4.13 h	2.27 h	0.13 h	0.03 h
35 - 45 %	95.59 h	75.78 h	88.03 h	165.89 h	323.96 h	184.46 h	116.02 h	22.31 h	[illegible]	[illegible]	[illegible]	[illegible]
25 - 35 %	267.51 h	126.88 h	106.09 h	155.04 h	326.59 h	153.99 h	94.17 h	17.91 h	4.47 h	1.64 h	0.10 h	0.01 h
15 - 25 %	679.67 h	226.76 h	198.54 h	164.80 h	225.76 h	86.82 h	47.82 h	12.18 h	4.33 h	1.38 h	0.08 h	
2 - 15 %	3,077.42 h	275.47 h	314.08 h	268.32 h	240.19 h	87.41 h	42.43 h	12.86 h	4.22 h	1.28 h	0.04 h	
0 - 2 %	139.19 h	344.76 h	369.02 h	356.67 h	318.27 h	138.13 h	68.55 h	19.91 h	7.28 h	2.83 h	0.16 h	0.01 h

1.2%

1.4%

0.05%

Revolutions per minute (RPM)

Fig. 4.50 Engine load matrix with the percentage time a Scania heavy-truck operating in the critical regions related to the compressor wheel LCF failure (marked with red and identified by Scania). It can be observed that the truck operates only 2.65% of the operational time in the critical regions which causes the compressor wheel LCF failure to be rare.

operating condition of turbocharger components, and thus affect the compressor wheel LCF [130]. The problem satisfies the assumption of the methodology due to the following: the failure mechanism that needs predicting is LCF and it causes equipment to fail under degradation, hence the failure is not random [130].

In the remainder of this section, a description of how the methodology is implemented using the effect of harsh use of equipment on failure mechanisms auxiliary information kind to generate plausible turbocharger failure data, and thus address the problem of limited failure data availability for Scania heavy-truck turbocharger prognostics is provided.

Phase 1: Auxiliary information processing

In this phase, suitable pieces of auxiliary information pertaining to the effect of harsh acceleration and braking on the LCF failure mechanism of turbocharger compressor wheel is identified, validated and converted into the vector form which allows constructing the conditional generator and discriminator of the CGAN.

(I) Auxiliary information identification

The model constructed for the turbocharger failure process is shown in Fig. 4.51. Using this model, expert knowledge pertaining to different stages in the turbocharger failure process is obtained from turbocharger component designers and maintenance engineers at Scania. The following auxiliary information is identified from this expert knowledge: *historically, majority of the Scania heavy-trucks that have failed due to the turbocharger compressor wheel LCF have spent more operational time in certain regions (i.e. critical regions shown in Fig. 4.50) of the engine load matrix compared to the trucks that have not failed due to the turbocharger compressor wheel LCF.*

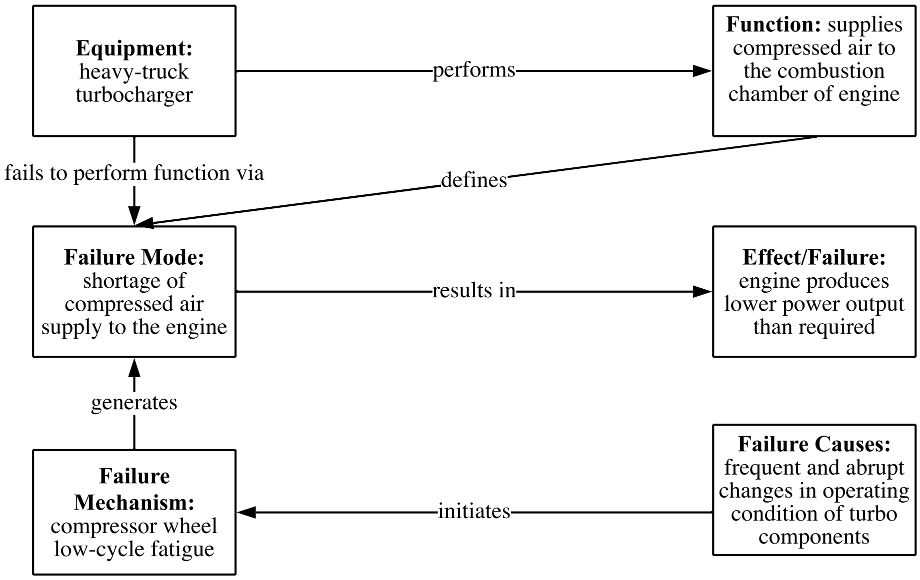


Fig. 4.51 Diagram depicting a model of the turbocharger failure process which is based on the turbocharger failure definition.

(II) Auxiliary information validation

In order to validate the auxiliary information, the average operational time a failed in the prognostics modelling dataset spent in critical regions of the engine load matrix is compared to the average operational time a healthy truck spent in the critical regions. It was identified that a failed truck spent 2.8% of the operational time in the critical regions whereas a healthy truck spent only 0.3% of the operational time in the critical regions.

In order to conduct a more robust experiment, the Welch's t-test is performed on the following null hypothesis: *the operational time a failed truck spent in critical regions of the engine load matrix is significantly higher than the operational time a healthy truck spent in the critical regions*. The rationale for this hypothesis is as follows: according to the auxiliary information identified from expert knowledge, Scania heavy-trucks that have failed due to the turbocharger failure have spent more operational time in critical regions of the engine load matrix than healthy trucks. The objective of Welch's t-test is to identify whether the corresponding null hypothesis can be rejected. The p-value obtained for the test is 0.67, which means there is weak evidence against the null hypothesis thus it is retained. Therefore, the following piece of auxiliary information which is based on the effect of harsh acceleration and braking on the compressor wheel LCF failure mechanism can be used in the methodology for generating plausible turbocharger failure data:

Auxiliary information. Scania heavy-trucks that have failed due to the turbocharger compressor wheel LCF have spent significantly more operational time in the critical regions of the engine load matrix than healthy trucks.

(III) Auxiliary information conversion

Constructing a vector of auxiliary information from class labels approach introduced in Sec. 3.3.2 is used in this case study to convert the validated auxiliary information. First, all the failed trucks are grouped into three risk groups depending on the operational time they spent in critical regions of the engine load matrix. More specifically, failed trucks operated between 0% to 1% of their operational time is given label 1, 1.1% to 2% is given label 2, and 2.1% to 3% is given label 3. The healthy trucks are given the reserved label 0 which refers to no-risk group. Then a new label column indicating to which risk group each truck belongs to is added to the Scania turbocharger prognostics modelling dataset as shown in Fig. 4.52. This means, the dataset now stores data that encode the validated auxiliary information. This column is also a vector representation of the auxiliary information: class labels that represent the no-risk group and the three risk groups for each heavy-truck with natural numbers 0,1,2 and 3 are collectively is a vector of natural numbers $Y = \{y \in \mathbb{N} | 0 \leq y \leq 3\}$.

Fig. 4.52 Diagram depicting the Scania heavy-truck turbocharger prognostics modelling dataset with the converted vector of auxiliary information pertaining to the effect of harsh use of equipment on failure mechanisms. The *Label* column corresponds to the healthy trucks (i.e. label 0) and failed trucks (i.e. label 1).

Phase 2: Conditional generative model estimation

In this phase, the vector of auxiliary information and the limited amount of turbocharger failure data samples available are integrated into the CGAN to estimate a generative model that is capable of generating plausible turbocharger failure data samples. The same approach used in the previous case study is followed to preprocess and structure the prognostics modelling dataset. Thus, the following five datasets are produced:

1. Stratified train set: the stratified train split of the turbocharger prognostics modelling dataset which also includes the vector of auxiliary information which encodes the effect of harsh acceleration and braking on the compressor wheel LCF.

2. Stratified validation set: the stratified validation split of the turbocharger prognostics modelling dataset which also includes the vector of auxiliary information which encodes the effect of harsh acceleration and braking on the compressor wheel LCF.

3. Stratified test set: the stratified test split of the turbocharger prognostics modelling dataset which also includes the vector of auxiliary information which encodes the effect of harsh acceleration and braking on the compressor wheel LCF.

4. Train failure data subset: a subset of stratified train set that only contains turbocharger failure data samples in the stratified train set.

5. Train non-failure data subset: a subset of stratified train set that only contains normal data samples in the stratified train set.

The conditional generative model estimation is similar to the previous case study apart from the following: the failure data samples input X is the aforementioned train failure data subset and the auxiliary information input Y is the vector of auxiliary information that encodes the effect of harsh acceleration and braking on the turbocharger compressor wheel LCF failure mechanism. Conditioning the generator and discriminator on auxiliary information pertaining to the turbocharger failure allows controlling and directing the plausible failure data generation process. Thus, the estimated generative model is able to generate new turbocharger failure data samples that conform to the following condition: *a newly generated turbocharger failure data sample should belongs to one of four risk groups*.

The CGAN is trained using the minibatch stochastic gradient descent algorithm presented in Algorithm 1. The hyperparameters used to train the CGAN in this case study are outlined in Table. 4.21.

Table 4.21 Hyperparameters used to train the CGAN for estimating a generative model that is capable of generating plausible turbocharger failure data.

Hyperparameter	Value
Optimiser	Adam
Learning rate	$2e-4$
Weight decay	$1e-4$
Exponential decay rate for first moment estimate	0.2
Exponential decay rate for second moment estimate	0.8
Number of batches	32
Number of epochs	60
Number of training iterations	2800

Fig. 4.53 shows the learning curves of CGAN which provide a visual indication of whether the convergence is achieved. The following can be observed from these learning curves: (i) discriminator loss on real and generated turbocharger failure data samples is stable around 1/2 which indicates that the discriminator cannot discriminate between real and generated turbocharger failure data samples;

(ii) generator loss has become stable as the number of training iterations increases which indicates the generator is consistent; (iii) variance of generator and discriminator loss is modest. As discussed in Sec. 3.3.3, these properties are expected from learning curves of a CGAN that has achieved an equilibrium between its generator and discriminator.

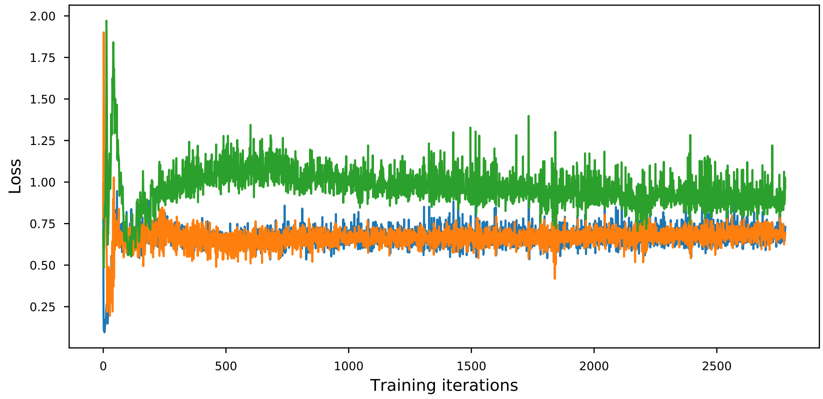


Fig. 4.53 The generator (G) and discriminator (D) learning curves of the CGAN trained for Scania heavy-truck turbocharger failure data generation.

Since the analysis of learning curves is only a qualitative evaluation, the K-S test is performed to conduct a more robust and quantitative evaluation. Table 4.22 outlines the K-S test results obtained for a set of features in the turbocharger prognostics modelling dataset. It can be observed that all the p-values produced by the K-S test are above 0.05 which indicates weak evidence against the null hypothesis. Thus, the null hypothesis cannot be rejected and therefore the CGAN has achieved the convergence. More specifically, the generative model trained in the CGAN, $G^*(Z \mid Y)$ can replicate the real turbocharger failure data distribution, and hence capable of generating plausible turbocharger failure data samples.

Table 4.22 K-S test results obtained for the set of features when statistically comparing the real turbocharger failure data distribution to the generated turbocharger failure data distribution.

Feature	K-S test p-value
Engine Load	0.92
Engine RPM	0.93
Turbocharger thrust load	0.87
Turbocharger turbine speed	0.91
Boost air pressure	0.78
Atmospheric air pressure	0.77
Ambient temperature	0.59

The MMD test outlined in Sec. 3.3.3 is employed to evaluate whether the convergence performance is affected by generator overfitting to real failure data samples in the train failure data subset. The p-value observed for this statistical test is 0.43 (> 0.05), which indicates weak evidence against the null hypothesis of the MMD test. This means the null hypothesis cannot be rejected, thus the generative model, $G^{*}(Z \mid Y)$ is not overfitting to real failure data. The convergence performance of the estimated generative model is therefore not affected by overfitting.

Phase 3: Plausible failure data generation

The generative model estimated by conditioning the failure data generation process on the effect of harsh acceleration and braking on the turbocharger compressor wheel LCF failure mechanism is used to generate plausible turbocharger failure data samples in this phase. The process followed for generating plausible failure data is similar to the one discussed in the previous example, however, uses the following input parameters: X is train failure data subset, Y is vector of auxiliary information that encodes the effect of harsh acceleration and braking on the turbocharger compressor wheel LCF failure mechanism, and Z is a noise vector.

In order to produce augmented train sets, samples from each plausible turbocharger failure dataset generated are added to a copy of the original train set (also see Fig. 3.12). The augmented train sets are used in the next phase for turbocharger prognostics modelling, and the set that produced the best prognostics performance is selected for developing a PdM model for Scania heavy-truck turbocharger maintenance under the conditions of limited failure data availability.

Phase 4: Prognostics modelling

In this phase, a suitable prognostics model for Scania heavy-truck turbocharger failure prediction is estimated using the augmented train sets and evaluated on the original test set.

(I) Prognostics modelling approach establishment

The binary classification-based approach is used in this case study to develop a prognostics model. The reason for this is that the prognostics modelling dataset is structured and labelled for the binary classification problem by Scania as their prognostics objective is to predict whether a data sample captured from the onboard condition monitoring system of a heavy-truck is pertaining to an imminent turbocharger failure (i.e. turbocharger failure within next 15 days) or not. The objective of Scania heavy-truck turbocharger prognostics modelling is therefore to predict the turbocharger failure probability within a prediction horizon of 15 days for the trucks in the test set using a probabilistic binary classifier, and then calculate the reliability of these trucks within the prediction horizon.

(II) Prognostics model estimation and evaluation

In the previous phase, multiple augmented train sets are produced and each of these has a different number of plausible failure data samples. These augmented train sets, and the unchanged validation and test sets are used in the process recommended by the methodology (see Fig. 3.16) to estimate a prognostics model for Scania turbocharger prognostics. The augmented train set that produced the best prognostics performance is the set containing 33000 plausible turbocharger failure data samples. In other words, when the data imbalance ratio is 34500:110000 compared to 1500:110000 in the original train set. More formally, the normalised Shannon entropy of the augmented train set with 33000 plausible turbocharger failure data samples and 1500 real turbocharger failure data samples is 0.8 compared to the 0.1 in the original train set. Thus, the 33000 plausible turbocharger failure data samples generated by the methodology reduced the extent of limited turbocharger failure data availability problem.

In the following, prognostics performance obtained when trained on the augmented train set and increase in prognostics performance provided by the methodology compared to the performance on original dataset are discussed:

(a) Prognostics performance on augmented train set

Table 4.23 summarises the prognostics performance obtained by different prognostics models when trained on the augmented train set and evaluated on the test set. It can be observed that the GBM and RF-based prognostics models obtained a Kappa statistic higher than 0.61, which means their prognostics predictions are associated with minimal uncertainty (also see Table 3.3).

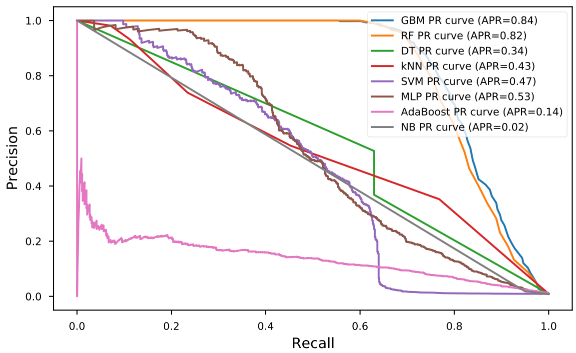
Table 4.23 Prognostics performance obtained by models when trained on the augmented train set and evaluated on the test set.

Prognostics model	Kappa statistic	Precision		Recall	
		Failure data class	Normal data class	Failure data class	Normal data class
GBM	0.72	0.9976	0.9959	0.5603	0.9999
RF	0.71	0.9999	0.9959	0.5509	0.9999
DT	0.57	0.5269	0.9966	0.63	0.9948
MLP	0.42	0.9266	0.9933	0.2707	0.9998
kNN	0.35	0.7383	0.993	0.2346	0.9993
SVM	0.12	0.9999	0.9915	0.0643	0.9999
AdaBoost	0.05	0.2857	0.9912	0.0294	0.9993
NB	0.01	0.0119	0.9994	0.9839	0.252

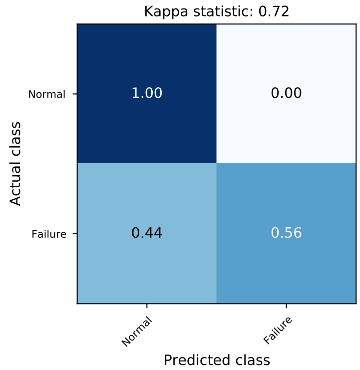
Table 4.23 also outlines the precision and recall obtained by the prognostics models. Apart from the NB-based model, all the other models have obtained a high precision and recall for the normal data class since there is a large number of data samples pertaining to the normal condition of the turbocharger (i.e. 110000 data samples). The precision-recall curves of all the prognostics models are shown in Fig. 4.54a. The GBM-based prognostics model has obtained the best overall performance since it has the highest Kappa statistic (i.e. prognostics predictions are associated with minimal uncertainty), comparable precision to the RF-based model and more importantly, a higher recall for the failure data class which means this model causes the lowest number of undetected failures. The confusion matrix obtained by the GBM-based prognostics model is shown in Fig. 4.54b.

(b) Prognostics performance compared to original train set

The impact of the methodology on Scania heavy-truck turbocharger prognostics can be observed from Fig. 4.55, Fig. 4.56, Fig. 4.57. The figures compare Kappa statistic, precision and recall obtained by the prognostics models when trained on the original and augmented train sets.



(a) Precision-recall curves of all prognostics models.



(b) GBM confusion matrix.

Fig. 4.54 Prognostics performance obtained by all the prognostics models and the confusion matrix obtained by the best model (i.e. GBM-based model). The Average Precision-Recall (APR) is used as the metric to quantify the prognostics performance shown in the precision-recall curves.

Compared to when trained on the original train set, the Kappa statistic and the recall of failure data class are significantly higher when trained on the augmented train set. More specifically, the model has obtained a 60% increase in Kappa statistic (i.e. reduced the uncertainty associated with prognostics predictions) and 93% increase in recall on the failure data class (i.e. reduced the number of undetected failures) when trained on the augmented train set. This is since, in addition to the real failure data samples, the augmented train set includes plausible failure data samples which are generated by conditioning the failure data generation process on the effect of harsh use of equipment on failure mechanisms auxiliary information kind.

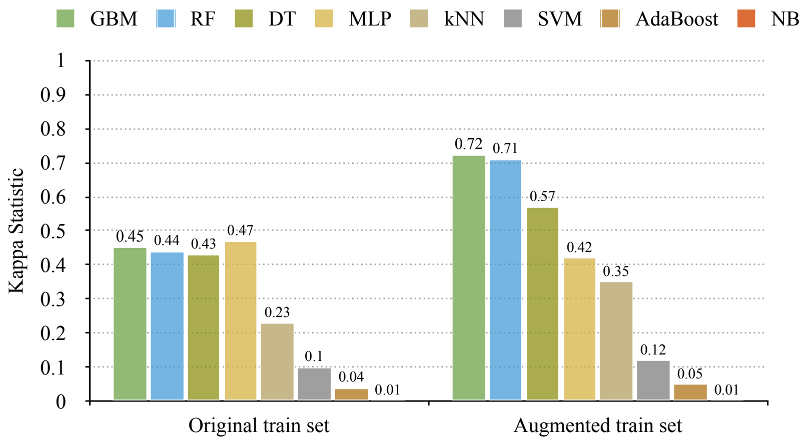
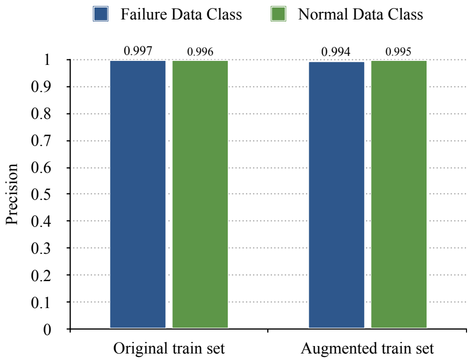
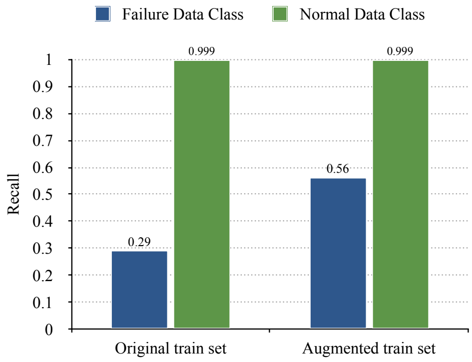


Fig. 4.55 Kappa statistic obtained by prognostics models when trained on the original and augmented train sets. The models are evaluated on the original test set.

The confusion matrixes of the GBM-based prognostics model are shown in Fig. 4.58. It can be observed that when trained on the augmented train set, the model produces 93% more failure detections compared to when trained on the original train set.

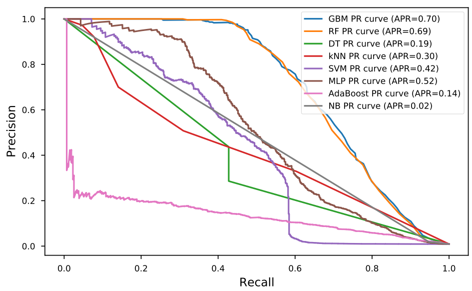


(a) Precision (higher the precision lower the number of false alarms)

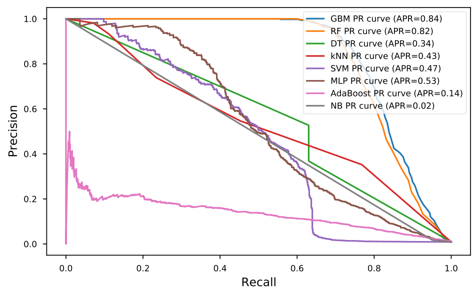


(b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.56 Prognostics performance obtained by the GBM-based prognostics model when trained on the original and augmented train sets. The models are evaluated on the original test set.

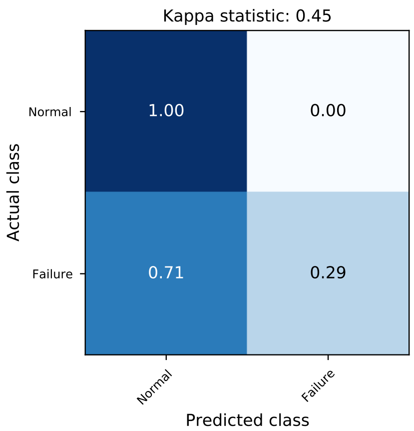


(a) Original train set

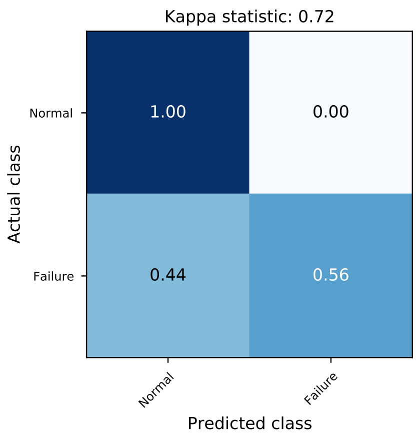


(b) Augmented train set

Fig. 4.57 Precision-recall curves obtained by the prognostics models when trained on the original and augmented train sets. The models are evaluated on the original test set.



(a) Original train set



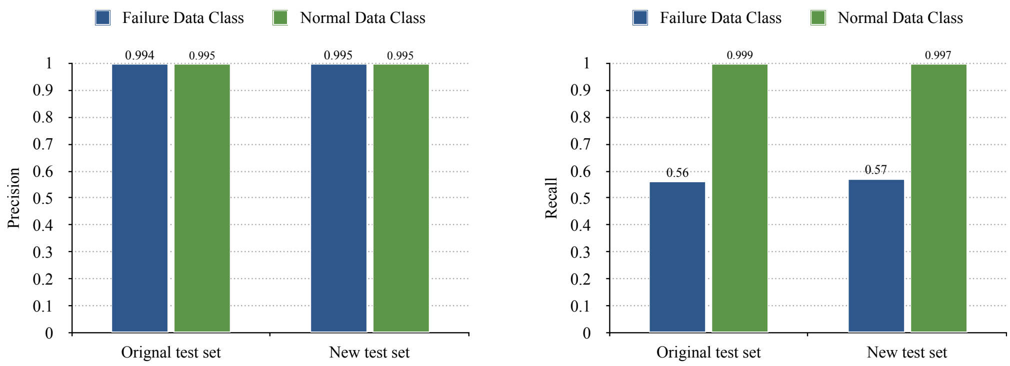
(b) Augmented train set

Fig. 4.58 Confusion matrixes obtained by the GBM-based prognostics model when trained on the original and augmented train sets. The models are evaluated on the original test set.

Validation of prognostics performance

Similar to the previous two case studies, the results discussed in this case study are also obtained using a test set that is not previously seen by the prognostics models. Nevertheless, in this case study also it is worth further validating the prognostics model performance and its consistency by collecting another test set from the industrial scenario. To this end, after the models were developed and evaluated on the original test set, another test set is collected from the Scania heavy-trucks studied in this case study.

Fig. 4.59 provides a comparison of prognostics performance obtained for the original and newly collected test sets. It can be observed that the prognostics performance remains consistent across both test sets, which demonstrates the validity of the prognostics model and the consistency of prognostics performance.



(a) Precision (b) Recall

Fig. 4.59 Comparison of prognostics performance obtained by the GBM-based Scania turbocharger prognostics model when evaluated on the original test set and new test set.

Summary of methodology application using the effect of harsh use of equipment

The problem statement of this case study stated in Sec. 4.3.1, which aligns with the two research questions of this thesis involves addressing the problem of limited failure data availability for prognostics and quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques.

In this case study, the methodology which addresses the first research question of this thesis is applied to generate plausible failure data samples using the effect of harsh use of equipment on failure mechanisms auxiliary information kind, and thus address the problem of limited failure data availability for prognostics in practice. It is shown that the methodology is capable of addressing this problem by producing a prognostics model that is capable of estimating accurate prognostics predictions with minimal uncertainty. Thus, this case study has shown that the application of the methodology using the effect of harsh use of equipment on failure mechanisms auxiliary information kind addresses the problem of limited failure data availability for prognostics in practice.

4.3.3 Analysis of prognostics impact and key factors influencing effectiveness

In this section, insights are produced using the Scania heavy-truck turbocharger prognostics case study to quantify the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques, and to identify the key factors influencing the effectiveness of methodology when applied using the effect of harsh use of equipment on failure mechanisms auxiliary information kind.

Impact on prognostics under the conditions of limited failure data availability

The GBM algorithm is used to model prognostics as it produced the best prognostics performance. Fig. 4.60 provides a comparison of the uncertainty associated with prognostics predictions produced by the methodology and existing techniques. Similar to the analyses performed in previous case studies, it can be observed that the methodology provides a significant reduction in uncertainty compared to the existing techniques when applied using the effect of harsh use of equipment on failure mechanisms auxiliary information kind. More specifically, the uncertainty reductions outlined in Table 4.24 are provided by the methodology compared to the existing techniques.

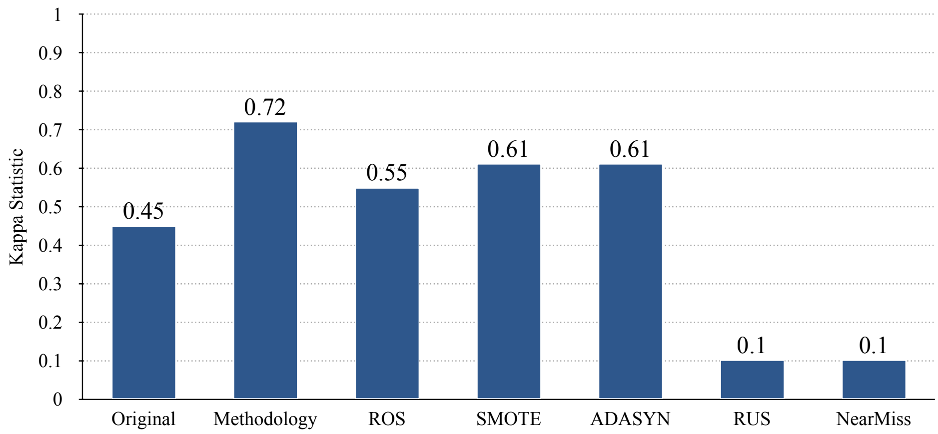
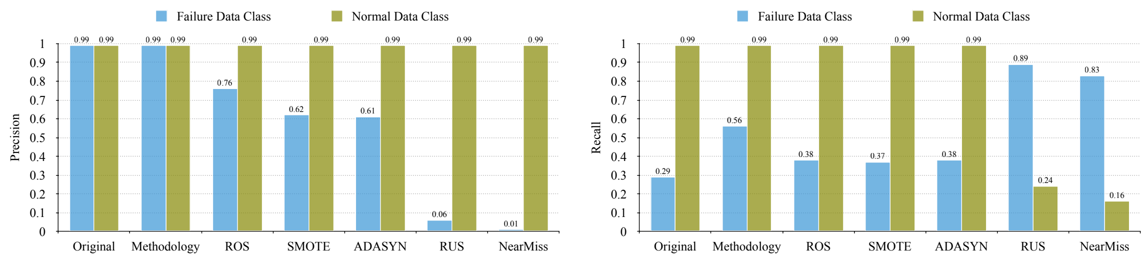


Fig. 4.60 Comparison of uncertainty associated with prognostics predictions produced by the methodology and existing techniques. The uncertainty produced on the original dataset is denoted as Original.

Table 4.24 Reduction in uncertainty (i.e. increase in Kappa statistic) provided by the methodology compared to the existing techniques.

Reduction in uncertainty				
ROS	SMOTE	ADASYN	RUS	NearMiss
31%	18%	18%	620%	620%

In Fig. 4.61, the prognostics performance produced by the methodology is compared to the existing techniques (also see Table 4.25). It can be observed that the methodology allows predicting a higher number of failures (i.e. has a higher recall on failure data class) and causes a lower number of false alarms (i.e. has a higher precision on failure data class) compared to the oversampling techniques (i.e. ROS, SMOTE and ADASYN). The undersampling techniques (i.e. RUS and NearMiss) have obtained a higher recall on the failure data class since they caused the prognostics model to overfit to the failure data class after undersampling. This is evident from the poor performance obtained by these techniques for the precision on the failure data class (see Fig. 4.61a). This means undersampling techniques caused the prognostics model to predict samples pertaining to the normal and failure conditions as samples pertaining to the failure condition, which leads to a large number of false alarms.



(a) Precision (higher the precision lower the number of false alarms)

(b) Recall (higher the recall lower the number of undetected failures)

Fig. 4.61 Comparison of prognostics performance produced by methodology and existing techniques.

Table 4.25 Decrease in the number of false alarms (i.e. increase in precision) and decrease in the number of undetected failures (i.e. increase in recall) provided by the methodology for turbocharger prognostics compared to the existing techniques.

	Failure data class	
	Precision	Recall
ROS	30%	47%
SMOTE	60%	51%
ADASYN	62%	47%
RUS	1550%	-37%
NearMiss	9800%	-33%

As discussed in the analyses performed in previous case studies, the reason for the significant increase in prognostics performance is since the methodology is capable of addressing the problem of limited failure data availability by generating new and plausible failure data samples. In contrast, as discussed in the literature review chapter, the existing techniques either duplicate existing failure data or randomly generate data (oversampling techniques) or do not increase the number of failure data samples available for prognostics modelling (undersampling techniques). Hence, they do not address the fundamental problem of limited failure data availability for prognostics. The following insight can be gained from this analysis:

Insight. When the methodology is applied using the effect of harsh use of equipment on failure mechanisms auxiliary information kind for prognostics under the conditions of limited failure data availability, it produces effective prognostics predictions and outperforms existing techniques used in the literature by a large margin.

Key factors influencing the effectiveness of methodology

In this section, an analysis is performed from a practical perspective to identify the influence of real failure data availability and auxiliary information availability factors on the effectiveness of methodology in practice (i.e. the influence on the prognostics performance) when it is applied using the effect of harsh use of equipment on failure mechanisms auxiliary information kind. The scenarios used in this analysis are similar to the ones used in the previous case studies and are described in Table 4.26.

Table 4.26 Description of scenarios considered in the analysis.

Scenario name	Parameters	Variation
Base scenario	X is failure data in Turbo_1 (see Fig. 4.62), Y is the vector encoding the effect of harsh acceleration and braking on compressor wheel LCF and is given, $Z \sim N(0,1)$ and is fixed	
Scenario 1		$X \rightarrow$ failure data in Turbo_2
Scenario 2		$X \rightarrow$ failure data in Turbo_3
Scenario 3		$X \rightarrow$ failure data in Turbo_4
Scenario 4		$X \rightarrow$ failure data in Turbo_5
Scenario 5		$X \rightarrow$ failure data in Turbo_6
Scenario 6		Y is given $\rightarrow Y$ is not given (not conditioned)
Scenario 7		Y is auxiliary $\rightarrow Y$ is arbitrary

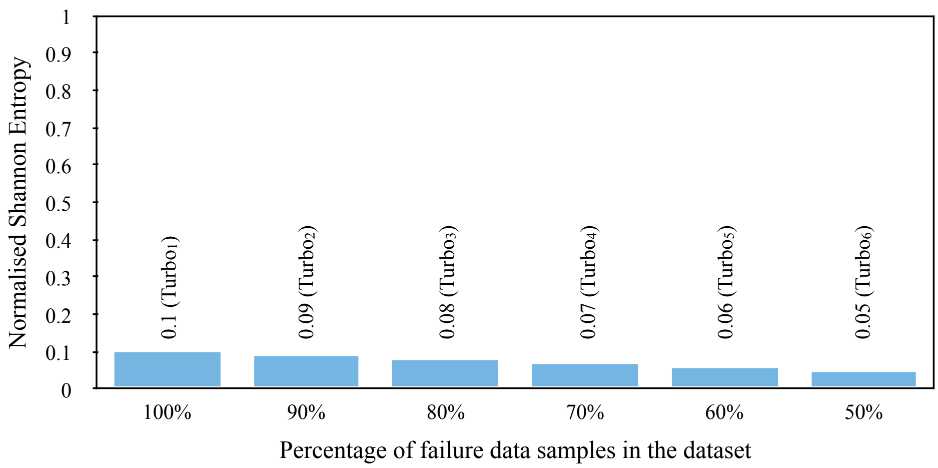


Fig. 4.62 Plot depicting how the extent of limited failure data availability problem is increased for analysing the effect of real failure data samples. The failure data are incrementally removed from the turbocharger dataset and resulting datasets are denoted Turbo_1, Turbo_2, Turbo_3 and so on.

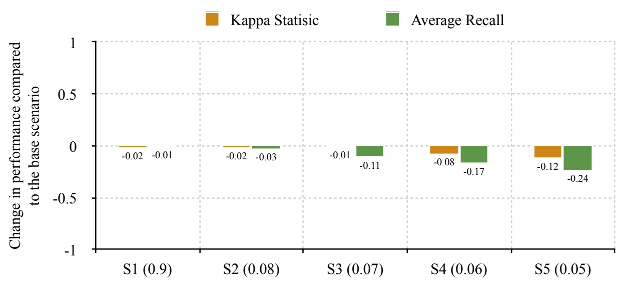
(I) Effect of real failure data availability on prognostics performance

The extent of the limited failure data availability problem in the turbocharger prognostics modelling dataset is high (i.e. has a normalised Shannon entropy of 0.1). As shown in the previous analysis, the methodology produced effective prognostics predictions using this dataset and outperformed the existing techniques. In this analysis, the extent of the problem is further increased to identify the effect of real failure data availability on the prognostics performance provided by the methodology.

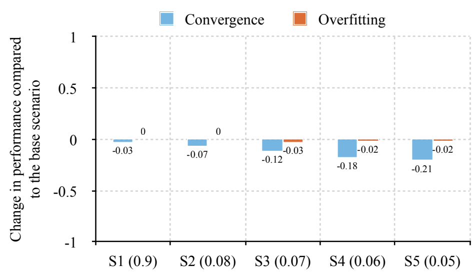
The effect on prognostics performance when scenarios 1 to 5 are applied to turbocharger prognostics modelling is shown in Fig. 4.63. It can be observed that the amount of real failure data samples has an effect on prognostics performance produced by the methodology and when the real failure data availability is getting reduced, the prognostics performance starts to degrade rapidly (see scenarios 3 to 5). The reason for this can be observed from the convergence performance which shows that the convergence of the CGAN model starts to degrade rapidly for scenarios 3 to 5.

Fig. 4.64 shows the effect of real failure data availability on prognostics performance produced by the methodology compared to the existing techniques. It can be observed that whilst the methodology outperforms existing techniques in achieving a low uncertainty and a lower number of undetected failures for scenarios 1 to 3, its Kappa statistic value and recall starts to degrade rapidly for scenarios 4 and 5. The following insight which is also produced in the previous case studies can be gained from this analysis:

Insight. In practice, when the availability of real failure data samples is approximately equal to or less than 0.1 normalised Shannon entropy, the prognostics performance of the methodology starts to become less effective and eventually becomes comparable to the existing techniques as the availability of real failure data is reduced.



(a) Prognostics performance

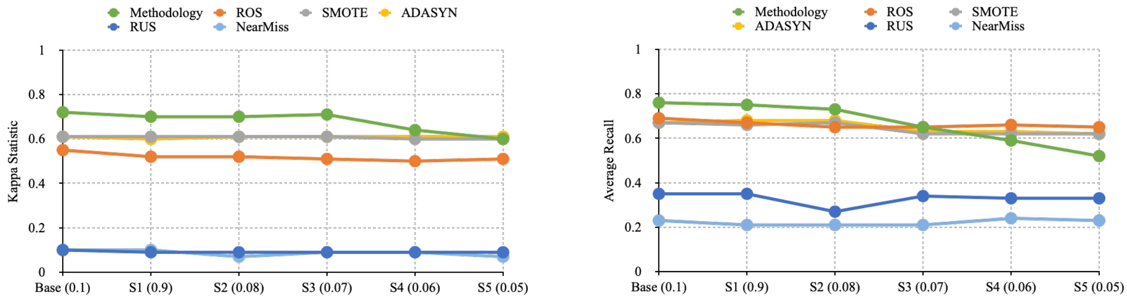


(b) Convergence performance

Fig. 4.63 Turbocharger convergence and prognostics performances produced by the methodology for scenarios (denoted S) 1 to 5 compared to the base scenario. The x-axis shows the scenario and corresponding extent of limited failure data availability (i.e. the normalised Shannon entropy).

(II) Effect of auxiliary information availability on prognostics performance

It can be observed from Fig. 4.65, the prognostics performance produced by the methodology significantly reduces when auxiliary information pertaining to failure mechanisms is not integrated for controlling and directing the plausible failure data generation process. The effect of auxiliary

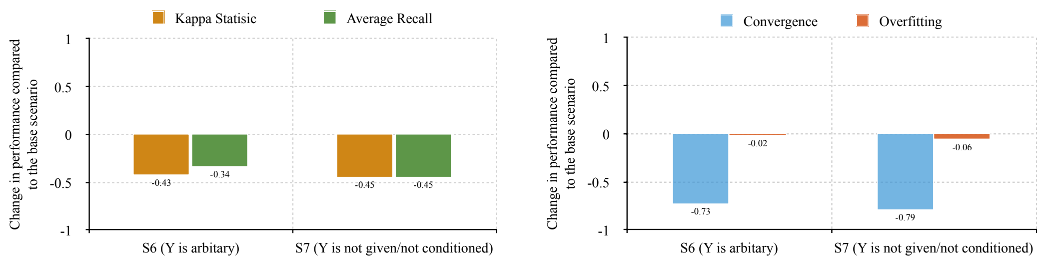


(a) Uncertainty (b) Recall

Fig. 4.64 Turbocharger prognostics performance comparison for scenarios 1 to 5. Higher the precision lower the number of false alarms, and higher the recall lower the number of undetected failures.

information availability on prognostics performance compared to the existing techniques is shown in Fig. 4.66. It can be observed that the prognostics performance produced by the methodology for Scenario 6 and 7 is significantly low compared to the base scenario and existing oversampling techniques. The following insight which is also produced in the previous case studies and theoretical analysis can be gained from this analysis:

Insight. The integration of auxiliary information pertaining to failure mechanisms which allows controlling and directing the plausible failure data generation process is crucial to the methodology for producing effective prognostics under the conditions of limited failure data availability.



(a) Prognostics performance (b) Convergence performance

Fig. 4.65 Turbocharger convergence and prognostics performances produced by the methodology for scenarios 6 and 7 compared to the base scenario.

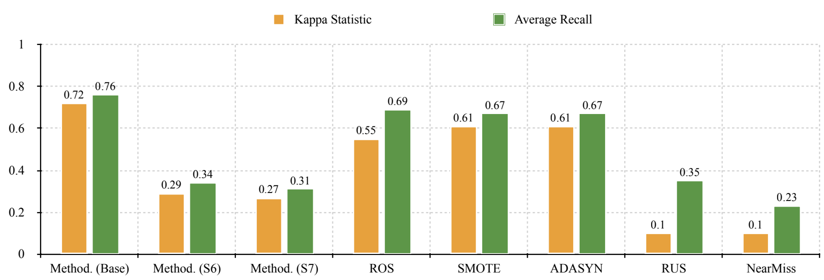


Fig. 4.66 Turbocharger prognostics performance produced by the methodology for scenarios 6 and 7 compared to the base scenario and existing techniques.

4.4 Conclusion

In this chapter, how the methodology for prognostics under the conditions of limited failure data availability can be implemented in practice was demonstrated. To this end, three case studies were provided and they showed how to implement the methodology using the effect of environmental conditions on failure mechanisms, effect of harsh use of equipment on failure mechanisms and similarity between equipment that has failed under a single failure mode auxiliary information kinds. The case studies were developed using real-world industrial scenarios involving BT residential broadband line prognostics and Scania heavy-truck component prognostics.

In this chapter, the second research question of this thesis was answered by quantifying the impact of methodology on prognostics under the conditions of limited failure data availability compared to the existing techniques, and by providing insights into the key factors influencing the effectiveness of methodology. In the next chapter, these insights are used to provide recommendations for applying the methodology in the industry.

Chapter 5

Recommendations for application in industry

This chapter provides recommendations for applying the methodology for prognostics under the conditions of limited failure data availability in the industry. To this end, the strengths and weaknesses of the methodology from theoretical and practical perspectives are outlined first. Then risks associated with the industrial application of the methodology and potential risk mitigation actions are discussed. The chapter is concluded by providing application criteria that allow identifying suitable and unsuitable industrial applications for the methodology. The criteria is developed using insights gained from the theoretical analysis in Chapter 3 and analyses performed using the real-world case studies in Chapter 4.

5.1 Strengths and weaknesses

In this section, strengths and weaknesses of the methodology considering the effective implementation of PdM are discussed. The methodology has the following strengths and weaknesses:

Strength 1: *Addresses the problem of limited failure data availability for prognostics by producing new and plausible failure data.* The theoretical results of the methodology proved that when conditioned on auxiliary information pertaining to failure mechanisms, the methodology is capable of converging the generated failure data distribution to the real failure data distribution, and hence capable of generating new and plausible failure data samples. Then in theoretical analysis and analyses performed from a practical perspective using the case studies showed that when prognostics models are trained on the augmented train set which includes plausible failure data generated by the methodology, the prognostics performance is significantly high compared to the existing techniques used to address the problem of limited failure data availability for prognostics.

This strength of the methodology is useful for the effective implementation of PdM in the industry since the ability to predict the probability that equipment operates without failure up to some future time and/or the TTF of equipment effectively enables developing an optimal PdM policy that balances the cost of maintenance, the risk of failure and the performance of the equipment.

Strength 2: *Integrates expert knowledge to enable the use of the data-driven approach for prognostics modelling under the conditions of limited failure data availability.* The methodology allows generating plausible failure data so that training datasets used for estimating data-driven prognostics models can be augmented to include an increased number of failure data samples, and thus develop effective data-driven models for prognostics under the conditions of limited failure data availability. As shown in the theoretical analysis and analyses performed from a practical perspective in case studies, the key to plausible failure data generation is the integration of auxiliary information pertaining to failure mechanisms (i.e. sound engineering knowledge of failure mechanisms), which allows controlling and directing the failure data generation process. In the methodology, the auxiliary information is identified from the expert knowledge of personnel such as reliability and maintenance engineers, component designers, equipment operators and applied research scientists.

This strength of the methodology enables the effective implementation of PdM by facilitating the use of the data-driven approach to prognostics modelling which has been highly successful for prognostics modelling when there is a sufficient amount of failure data samples is available.

Strength 3: *Has a robust theoretical foundation.* This thesis presented the theoretical foundation of the methodology which is an extension of the well-established theoretical foundation of the two-player minimax game [69]. The extension involved incorporating the joint probability distribution to the theoretical model of the two-player minimax game so that the value function of the minimax game includes the conditional generator $G(Z \mid Y)$ and the conditional discriminator $D(X \mid Y)$, where Z is random noise, X is real failure data and Y is the vector of auxiliary information pertaining to failure mechanisms. This extended theoretical foundation is mathematically proved in this thesis in Sec. 3.3.3. Moreover, the methodology is analysed from a theoretical perspective in Sec. 3.4 using a theoretical model of a system and failure to produce insights into the behaviour and working mechanism of the methodology.

This strength of the methodology is useful for the effective implementation of PdM since compared to the existing techniques which lack a theoretical basis, the methodology allows developing a rationale for prognostics and PdM model performance, and thus have confidence in prognostics and PdM predictions when applied in practice.

Strength 4: *Outperforms existing techniques used for prognostics under the conditions of limited failure data availability by a large margin.* The insights gained from case studies presented in this thesis showed that the methodology outperforms existing techniques used for prognostics under the conditions of limited failure data availability by a large margin. This is since compared to the existing techniques which either duplicate existing failure data or randomly generate data, the methodology generates new and plausible failure data by conditioning the failure data generation process on auxiliary information pertaining to failure mechanisms.

This strength of the methodology addresses the problem of prognostics under the conditions of limited failure data availability which has been the bottleneck for the effective implementation of PdM in the industry.

Strength 5: *Easily generalises into different kinds of industrial applications.* The methodology consists of four phases: (i) auxiliary information processing; (ii) conditional generative model estimation; (iii) plausible failure data generation; (iv) prognostics modelling. The latter three phases of the methodology are generalisable into different kinds of industrial applications since the only input parameter that depends on the industrial application in these phases is the historical condition monitoring and/or event dataset. Even though the auxiliary information processing phase depends on the kind of auxiliary information available, the methodology supports the following three kinds of auxiliary information which are widely available in practice through expert knowledge: (i) effect of environmental conditions on failure mechanisms; (ii) effect of harsh use of equipment on failure mechanisms; (iii) similarity between equipment that has failed under a single failure mode.

This strength of the methodology allows it to be applied to a wide range of industrial applications which enables the effective implementation of PdM for most industrial organisations.

Weakness 1: *Prognostics performance starts to degrade when the normalised Shannon entropy of the original dataset is approximately equal to or less than 0.1.* Chapter 4 showed that the methodology outperforms existing techniques by a large margin under the conditions of limited failure data availability. However, as shown in the theoretical analysis and analyses performed in the case studies, when the normalised Shannon entropy is approximately equal to or less than 0.1, the convergence performance of the CGAN model implemented for the methodology starts to degrade. This subsequently causes the improved prognostics performance produced by the methodology to degrade and eventually become comparable to the performance provided by the existing techniques.

Hence, it is recommended that the application of methodology should be avoided in industrial scenarios in which the extent of limited failure data availability problem for prognostics is approximately equal to or less than 0.1 normalised Shannon entropy. This is since the methodology requires more effort to implement compared to the existing techniques as discussed below. However as shown in the analyses performed in case studies, when the normalised Shannon entropy is approximately equal to or less than 0.1, the use of existing techniques should also be avoided since the prognostics performance produced by them are associated with high uncertainty, a large number of undetected failures and false alarms, which leads to the ineffective implementation of PdM in practice.

Weakness 2: *More effort is required to implement compared to the existing techniques.* From the implementation perspective, the majority of existing techniques proposed in the literature (e.g. random oversampling, SMOTE, ADASYN, random undersampling and NearMiss) can simply be part of a data preprocessing pipeline, which allows them to be implemented with less effort. However, the methodology proposed in this thesis is not a “plug and play” solution to the problem of prognostics under the conditions of limited failure data availability. The reason for this is that it requires not only the condition monitoring and/or event data but also the auxiliary information pertaining to failure mechanisms. Auxiliary information needs to be identified from expert knowledge obtained from maintenance and reliability engineers, component designers, equipment operators and applied research scientists. Then it needs to be validated and converted into a suitable form for integrating into the plausible failure data generation process. Hence, the methodology requires more effort to implement than existing techniques.

However, the prognostics and PdM cost benefits produced by the methodology compensate for the increase in effort required for its implementation. More importantly, once implemented, the methodology enables the effective implementation of PdM under the conditions of limited failure data availability for prognostics, which is a problem existing techniques have so far failed to address.

5.2 Risks and potential mitigation actions

In this section, the risks associated with the application of methodology and potential risk mitigation actions are discussed.

Risk 1: *Methodology can cause prognostics models to overfit.* Since overall, there is a limited amount of failure data samples available to start with, it could be possible in some applications the plausible failure data generated by the methodology represent only a few degradation patterns pertaining to the failure mode that needs predicting. The prognostics models estimated on this data could therefore overfit to certain degradation patterns, and hence not predict the failure effectively in practice.

In order to mitigate this risk, the incremental learning method can be used [131]. Incremental learning allows data-driven models to be trained continuously or periodically on new data so that the model adapts (i.e. updates model weights and biases) to the changes in the domain [131]. When implementing the methodology for an industrial application, incremental learning can be considered as part of the data-driven model training workflow design. Alternatively, if the workflow is developed using a cloud service provider, out-of-the-box incremental learning modules can be used (e.g. Amazon Web Services SageMaker Incremental Learning [132] and

Microsoft AzureML Data Factory [133]. Incorporating incremental learning into the data-driven model training workflow allows the methodology to re-train its CGAN model as new condition monitoring and/or event data is fed in, and thus learn to generate plausible failure data samples that represent a wider range of degradation patterns pertaining to the failure mode that needs predicting. This allows the prognostics models trained on the generated failure data to improve their generalisation capability, and hence predict the failure effectively in practice.

Risk 2: *Changing auxiliary information can cause implausible failure data to be generated.* In some industrial applications, the possibility of changing auxiliary information might need to be taken into account. For example, in the case of the effect of harsh use of equipment on failure mechanisms auxiliary information kind, it could be possible that the heavy-truck drivers are no longer performing harsh acceleration and harsh braking due to improved driver training programs. Similarly, components might not get affected by harsh acceleration and harsh braking anymore due to the advances in component designs that help produce components that can withstand harsh acceleration and braking. If not accounted, this change in auxiliary information will cause the methodology to generate implausible failure data since the failure data generation process is now conditioned on auxiliary information not pertaining to the failure mode that needs predicting (i.e. arbitrary auxiliary information). As shown in the theoretical analysis and analyses performed in case studies, conditioning the plausible failure data generation process on arbitrary auxiliary information significantly degrades the convergence and prognostics performances of the methodology.

This risk can be mitigated by introducing processes that combine human intervention and automation. For example, it is possible to automate the auxiliary information validation step by integrating it into the aforementioned data-driven model training workflows, so that the validity of auxiliary information is continuously or periodically measured by a system and observed by human experts. When the validity of auxiliary information compared to the latest failure instances starts to degrade, the auxiliary information processing phase of the methodology can be repeated to update the vectors of auxiliary information used in the conditional generative model estimation phase of the methodology.

5.3 Application criteria

In this section, application criteria that allow identifying suitable and unsuitable industrial applications for the methodology is presented. The criteria are developed using the insights gained from the analysis performed from a theoretical perspective and a practical perspective in Chapter 3 and Chapter 4 respectively. The recommended application criteria of the methodology for prognostics under the conditions of limited failure data availability are as follows:

Criterion 1: *Is the extent of limited failure data availability supported?* The improved prognostics performance produced by the methodology starts to degrade when the normalised Shannon entropy is 0.1 or below. This is since the convergence performance of the CGAN starts to degrade and therefore causes implausible failure data to be generated as shown in the theoretical analysis and analyses performed in the case studies. Hence, the methodology is most suitable to be applied when the extent of the limited failure data availability problem is above 0.1 normalised Shannon entropy.

Criterion 2: *Is auxiliary information pertaining to the failure mode available?* The methodology requires at least one of the following three kinds of auxiliary information to be available in the industrial scenario in addition to the condition monitoring and/or event data which are limited: (i) effect of environmental conditions of failure mechanisms; (ii) effect of harsh use of equipment on failure mechanisms; (iii) similarity between equipment that has failed under a single failure mode. This is since the methodology uses auxiliary information to control and direct the failure data generation process to generate plausible failure data samples.

Criterion 3: *Can the auxiliary information be verified?* Before using the available auxiliary information for generating plausible failure data it needs to be verified. More specifically, whether the available auxiliary information is suitable for generating plausible failure data needs to be identified statistically using the statistical test framework.

Criterion 4: *Can the auxiliary information be converted?* The methodology provides two approaches for converting auxiliary information into the vector form: (i) constructing an auxiliary information vector from a continuous distribution; (ii) constructing an auxiliary information vector from class labels. It could be possible that certain forms of available auxiliary information cannot be converted into the vector form using these approaches or are not suitable for converting into the vector form.

5.4 Conclusion

This chapter provided recommendations for applying the methodology for prognostics under the conditions of limited failure data availability in the industry. These recommendations are based on the insights gained from the theoretical analysis in Chapter 3 and analyses performed using the real-world case studies in Chapter 4. The recommendations include strengths and weaknesses of the methodology and their impact on industrial application, risks and potential mitigation actions, and application criteria that allow identifying suitable and unsuitable industrial applications for the methodology.

Chapter 6

Conclusion

In this chapter, a summary of the research presented in this thesis and its novelty are discussed first. A recap of the research questions, contributions of the research to the academic knowledge and industrial practice, awards and recognition the research has received in academia and industry are discussed next. The chapter is concluded by outlining recommendations for future research.

6.1 Research summary and novelty

The literature review of this thesis showed that a systematic approach for prognostics modelling under the conditions of limited failure availability is a research gap. Further, the industry review of the thesis showed that this research gap is the bottleneck for the effective implementation of PdM for industrial organisations. Therefore, the importance of a solution that addresses the research gap is that it enables the effective implementation of PdM, and hence realise the vision of an optimal maintenance policy for industrial organisations. In other words, when this research gap is addressed, industrial organisations will be able to prevent costs due to under-maintenance and over-maintenance of equipment and false alarms.

This thesis was therefore dedicated to addressing the problem of prognostics under the conditions of limited failure data availability. To address this problem, the thesis presented a methodology that allows developing prognostics models under the conditions of limited failure data availability. The methodology integrated sound engineering knowledge of failure mechanisms as auxiliary information, the limited amount of failure data samples available and noise to estimate a generative model in a minimax game. The estimated generative model was then used to produce new and plausible failure data samples, which allows augmenting historical datasets used for prognostics modelling to include an increased amount of failure data samples. Using real-world case studies, the thesis showed that the plausible failure data generated by the methodology allows developing effective prognostics models under the conditions of limited failure data availability.

The analyses performed in this thesis showed that the methodology outperforms existing techniques proposed in the literature. The reason for this is that compared to the existing techniques which duplicate existing failure data or randomly generate data, the methodology is capable of generating new and plausible failure data samples. Thus, the methodology allows addressing the fundamental problem of limited failure data availability for prognostics. The methodology is capable of generating new and plausible failure data is due to the conditioning of failure data generation process on the sound engineering knowledge of failure mechanisms. That is, the integration of the expert knowledge pertaining to the effect of environmental conditions on failure mechanisms, the effect of harsh use of equipment on failure mechanisms, and similarity between equipment that has failed under a single failure mode. The novelty of the research presented in this thesis also lies in this aspect of the methodology. More specifically, the novelty of the research is as follows:

Research novelty. The systematic identification, conversion and integration of sound engineering knowledge of failure mechanisms to estimate a generative model that is capable of generating new and plausible failure data samples for prognostics under the conditions of limited failure data availability.

6.2 Recap of research questions

In this section, a recap of the two research questions answered in this thesis is provided.

1. How can equipment failure be predicted under the conditions of limited failure data availability?

   This research question is answered in Chapter 3 by providing a methodology for prognostics under the conditions of limited failure data availability. The first part of the chapter provided a detailed description of the prerequisite, assumption and the four phases of the methodology: auxiliary information processing, conditional generative model estimation, plausible failure data generation and prognostics modelling. The theoretical results of the integral part of the methodology, that is, the conditional generative model estimation using auxiliary information pertaining to failure mechanisms is also provided with the methodology description. The second part of the chapter discussed insights produced from a theoretical analysis that aimed to identify the behaviour and working mechanism of the methodology. In this analysis, the methodology is evaluated in a theoretical setting which involved generating plausible failure data for a theoretical model of a vehicle transmission system with a gear tooth fault. The insights produced from this analysis is used to discuss the behaviour and working mechanism of the methodology, and thus develop a rationale for its capabilities from a theoretical perspective.

2. What impact does the proposed solution have on prognostics compared to existing techniques under the conditions of limited failure data availability?

   This research question is answered in Chapter 4 by quantifying the impact of methodology on prognostics conditions of limited failure data availability compared to the existing techniques proposed in the literature, and by producing insights into the key factors influencing the effectiveness of the methodology. The chapter provided three case studies of the methodology in real-world industrial scenarios. The BT residential broadband line prognostics case study showed how to apply the methodology using the effect of environmental conditions on failure mechanisms auxiliary information kind. The Scania air processing system and turbocharger prognostics case studies showed how to apply the methodology using the similarity between equipment that has failed under a single failure mode and the effect of harsh use of equipment on failure mechanisms auxiliary information kinds respectively. It is shown that the methodology outperforms existing techniques used in the literature for addressing the problem of prognostics under the conditions of limited failure data availability by a large margin. In each case study, an analysis is also performed to identify the key factors influencing the effectiveness of the methodology. It is shown that the availability of a sufficient amount of real failure data samples and the availability of auxiliary information pertaining to failure mechanisms are the key factors influencing the effectiveness of the methodology.

In addition to answering the two research questions, the thesis also provided recommendations for methodology application in the industry in Chapter 5. These recommendations are developed using the insights gained from the analyses performed from a theoretical perspective in Chapter 3 and a practical perspective in Chapter 4. The recommendations include the strengths and weaknesses of the methodology, potential risks associated with applying the methodology in the industry and risk mitigation actions, and application criteria that allow identifying suitable and unsuitable applications for the methodology.

6.3 Research contributions

In this section, the contributions made by this research to the academic knowledge and industrial practice are outlined.

6.3.1 Contributions to academic knowledge

The following contributions are made by this research to the academic knowledge:

- Provides a methodology for developing prognostics models under the conditions of limited failure data availability, and insights into its behaviour and working mechanism, its impact on prognostics compared to the existing techniques and key factors influencing its effectiveness.
- Provides a systematic approach to identifying and validating auxiliary information pertaining to failure mechanisms which are obtained from expert knowledge and are suitable for generating plausible failure data using the concept of conditional generative modelling.
- Provides two approaches for converting auxiliary information into the vector form which allows integrating auxiliary information pertaining to failure mechanisms into the conditional generative modelling framework: constructing an auxiliary information vector from a continuous distribution and constructing an auxiliary information vector from class labels.

- Provides a theoretical foundation for the methodology for prognostics under the conditions of limited failure data availability, and thus allows developing a rationale for prognostics and PdM results produced by the methodology when implemented in industrial scenarios.
- Provides the formal proof of conditional generative modelling using auxiliary information pertaining to failure mechanisms which shows the following: when conditioned on auxiliary information pertaining to failure mechanisms, the CGAN is capable of estimating a generative model that has captured the semantic features of the failure mode that needs predicting.

In addition, this research has contributed to the academic knowledge through the following academic publications:

- A Methodology for Prognostics Under Limited Failure Data Availability, journal article, IEEE Access Journal [123].
- Using Expert Knowledge to Generate Failure Data for Broadband Line Prognostics Under the Conditions of Limited Failure Data Availability, conference paper, 4th IFAC Workshop on Advanced Maintenance Engineering, Services and Technologies [134].
- Generating Real-valued Failure Data for Prognostics Under Limited Failure Data Availability, conference paper, 2019 IEEE Conference on Prognostics and Health Management, San Francisco, USA [135].

6.3.2 Contributions to industrial practice

The contributions of this research to the industrial practice are as follows:

- Enables the effective implementation of PdM under the conditions of limited failure data availability for prognostics, and thus allows preventing costs due to over-maintenance and under-maintenance of equipment and false alarms.
- Provides guidance on how to use expert knowledge of failure mechanisms that is available within industrial organisations through the maintenance engineers, reliability engineers, components designers, equipment operators and applied research scientists to overcome the problem of limited failure data availability for prognostics.
- Provides case studies and recommendations for applying the methodology for prognostics under the conditions of limited failure data availability in industrial scenarios.

6.4 Awards and recognition in academia and industry

During its elaboration, the research presented in this thesis has gained the following notable awards and recognition in academia and industry:

- Received the Best Paper Award at the 2019 IEEE International Conference on Prognostics and Health Management, San Francisco, USA.
- Popularity of this research in the industry led to starting the research collaboration between Scania Commercial Vehicles and University of Cambridge. This collaboration to date has allowed other researchers at the Institute for Manufacturing, University of Cambridge to obtain prognostics modelling datasets and expert knowledge from the industry and validate their research in real-world industrial scenarios.

6.5 Recommendations for future research

The following three research directions are recommended for future research: (i) identification of other kinds of auxiliary information that are suitable for conditioning the plausible failure data generation process of the methodology; (ii) enabling the application of the methodology in industrial scenarios in which the extent of limited failure data availability problem is approximately equal to or less than 0.1 normalised Shannon entropy; (iii) reducing the effort required to implement the methodology by automating the auxiliary information identification, validation and conversion steps. In the remainder of this section, these future research directions are discussed in detail.

Research direction 1: *Identification of other kinds of auxiliary information.* The research identified three kinds of auxiliary information that are suitable for conditioning the plausible failure data generation process of the methodology. Nevertheless, there are other kinds of auxiliary information that can be identified from expert knowledge in the prognostics and PdM domains, which could be suitable for conditioning the plausible failure data generation process. Investigating the suitability of other kinds of auxiliary information pertaining to failure mechanisms is recommended as a future research direction since it allows increasing the number of suitable industrial scenarios supported by the prerequisite of the methodology.

The other kinds of auxiliary information worth investigating could include the following: the effect of improper equipment configuration on failure mechanisms, the effect of improper repair and replacement actions on failure mechanisms, the effect of introducing new technologies into the legacy systems on failure mechanisms (e.g. extending legacy telecommunications infrastructure to support the 5G technology); and the effect of supply chain disruptions on failure mechanisms due to health (e.g. COVID-19) and political events (e.g. Brexit).

Research direction 2: *Improve methodology application.* As discussed in Sec. 5.1, one of the weaknesses of the methodology is that the prognostics performance starts to degrade when the normalised Shannon entropy is approximately equal to or less than 0.1. Hence, currently, it is recommended that the application of methodology should be avoided in industrial scenarios in which the extent of limited failure data availability problem for prognostics is approximately equal to or less than 0.1 normalised Shannon entropy.

Investigating extensions to the conditional generative model estimation aspect of the methodology that allows preventing the degradation of convergence performance of the CGAN model when the normalised Shannon entropy is approximately equal to or less than 0.1 is recommended as a future research direction. For example, the adaptive discriminator augmentation mechanism which has emerged recently in the image recognition domain for stabilising GAN training in limited data regimes could be a suitable extension to address this weakness of the methodology (see Karras et al. [136]).

Research direction 3: *Reduce effort required for methodology implementation.* As discussed in Sec. 5.1, the second weakness of the methodology is that it requires more effort to implement compared to the existing techniques. This is since auxiliary information needs to be identified from expert knowledge obtained from maintenance and reliability engineers, component designers, equipment operators and applied research scientists. Then it needs to be validated and converted into a suitable form for integrating into the plausible failure data generation process.

Investigating ways to automate the auxiliary information identification, validation and conversions steps of the methodology is a potential future research direction. For example, one could investigate how to develop a process to automate the auxiliary information validation step by integrating it into data-driven model training workflows, so that the validity of auxiliary information is continuously or periodically measured by an automated process.

This thesis presented a solution to the long-lasting problem of prognostics under the conditions of limited failure data availability. This allows predictive maintenance to deliver on the promise of an optimal maintenance policy for industrial organisations. Thus, this thesis provided a major contribution for realising the vision of the modern industry that aims to reduce costs and waste whilst increasing the productivity and added value.

References

[1] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006.

[2] Ying Peng, Ming Dong, and Ming Jian Zuo. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, 2010.

[3] Yaguo Lei, Naipeng Li, Liang Guo, Ningbo Li, Tao Yan, and Jing Lin. Machinery health prognostics: A systematic review from data acquisition to rul prediction. *Mechanical Systems and Signal Processing*, 104:799–834, 2018.

[4] Tangbin Xia and Lifeng Xi. Manufacturing paradigm-oriented phm methodologies for cyber-physical systems. *Journal of Intelligent Manufacturing*, 30(4):1659–1672, 2019.

[5] Sze-jung Wu, Nagi Gebraeel, Mark A Lawley, and Yuehwern Yih. A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(2):226–236, 2007.

[6] Evgeny Burnaev. Rare failure prediction via event matching for aerospace applications. In *2019 3rd International Conference on Circuits, System and Simulation (ICCSS)*, pages 214–220. IEEE, 2019.

[7] Bram de Jonge and Philip A Scarf. A review on maintenance optimization. *European Journal of Operational Research*, 2019.

[8] Yasmin Fathy, Mona Jaber, and Alexandra Brintrup. Learning with imbalanced data in smart manufacturing: A comparative analysis. *IEEE Access*, 2020.

[9] Jinjiang Wang, Yuanyuan Liang, Yinghao Zheng, Robert X Gao, and Fengli Zhang. An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. *Renewable energy*, 145:642–650, 2020.

[10] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239, 2017.

[11] Gary M Weiss. Mining with rarity: a unifying framework. *ACM Sigkdd Explorations Newsletter*, 6(1):7–19, 2004.

[12] JZ Sikorska, Melinda Hodkiewicz, and Lin Ma. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical systems and signal processing*, 25(5):1803–1836, 2011.

[13] Linxia Liao and Felix 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*, 63(1):191–207, 2014.

[14] Paul Johannesson and Erik Perjons. An introduction to design science. pages 75–89, 2014.

[15] Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[16] Kamran Javed, Rafael Gouriveau, and Noureddine Zerhouni. State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mechanical Systems and Signal Processing*, 94:214–236, 2017.

[17] Nicola Pugno, M Ciavarella, Pietro Cornetti, and Alberto Carpinteri. A generalized paris' law for fatigue crack growth. *Journal of the Mechanics and Physics of Solids*, 54(7):1333–1349, 2006.

[18] Elias Keedy and Qianmei Feng. A physics-of-failure based reliability and maintenance modeling framework for stent deployment and operation. *Reliability Engineering & System Safety*, 103:94–101, 2012.

[19] Elinirina I Robinson, Julien Marzat, and Tarek Raïssi. Model-based prognosis of fatigue crack growth under variable amplitude loading. *IFAC-PapersOnLine*, 51(24):176–183, 2018.

[20] Pavle Boškoski, Boštjan Dolenc, Bojan Musizza, and D Juričić. Model-based prediction of the remaining useful life of the machines. *IFAC-PapersOnLine*, 50(1):12803–12808, 2017.

[21] Runqing Huang, Lifeng Xi, Xinglin Li, C Richard Liu, Hai Qiu, and Jay Lee. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical systems and signal processing*, 21(1):193–207, 2007.

[22] Fraser Ewing, Philipp R Thies, Jonathan K Shek, and Claudio Bittencourt. A physics-based prognostics approach for tidal turbines. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 1–7. IEEE, 2019.

[23] Florent Aubry, Virginie Chameroy, Alain Giron, R Di Paola, B Gibaud, D Vital, R Lienard, A Todd-Pokropek, R Kanz, F Deconinck, et al. Mimosa: Conceptual modelling of medical image management in an open system architecture. In *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 1199–1201. IEEE, 1992.

[24] Diomidis H Stamatis. *Failure mode and effect analysis: FMEA from theory to execution*. Quality Press, 2003.

[25] J-SR Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3):665–685, 1993.

[26] Enrico Zio and Francesco Di Maio. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliability Engineering & System Safety*, 95(1):49–57, 2010.

[27] Jay Lee, Jun Ni, Dragan Djurdjanovic, Hai Qiu, and Haitao Liao. Intelligent prognostics tools and e-maintenance. *Computers in industry*, 57(6):476–489, 2006.

[28] L Botman. *Machine failure diagnostics and prognostics using error information and sensor data*. PhD thesis, Master thesis, Eindhoven University of Technology, Eindhoven, 2017.

[29] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou. Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research*, 213(1):1–14, 2011.

[30] Tianmei Li, Hong Pei, Zhenan Pang, Xiaosheng Si, and Jianfei Zheng. A sequential bayesian updated wiener process model for remaining useful life prediction. *IEEE Access*, 2019.

[31] Jerry Lawless and Martin Crowder. Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime data analysis*, 10(3):213–227, 2004.

[32] Chih-Chun Tsai, Sheng-Tsaing Tseng, and Narayanaswamy Balakrishnan. Optimal burn-in policy for highly reliable products using gamma degradation process. *IEEE Transactions on Reliability*, 60(1):234–245, 2010.

[33] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[34] Mikhail Nikulin, Hong-Dar Isaac Wu, et al. *The Cox model and its applications*. Springer, 2016.

[35] V Makis and X Jiang. Optimal replacement under partial observations. *Mathematics of operations research*, 28(2):382–394, 2003.

[36] Mark Schwabacher and Kai Goebel. A survey of artificial intelligence for prognostics. In *AAAI Fall Symposium: Artificial Intelligence for Prognostics*, pages 108–115, 2007.

[37] Warren S Sarle. Neural networks and statistical models. 1994.

[38] Zeqi Zhao, Bin Liang, Xueqian Wang, and Weining Lu. Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliability Engineering & System Safety*, 164:74–83, 2017.

[39] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9 (8):1735–1780, 1997.

[40] Charu C Aggarwal. Instance-based learning: A survey. *Data classification: algorithms and applications*, 157, 2014.

[41] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[42] Zhenyu Wu, Wenfang Lin, and Yang Ji. An integrated ensemble learning model for imbalanced fault diagnostics and prognostics. *IEEE Access*, 6:8394–8402, 2018.

[43] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, Michael S Lauer, et al. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.

[44] Erik Frisk, Mattias Krysander, and Emil Larsson. Data-driven lead-acid battery prognostics using random survival forests. Technical report, Linkoping University Linkoping Sweden, 2014.

[45] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[46] Renato de Pádua Moreira and Cairo Lúcio Nascimento. Prognostics of aircraft bleed valves using a svm classification algorithm. In *2012 IEEE Aerospace Conference*, pages 1–8. IEEE, 2012.

[47] Vasilis Sotiris and Michael Pecht. Support vector prognostics analysis of electronic products and systems. In *AAAI Fall Symposium on Artificial Intelligence for Prognostics*, pages 120–127, 2007.

[48] Adnan Nuhic, Tarik Terzimehic, Thomas Soczka-Guth, Michael Buchholz, and Klaus Dietmayer. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *Journal of power sources*, 239:680–688, 2013.

[49] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521 (7553):452, 2015.

[50] Kai Goebel, Bhaskar Saha, and Abhinav Saxena. A comparison of three data-driven techniques for prognostics. In *62nd meeting of the society for machinery failure prevention technology (mfpt)*, pages 119–131, 2008.

[51] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, 2003.

[52] Daeil Kwon, Michael H Azarian, and Michael Pecht. Remaining-life prediction of solder joints using rf impedance analysis and gaussian process regression. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 5(11):1602–1609, 2015.

[53] Michael E Tipping. The relevance vector machine. In *Advances in neural information processing systems*, pages 652–658, 2000.

[54] Datong Liu, Jianbao Zhou, Dawei Pan, Yu Peng, and Xiyuan Peng. Lithium-ion battery remaining useful life estimation with an optimized relevance vector machine algorithm with incremental learning. *Measurement*, 63:143–151, 2015.

[55] Aiwina Heng, Sheng Zhang, Andy CC Tan, and Joseph Mathew. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical systems and signal processing*, 23(3):724–739, 2009.

[56] Jay Lee, Fangji Wu, Wenyu Zhao, Masoud Ghaffari, Linxia Liao, and David Siegel. Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mechanical systems and signal processing*, 42(1-2):314–334, 2014.

[57] Samir Khan and Takehisa Yairi. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 107:241–265, 2018.

[58] Minh-Thang Luong, Quoc V Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser. Multi-task sequence to sequence learning. *arXiv preprint arXiv:1511.06114*, 2015.

[59] Ansi Zhang, Honglei Wang, Shaobo Li, Yuxin Cui, Zhonghao Liu, Guanci Yang, and Jianjun Hu. Transfer learning with deep recurrent neural networks for remaining useful life estimation. *Applied Sciences*, 8(12):2416, 2018.

[60] PABLO M Granitto, FLAVIA Gasperi, FRANCO Biasioli, E Trainotti, and Cesare Furlanello. Modern data mining tools in descriptive sensory analysis: A case study with a random forest approach. *Food Quality and Preference*, 18(4):681–689, 2007.

[61] Jérôme Paul, Michel Verleysen, Pierre Dupont, et al. Identification of statistically significant features from random forests. In *ECML workshop on Solving Complex Machine Learning Problems with Ensemble Methods*, pages 69–80, 2013.

[62] Fen Miao, Yun-Peng Cai, Yuan-Ting Zhang, and Chun-Yue Li. Is random survival forest an alternative to cox proportional model on predicting cardiovascular disease? In *6TH European conference of the international federation for medical and biological engineering*, pages 740–743. Springer, 2015.

[63] Tangbin Xia, Yifan Dong, Lei Xiao, Shichang Du, Ershun Pan, and Lifeng Xi. Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliability Engineering & System Safety*, 2018.

[64] Ed Snelson. Tutorial: Gaussian process models for machine learning. *Gatsby Computational Neuroscience Unit, UCL*, 2006.

[65] Janell Duhaney, Taghi M Khoshgoftaar, and Amri Napolitano. Studying the effect of class imbalance in ocean turbine fault data on reliable state detection. In *2012 11th International Conference on Machine Learning and Applications*, volume 1, pages 268–275. IEEE, 2012.

[66] Ajinkya More. Survey of resampling techniques for improving classification performance in unbalanced datasets. *arXiv preprint arXiv:1608.06048*, 2016.

[67] Shuai Zheng, Ahmed Farahat, and Chetan Gupta. Generative adversarial networks for failure prediction. *arXiv preprint arXiv:1910.02034*, 2019.

[68] Taisheng Zheng, Lei Song, Jianxing Wang, Wei Teng, Xiaoli Xu, and Chao Ma. Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings. *Measurement*, page 107741, 2020.

[69] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[70] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[71] Adrià Salvador Palau, Zhenglin Liang, Daniel Lütgehetmann, and Ajith Kumar Parlikad. Collaborative prognostics in social asset networks. *Future Generation Computer Systems*, 92: 987–995, 2019.

[72] Chao Jin, Dragan Djurdjanovic, Hossein D Ardakani, Keren Wang, Matthew Buzza, Behrad Begheri, Patrick Brown, and Jay Lee. A comprehensive framework of factory-to-factory dynamic fleet-level prognostics and operation management for geographically distributed assets. In *2015 ieee international conference on automation science and engineering (case)*, pages 225–230. IEEE, 2015.

[73] Maharshi Dhada, Amit Jain, Manuel Herrera, Marco Hernandez, and Ajith Parlikad. Secure and communications efficient collaborative prognosis. 2020.

[74] Ying Lin, Kaibo Liu, Eunshin Byon, Xiaoning Qian, Shan Liu, and Shuai Huang. A collaborative learning framework for estimating many individualized regression models in a heterogeneous population. *IEEE Transactions on Reliability*, 67(1):328–341, 2017.

[75] Adrià Salvador Palau. *Distributed Collaborative Prognostics*. PhD thesis, University of Cambridge, 2020.

[76] Abhinav Saxena, Kai Goebel, Don Simon, and Neil Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management*, pages 1–9. IEEE, 2008.

[77] Aida Rezaei and Azzedine Dadouche. Development of a turbojet engine gearbox test rig for prognostics and health management. *Mechanical systems and signal processing*, 33:299–311, 2012.

[78] Balaji Sankar, Thennavarajan Subramanian, Brijeshkumar Shah, Vijayendranath Vanam, Soumendu Jana, Srinivisan Ramamurthy, Radhakant Satpathy, Benudhar Sahoo, and Satish Yadav. Aero-thermodynamic modelling and gas path simulation for a twin spool turbo jet engine. In *ASME 2013 Gas Turbine India Conference*. American Society of Mechanical Engineers Digital Collection, 2013.

[79] Wilfried PJ Visser and Michael J Broomhead. Gsp a generic object-oriented gas turbine simulation environment. 2000.

[80] Edward L Melnick and Brian S Everitt. *Encyclopedia of quantitative risk analysis and assessment*, volume 1. John Wiley & Sons, 2008.

[81] Stephan Spiegel, Fabian Mueller, Dorothea Weismann, and John Bird. Cost-sensitive learning for predictive maintenance. *arXiv preprint arXiv:1809.10979*, 2018.

[82] Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme. Cost-sensitive learning methods for imbalanced data. In *The 2010 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2010.

[83] Chao Chen, Andy Liaw, Leo Breiman, et al. Using random forest to learn imbalanced data. *University of California, Berkeley*, 110(1-12):24, 2004.

[84] S Alestra, C Brand, E Burnaev, P Erofeev, A Papanov, C Bordry, and C Silveira-Freixo. Rare event anticipation and degradation trending for aircraft predictive maintenance. In *11th World Congress on Computational Mechanics, WCCM*, volume 5, page 6571, 2014.

[85] Gary M Weiss and Haym Hirsh. Learning to predict extremely rare events. In *AAAI workshop on learning from imbalanced data sets*, pages 64–68. AAAI Press, 2000.

[86] Pradeep Ramuhalli, Jamie Coble, Ryan M Meyer, and Leonard J Bond. Prognostics health management and life beyond 60 for nuclear power plants. In *2012 Future of Instrumentation International Workshop (FIIW) Proceedings*, pages 1–4. IEEE, 2012.

[87] PV Varde and Michael G Pecht. Role of prognostics in support of integrated risk-based engineering in nuclear power plant safety. *Int. J. Progn. Health Manag*, 3:59, 2012.

[88] Jonas Biteus and Tony Lindgren. Planning flexible maintenance for heavy trucks using machine learning models, constraint programming, and route optimization. *SAE International Journal of Materials and Manufacturing*, 10(3):306–315, 2017.

[89] N Lybeck, B Pham, M Tawfik, JB Coble, RM Meyer, P Ramuhalli, and LJ Bond. Lifecycle prognostics architecture for selected high-cost active components. Technical report, Idaho National Laboratory (INL), 2011.

[90] AT&T Business Operations. Why predictive maintenance is key in manufacturing, 2020. URL https://www.business.att.com/learn/tech-advice/why-predictive-maintenance-is-key-in-manufacturing.html.

[91] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*, 2017.

[92] Jason Van Hulse, Taghi M Khoshgoftaar, Amri Napolitano, and Randall Wald. Feature selection with high-dimensional imbalanced data. In *2009 IEEE International Conference on Data Mining Workshops*, pages 507–514. IEEE, 2009.

[93] Michael Pecht. Prognostics and health management of electronics. *Encyclopedia of structural health monitoring*, 2009.

[94] Andrew KS Jardine and Albert HC Tsang. *Maintenance, replacement, and reliability: theory and applications*. CRC press, 2005.

[95] Michael Pecht and Jie Gu. Physics-of-failure-based prognostics for electronic products. *Transactions of the Institute of Measurement and Control*, 31(3-4):309–322, 2009.

[96] Sony Mathew, Diganta Das, Roger Rossenberger, and Michael Pecht. Failure mechanisms based prognostics. In *2008 International Conference on Prognostics and Health Management*, pages 1–6. IEEE, 2008.

[97] Jessy George Smith, Sai Kirthi Ponnuru, and Mandar Patil. Detection of aggressive driving behavior and fault behavior using pattern matching. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 207–211. IEEE, 2016.

[98] Ben Zhao, Leon Hu, Harold Sun, Ce Yang, Xin Shi, James Yi, Eric Curtis, and Abraham Engeda. Numerical investigation of a novel approach for mitigation of forced response of a variable geometry turbine during exhaust braking mode. In *Turbo Expo: Power for Land, Sea, and Air*, volume 49866, page V008T23A005. American Society of Mechanical Engineers, 2016.

[99] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[100] Conor C Taff, David Steinberger, Courtney Clark, Kara Belinsky, Hayley Sacks, Corey R Freeman-Gallant, Peter O Dunn, and Linda A Whittingham. Multimodal sexual selection in a warbler: plumage and song are related to different fitness components. *Animal Behaviour*, 84 (4):813–821, 2012.

[101] OpenWeather. Openweathermap api. https://openweathermap.org/api, 2020. (Accessed on 01/12/2020).

[102] MetOffice. Metoffice data point. https://www.metoffice.gov.uk/services/data/datapoint, 2020. (Accessed on 01/12/2020).

[103] Haibo He and Yunqian Ma. *Imbalanced learning: foundations, algorithms, and applications*. John Wiley & Sons, 2013.

[104] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python, 2011. URL https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.

[105] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Modeling tabular data using conditional gan. In *Advances in Neural Information Processing Systems*, pages 7335–7345, 2019.

[106] Mrinal Kanti Baowaly, Chia-Ching Lin, Chao-Lin Liu, and Kuan-Ta Chen. Synthesizing electronic health records using improved generative adversarial networks. *Journal of the American Medical Informatics Association*, 26(3):228–241, 2019.

[107] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex J Smola. A kernel method for the two-sample-problem. In *Advances in neural information processing systems*, pages 513–520, 2007.

[108] Olga Fink, Enrico Zio, and Ulrich Weidmann. A classification framework for predicting components' remaining useful life based on discrete-event diagnostic data. *IEEE Transactions on Reliability*, 64(3):1049–1056, 2015.

[109] Liangwei Zhang, Jing Lin, Bin Liu, Zhicong Zhang, Xiaohui Yan, and Muheng Wei. A review on deep learning applications in prognostics and health management. *Ieee Access*, 7: 162415–162438, 2019.

[110] Ke Li, Yalei Wu, Yu Nan, Pengfei Li, and Yang Li. Hierarchical multi-class classification in multimodal spacecraft data using dnn and weighted support vector machine. *Neurocomputing*, 259:55–65, 2017.

[111] Shaza M Abd Elrahman and Ajith Abraham. A review of class imbalance problem. *Journal of Network and Innovative Computing*, 1(2013):332–340, 2013.

[112] David MW Powers. What the f-measure doesn't measure: Features, flaws, fallacies and fixes. *arXiv preprint arXiv:1503.06410*, 2015.

[113] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[114] MathWorks MATLAB. Vehicle transmission system fault simulation, 2020. URL https://uk.mathworks.com/help/predmaint/ug/Use-Simulink-to-Generate-Fault-Data.html.

[115] MathWorks. Vibration analysis of rotating machinery. https://uk.mathworks.com/help/signal/ug/vibration-analysis-of-rotating-machinery.html, 2020. (Accessed on 01/12/2020).

[116] Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich, and Fred A Hamprecht. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1):213, 2009.

[117] Srikanth Sundaresan, Walter De Donato, Nick Feamster, Renata Teixeira, Sam Crawford, and Antonio Pescapè. Broadband internet performance: a view from the gateway. *ACM SIGCOMM computer communication review*, 41(4):134–145, 2011.

[118] Andy Valdar. *Understanding telecommunications networks*, volume 52. IET, 2006.

[119] David Yearling and David J Hand. Investigating temporal patterns of fault behaviour within large telephony networks. In *International Symposium on Intelligent Data Analysis*, pages 340–349. Springer, 2001.

[120] Michal Tencer and John Seaborn Moss. Humidity management of outdoor electronic equipment: methods, pitfalls, and recommendations. *IEEE Transactions on Components and Packaging Technologies*, 25(1):66–72, 2002.

[121] D Yearling and DJ Hand. A bayesian network datamining approach for modelling the physical condition of copper access networks. *BT technology journal*, 21(2):90–100, 2003.

[122] The Office of Communications. Investigation into bt's compliance with its obligations as a broadband universal service provider, 2020. URL https://www.ofcom.org.uk/about-ofcom/latest/bulletins/competition-bulletins/open-cases/cw_01256.

[123] Gishan D Ranasinghe, Tony Lindgren, Mark Girolami, and Ajith K Parlikad. A methodology for prognostics under the conditions of limited failure data availability. *IEEE Access*, 7: 183996–184007, 2019.

[124] UCI Machine Learning Repository. Scania air processing system failure dataset, 2016. URL https://archive.ics.uci.edu/ml/datasets/APS+Failure+at+Scania+Trucks.

[125] Camila Ferreira Costa and Mario A Nascimento. Ida 2016 industrial challenge: Using machine learning for predicting failures. In *International Symposium on Intelligent Data Analysis*, pages 381–386. Springer, 2016.

[126] Vítor Cerqueira, Fábio Pinto, Claudio Sá, and Carlos Soares. Combining boosted trees with metafeature engineering for predictive maintenance. In *International Symposium on Intelligent Data Analysis*, pages 393–397. Springer, 2016.

[127] Tony Lindgren, Håkan Warnquist, and Martin Eineborg. Improving the maintenance planning of heavy trucks using constraint programming. In *Proceedings of the 12th international workshop on constraint modelling and reformulation co-located with the 19th international conference on principles and practice of constraint programming (modref)*, pages 74–90, 2013.

[128] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[129] Christopher Gondek, Daniel Hafner, and Oliver R Sampson. Prediction of failures in the air pressure system of scania trucks using a random forest and feature engineering. In *International Symposium on Intelligent Data Analysis*, pages 398–402. Springer, 2016.

[130] T Raja, R Rajasekar, R Siva, N Karthik, and R Kumarasubramanian. Redesigning of the diesel engine turbocharger compressor wheel to improve the low-cycle fatigue life. *International Journal of Ambient Energy*, 40(8):821–826, 2019.

[131] Christopher P Diehl and Gert Cauwenberghs. Svm incremental learning, adaptation and optimization. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 4, pages 2685–2690. IEEE, 2003.

[132] Amazon Web Services. Incremental training in amazon sagemaker. https://docs.aws.amazon.com/sagemaker/latest/dg/incremental-training.html, 2021. (Accessed on 01/01/2021).

[133] Microsoft Azure ML. Retraining and updating azure machine learning models with azure data factory. https://azure.microsoft.com/en-gb/blog/retraining-and-updating-azure-machine-learning-models-with-azure-data-factory/, 2021. (Accessed on 01/01/2021).

[134] Gishan Don Ranasinghe, David Yearling, Mark Girolami, and Ajith Kumar Parlikad. Using expert knowledge to generate data for broadband line prognostics under limited failure data availability. *IFAC-PapersOnLine*, 53(3):265–270, 2020.

[135] Gishan Don Ranasinghe and Ajith Kumar Parlikad. Generating real-valued failure data for prognostics under the conditions of limited data availability. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 1–8. IEEE, 2019.

[136] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676*, 2020.