Data-driven Prognostics and Health Management for Maritime Systems Employing Trustworthy Digital Twins

Jaehan Jeon

A thesis submitted for the degree of Doctor of Philosophy

December 2024

Declaration of Authenticity and Author's Rights

Declaration of Authenticity and Author's Rights This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination, which has led to the award of a degree.

Signed: Jaehan Jeon

Date: 17th Dec, 2024

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Gerasimos Theotokatos, for his invaluable guidance throughout my PhD journey. I am also grateful to Maritime Safety Research Centre, Department of Naval Architecture, Ocean, and Marine Engineering, University of Strathclyde, and the AUTOSHIP project for their financial support. I heartfelt thank to my family for their continuous support, encouragement, and love throughout this journey. A special and heartfelt thanks to my wife, Yeonjin Jung, for her unwavering patience, love, and understanding, without which this journey would not have been possible.

Abstract

Maritime Autonomous Surface Ships (MASS) are expected to revolutionise the maritime and shipping industries benefiting the operations sustainability, whilst enhancing safety and resilience. However, the unmanned operations of MASS present immense challenges pertaining to emergency responses, including shipboard corrective maintenance. Prognostics and Health Management (PHM) systems and their applications to ship machinery have received increasing attention as means of addressing these challenges. PHM systems require data-driven models to estimate machinery health status, the development of which, however, is impeded by the scarcity of data representing anomaly conditions and stochastic degradation of machinery components. This study aims at advancing PHM in maritime systems by developing digital tools, namely physics-based digital twin and data-driven PHM model, as well as frameworks to evaluate and manage their trustworthiness. The first digital tool is the digital twin that integrates thermodynamic models, component degradation models, and sensor models, while ensuring its trustworthiness by using a framework that uses the steps of validation, verification, and robustness. The second digital tool is the data-driven models for fault diagnosis and health prognosis, including Health Indicator (HI) construction and forecast sub-tasks. The data-driven methods for each sub-model are systematically selected and integrated to develop a comprehensive PHM model, which subsequently supports maintenance decision making by estimating the Remaining Useful Life (RUL) of the components. The trustworthiness of data-driven models is managed throughout the engine lifetime considering accuracy and robustness metrics. A marine four-stroke engine is used as a reference system, assuming that its cylinder valves degrade according to different stochastic degradation patterns. The novelty in this research stems from the integration of methods to develop digital tools and frameworks that ensure their trustworthiness while addressing real-world challenges in the maritime domain.

Research Output

Journal Papers

J. Jeon, G. Theotokatos, A Framework to Assure the Trustworthiness of Physical Model-Based Digital Twins for Marine Engines, *Journal of Marine Science and Engineering* 2024, Vol. 12, Page 595 12 (2024) 595. https://doi.org/10.3390/JMSE12040595

Conference Papers

J. Jeon, G. Theotokatos, Datasets envelope impact on marine engines prognostics and health management models accuracy, in: *33rd European Safety and Reliability Conference*, 2023, pp. 2901–2906. https://doi.org/10.3850/978-981-18-8071-1_P592-cd

Contents

List of Figures

List of Tables

Acronyms & List of Symbols

Acronyms

ADO Advanced Direct Optimiser

AE Autoencoder

ANN Artificial Neural Network

ARIMA Autoregressive Integrated Moving Average

CI Confidence Interval

CNN Convolutional Neural Network

DDM Data-driven Model

DLM Degradation Law Modelling

DNN Deep Neural Network

DOE Design of Experiments

DT Digital Twin

DTR Decision Tree Regression

ELBO Evidence Lower Bound

EoL End of Life

ETA Estimated Time of Arrival

FMECA Failure Mode, Effects, and Criticality Analysis

GAN Generative Adversarial Network

GRU Gated Recurrent Unit

HI Health Indicator

HM Hybrid Model

KET Key Enabling Technology

KL Kullback-Leibler

KNN K-Nearest Neighbour

KNR K-Nearest Regression

LR Linear Regression

LSTM Long Short-term Memory

MASS Maritime Autonomous Surface Ship

MC Monte Carlo

MLP Multilayer Perceptron

MoE Margin of Error

MVEM Mean Value Engine Model

NASA National Aeronautics and Aerospace Administration

NSGA Non-dominated Sorting Genetic Algorithm

PDF Probability Density Function

PHM Prognostics and Health Management

PID Proportional Integral Derivative

PM Physics-based Model

RBF Radial Basis Function

RF Random Forest

RFR Random Forest Regression

RNN Recurrent Neural Network

RPN Risk Priority Number

RUL Remaining Useful Life

SOM Self Organising Map

SPM Stochastic Process Modelling

SVI Stochastic Variational Inference

SVM Support Vector Machine

SVR Support Vector Regression

TRL Technical Readiness Level

Abbreviations

0D Zero Dimensional

1D One Dimensional

3D Three Dimensional

Cyl Cylinder

Leak Leakage

savgol Savitzky-Golay Filter

Symbols

$\kappa[-]$ Outcome of Kernel Function

$\phi[-]$ Distribution Parameters

$\rho[-]$ Person's Correlation Coefficient

$\sigma[-]$ Outcome of Sigmoid Function

$\theta[-]$ BNN Parameters

$A[\%]$ Accuracy from Confusion Matrix

$E[\%]$ Percentage Error

$F[-]$ Outcome of Multidimensional Kernel Function

$FOC[g/kWh]$ Fuel Oil Consumption

$Ga[-]$ Gamma Distribution

$L[mm]$ Valve Lift

$L(t)[-]$ Labelled Data Value

$logit[-]$ Outcome of Logit Function

$Loss[-]$ Outcome of Loss Function

$Mon[-]$ Monotonicity

$N[rev/m]$ Rotational Speed

$p[bar]$ Pressure

$Prog[-]$ Prognosability

$r[mm]$ Valve Recession

$R^2[-]$ Coefficient of Determination

$R_m[-]$ Uncertainty Ratio

$r_s[-]$ Spearman's Coefficient

$R_u[-]$ Uncertainty Ratio for Robustness Criteria

$S^2[-]$ Variance

$sgn[-]$ Outcome of Signum Formula

$T[K]$ Temperature

$Trend[-]$ Trendability

$UL(t)[-]$ Unlabelled Data Value

$V[m^3]$ Wear Volume

$W[mm]$ Total Valve Wear

$X'[-]$ Normalised Value

Subscripts

Amb Ambient

CA Charge Air

EV Exhaust Valve

Exh Exhaust Gas

IV Intake Valve

max Maximum

TC Turbocharger

Chapter 1

Introduction

1.1 Digitalisation in the Maritime Industry

The maritime industry transports over 80% of the world trade by volume, a dependency that is anticipated to grow in the foreseeable future [1]. The maritime industry faces significant pressure to improve sustainability while remaining competitive, leading to an increased interest in digitalisation technologies [2]. Digitalisation transforms business models and introduces new value-producing opportunities through the implementation of cutting-edge technologies, including autonomous vehicles with robotics, artificial intelligence, big data, virtual/augmented/mixed reality, the internet of things, cloud computing, cyber security, and 3D printing [3].

Digitalisation in the maritime industry is expected to improve the productivity, efficiency, sustainability, and safety of ship operations [4, 5]. Despite the benefits of digitalisation, the maritime industry has a relatively slower adoption compared to other transport industries [6, 7]. This is primarily attributed to the conservatism of the maritime industry and the technical challenges posed by the dynamic nature of ship operations, as well as the extent of the ship systems operating envelope [8]. For reliable operations with digitalised systems in real-world scenarios, the system must demonstrate its capability across a wide operating envelope [9].

The advanced digitalisation technologies have effectively managed complex, high-dimensional, and noisy datasets [10]; however, their performances in a wide operating envelope have not been explored. Furthermore, the growing complexity of the maritime systems introduces new risks [11], necessitating thorough research to

identify the risks and manage the unforeseen safety issues with digitalised systems [12].

1.1.1 Maritime Autonomous Surface Ships

Maritime Autonomous Surface Ship (MASS) is one of the major applications in the maritime digitalisation research domain. Recent studies on autonomous ship technology classified Key Enabling Technology (KET) at Technical Readiness Level (TRL) 7 [13], indicating prototype testing in relevant environments. During the past decade, numerous projects have been conducted to advance and demonstrate KETs for autonomous shipping. Despite these efforts, unmanned operations remain unachieved, and autonomous operations still depend on remote human intervention [14].

Table 1.1 presents a summary of operating area and power sources in commercial projects that include shipbuilding. Current technology level limits the operating area of autonomous shipping to inland waterways or short-sea shipping [15]. The limitations in operational areas are attributable to the maintenance requirements of the machinery systems, as the probability of system failure increases with the duration of the voyage under unmanned conditions [16]. To reduce maintenance frequency, most projects, apart from SeaShuttle, employ battery packs as their main power source; however, their low energy density (1,224 kJ/kg) imposes limitations on the sailing distance [17]. A hydrogen-powered fuel cell is used for the SeaShuttle project; however, the economic feasibility of the fuel cell is low compared to the conventional marine engine system at the current price point [18]. Research on the design of machinery systems and their management strategy is required to expand the operating area of autonomous shipping.

The MASS is anticipated to improve the efficiency of maritime transport and address the current challenges of the maritime industry. Shipping companies are currently facing difficulties in recruiting high-quality seafarers. The harsh working and living conditions have led to dissatisfaction of the crews, resulting in their resignations [19]. The shortage of high-quality crews threatens the safety and sustainability of the shipping industry [20]. The implementation of MASS under the remote control concept is expected to attract high-quality workers to the shipping

Table 1.1: Commercial autonomous ship projects

Starting Year	Vehicle Name	Technology Provider	Control Method	Operation Area	Power Source
2017	YARA Birkeland	Kongsberg	Remote controlled	Short Sea (30 Nautical Miles)	Battery (7–9 MWh)
2020	ASKO	Kongsberg	Remote controlled	Short Sea (Coastal Terrain)	Battery (1.846 MWh)
2021	SEAFAR	Alewijnse	Remote controlled	Inland Waterway	–
2022	Reach Remote	Kongsberg	Remote controlled	Short Sea (Offshore Area)	Main Engine with Battery (369 kWh*2)
2022	SeaShuttle	Ocean Infinity	Remote controlled	Short Sea	Hydrogen-powered Fuel Cell (3.2 MW)
2022	Ekornes and DB Schenker	Kongsberg	Remote controlled	Short Sea (23 Nautical Miles)	Battery (600/1200kWh)

industry and mitigate the issue of crew shortages [21] by shifting hazardous and physically demanding onboard duties to shore-based monitoring duties [22].

Moreover, the MASS reduces operating cost and provides economical advantages. The removal of hotel systems for supporting crews not only reduces the cost of shipbuilding, but also reduces the weight of the vessel, energy consumption and maintenance requirements [23]. According to the findings of the DNV project "ReVolt", the deployment of autonomous ships equipped with state-of-the-art efficiency-improving technologies can result in a cost savings of 48 million USD over a 30-year operating lifetime [24].

From a safety perspective, the implementation of MASS is anticipated to reduce the incidence of maritime accidents, which serves as the primary motivation for initial research into autonomous ship technologies [25]. According to the 2023 annual overview of marine casualties and incidents report, human factors were involved in 80.7% of the investigated marine casualties and incidents [26]. The finding indicates that human attributes such as misinterpretation, negligence, and fatigue contribute significantly to a substantial number of maritime accidents. Autonomous shipping has the potential to eliminate human-related errors, thereby reducing maritime accidents.

Despite the aforementioned advantages, the implementation of autonomous shipping remains delayed by a significant challenge associated with emergency responses. In unmanned operational scenarios, the absence of onboard crews and corrective maintenance means that even minor malfunctions in machinery systems can lead to loss of propulsion power [27]. The deployment of MASS within maritime logistics necessitates that the reliability of their machinery systems matches or exceeds that of

conventional vessels [28, 29], as stakeholders are reluctant to entrust their services to unreliable transport [30]. Consequently, research on machinery health management systems is required to ensure the reliability of autonomous ship machinery systems [31].

1.1.2 Prognostics and Health Management System

Maintenance is of paramount importance for managing maritime machinery systems reliably and cost-effectively. During the past few decades, a range of maintenance strategies, including corrective, preventive, and predictive approaches, have evolved, as illustrated in Figure 1.1 [32].

The corrective maintenance strategy involves a run-to-fail approach in which repairs or replacements are performed upon component failure. This strategy is preferred when the consequence of failure is minimal; however, it increases maintenance cost and risk of downtime compared to other strategies [33].

The preventive maintenance strategy implements early actions to avoid potential future anomalies. There are two predominant approaches for scheduling maintenance: time-based and usage-based approaches. The former relies on pre-established time intervals, whereas the latter considers actual operational cycles [34].

The predictive maintenance strategy includes three approaches: Reliability-centred, Risk-based, and Condition-based approaches. The reliability-centred approach determines the maintenance requirements to ensure the continuous operation of critical components [35]. The risk-based approach prioritises the likelihood of any failure and its consequences [32]. This approach aims to minimise the frequency of failure and mitigate overall risk [33]. The condition-based approach aims to detect early signs of failure and perform maintenance based on the estimated health condition of a reference system [36]. Increasing interest in multi-component system management, coupled with rapid advances in data-driven technologies, makes the condition-based strategy a leading maintenance strategy for maintenance scheduling [37]. This approach has evolved into Prognostics and Health Management (PHM), including a comprehensive process from data acquisition to decision making [38].

Recent studies on maintenance strategies have evolved towards prescriptive maintenance, offering actionable recommendations and optimising maintenance

processes [39]. Nonetheless, these methods remain underdeveloped and are not yet widely adopted.

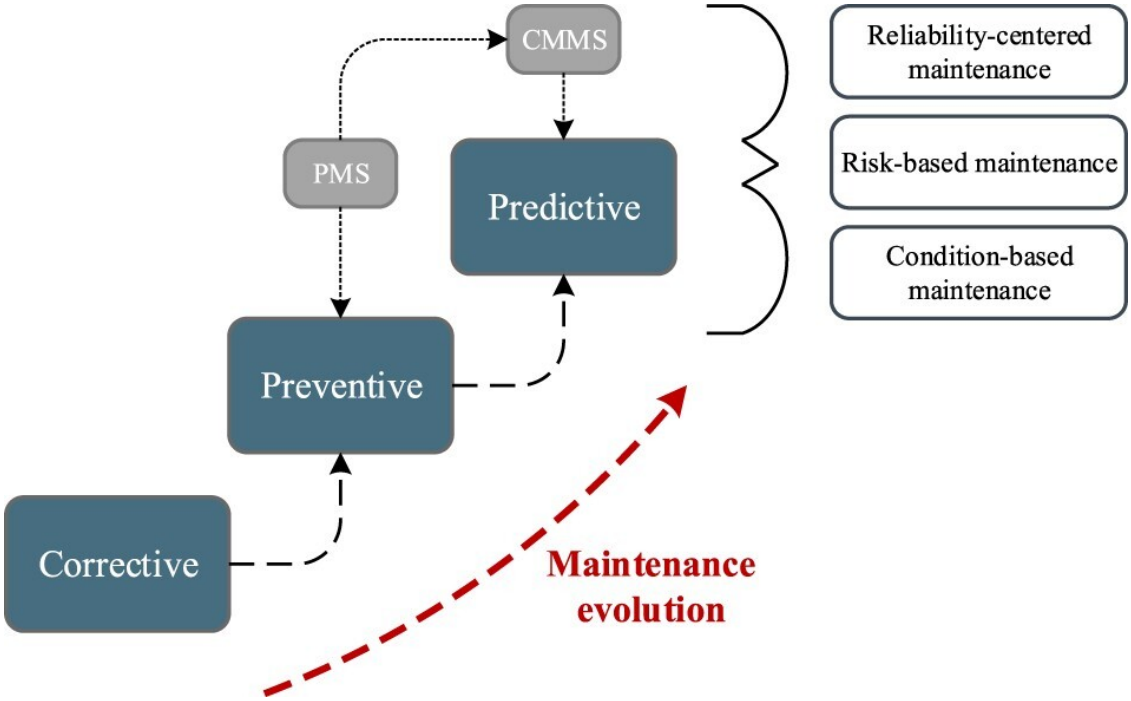


Figure 1.1: Maintenance strategy evolution

The PHM system can assess the current and future health status of ship machinery systems [40], thereby enabling optimal maintenance decision making [41]. PHM systems increase the reliability of ship machinery systems and reduce system downtime [42], subsequently supporting MASS operations under unmanned conditions [43]. However, the advancement of PHM technology is impeded by data-centric challenges, including the scarcity and incompleteness of measurement datasets [37]. Acquiring datasets that comprehensively represent machinery health conditions is a massive challenge due to the varying operational environments within a wide operating envelope of ship machinery systems [44]. The operations of ship machinery systems are influenced by environmental conditions, as well as the presence of faults and degradation within sub-systems. Typically, shipboard-measured datasets represent only a limited operating envelope in healthy conditions [45, 46] of well-maintained machinery systems [47].

To address the challenge of limited data availability, several studies employ data augmentation techniques, while frequently neglecting interactions with other system components [48]. Moreover, some researchers employ simulations with Digital Twin (DT) to generate data; however, their trustworthiness is often insufficiently addressed.

These artificial datasets do not adequately represent the complex behaviour of ship machinery systems, as machinery operation is an intricate phenomenon governed by both internal and external variables. To effectively augment anomaly data, it is essential to consider a system-level perspective that comprehensively accounts for interactions and interdependencies among system components [49].

1.2 Maritime Industry Outlook

Maritime 4.0 is defined as the integrated implementation of digital technologies in vessel design, vessel construction, operation, and management [50]. Since 2011, there has been a significant increase in research publications related to Maritime 4.0; however, the technologies supporting Maritime 4.0 remain underdeveloped and require extensive research [51].

With respect to MASS, the benefits of autonomous shipping are widely acknowledged among stakeholders; however, technological maturity [52] and extensive changes in the business model [53] require a more extended time frame. A major upgrade or replacement of the existing system leads to high investment costs and risks; however, their pay-off from the investment is not guaranteed to the maritime stakeholders [54]. The realisation of fully autonomous shipping is undoubtedly a long-term goal [55]; however, in progression towards this goal, the maritime industry will implement digitalisation technologies and reduce onboard crew numbers [5]. Favourable experiences derived from the implementation of semi-automated systems during this transition period will accelerate the progress toward fully autonomous shipping [15].

Toward the digitalisation of the marine industry, the research opportunities are identified as follows:

1. The maritime industry has employed automation technologies to assist human operators and gradually replace the human operator's role. The advancement of automation technologies transforms the operational paradigms within the maritime industry, while concurrently introducing unforeseen safety issues [12, 56]. In-depth risk analysis studies are required to address the transformed operational paradigms.

2. The existing rules and regulations are developed based on the designs and operational paradigms of conventional ships [15]. Significant amendment and clarification of these regulatory frameworks are required, particularly in addressing operator responsibility, to integrate unmanned vessels into maritime logistics [57].

3. In the context of remote control systems, connectivity between vessels and shore control centres is of paramount importance. However, remote communications are vulnerable to cyber attacks [58]. Consequently, there is a demand for specialised research on communication systems for remote vessels to mitigate the risks associated with remote operations.

4. Current research projects and studies in the MASS research domain have not determined a suitable machinery system design and management strategy for deep-sea autonomous shipping [56]. Consequently, comprehensive research on the design and management of machinery systems is required in accordance with the distinctive operational environment of autonomous ships.

5. The human operators in the shore control centre require high-quality and large-quantity data to understand the health status of remote ships; however, the data transmission has limitations such as latency and transmission frequency [29]. These information gaps pose significant risks of inadequate decision making by remote operators. Research on decision support and making systems is required to bridge these information gaps and provide processed information to human operators, thereby mitigating the risks of inadequate decision making induced by these information gaps.

1.3 Research Question

In summary, the management of the ship machinery system constitutes a significant obstacle to the digitalisation of the maritime industry and the implementation of autonomous shipping. The PHM system is anticipated to mitigate this obstacle; however, the distinctive operational environment of ship machinery systems poses challenges to the development of the PHM model, including limited data availability

and stochasticity in degradation patterns.

The following questions have been formulated to guide the research.

How can a PHM system be used for managing the ship machinery health condition during autonomous operations?

1. How can the development of data-driven PHM models address the challenges posed by the distinctive operating conditions of ships?
2. How can DT-generated simulation data effectively supplement shipboard measurement data?
3. How can the stochasticity in component degradation patterns be managed by data-driven PHM models?

1.4 Aim & Objectives

This research aims at advancing PHM in maritime systems by developing digital tools, as well as frameworks to ensure their trustworthiness. This study sets high-level objectives to address the aim as follows.

1. Conduct a comprehensive literature review to identify the outlook, research gaps, and state-of-the-art methods for PHM models
2. Use a marine engine as a reference system and analyse critical anomalies (degradation) of the engine
3. Develop a physics-based DT of the reference system
4. Develop a framework to assure trustworthiness of the physics-based DT
5. Generate simulation datasets using the trustworthy DT considering degradation of reference system components
6. Develop a data-driven PHM model to estimating RUL of the degradation components
7. Develop a framework for managing the trustworthiness of data-driven PHM model throughout the reference system lifetime

8. Demonstrate the digital tools (Obj. 3 & 6), simulation datasets (Obj. 5), and frameworks to manage the trustworthiness of digital tools (Obj. 4 & 7) employing various operating profiles

1.5 Research Scope & Boundary

The development of digital tools and frameworks considers the distinctive operational environment of ship machinery systems, especially the unavailability of labelled data during operations. Health labels in the simulation-generated datasets are isolated from the training phases and used to demonstrate the developed research materials.

This study focusses on the demonstration of the developed digital tools and frameworks. The developed digital tools and frameworks can employ any degradation components; however, this study used a limited number of degradation components. The implementation of digital tools at the system level is beyond the scope of this research. In this context, algorithms for data-driven models are used in their most basic form to emphasise enhancing performance through the model management framework, rather than the algorithm itself.

The operational duration of the demonstration is restricted prior to maintenance activity, as the effects of maintenance on the machinery health condition fall outside the scope of this research.

1.6 Dissertation Layout

The layout of the dissertation is structured in eight chapters, an overview of which is provided as follows.

Chapter 1: Introduction to maritime digitalisation and autonomous shipping, as well as the aim, objectives, scope, and boundaries of the research.

Chapter 2: The literature review is contained in this chapter, the outcome of which includes the research gaps.

Chapter 3: The research approach and methodology are outlined in this chapter, including eight research objectives and six research phases. The objectives and phases are briefly described with figures.

Chapter 4: Methods and tools for developing trustworthy DT are described in this chapter. The development includes physics-based DT development, operating envelope extension, degradation model addition, DT integration, and calibration. In addition, a framework for ensuring the trustworthiness of the DT is proposed in this chapter.

Chapter 5: Methods and tools for developing data-driven PHM models are described in this chapter. A framework for the development and management of PHM models is proposed, including data-driven methods comparative assessment, data-driven sub-model development, RUL estimation & maintenance decision-making, and data-driven sub-model management throughout the asset lifetime.

Chapter 6: This chapter contains a description of the reference engine system and its anomalies, as well as case studies to demonstrate the tools and frameworks developed in Chapters 4 and 5.

Chapter 7: In this chapter, the results for each case study are presented and discussed to evaluate the effectiveness of the frameworks, tools, and PHM models that have been proposed and developed in the preceding chapters.

Chapter 8: This chapter summarises the research by highlighting its main findings, novelties & contributions, analysing the achievement of aim & objectives, and concluding with recommendations for future study.

Chapter 2

Literature Review

2.1 Overview

Figure 2.1 presents the keywords utilised in the literature review. An extensive literature review is performed using the keywords to identify research trends, state-of-the-art technologies, application challenges, and existing research gaps.

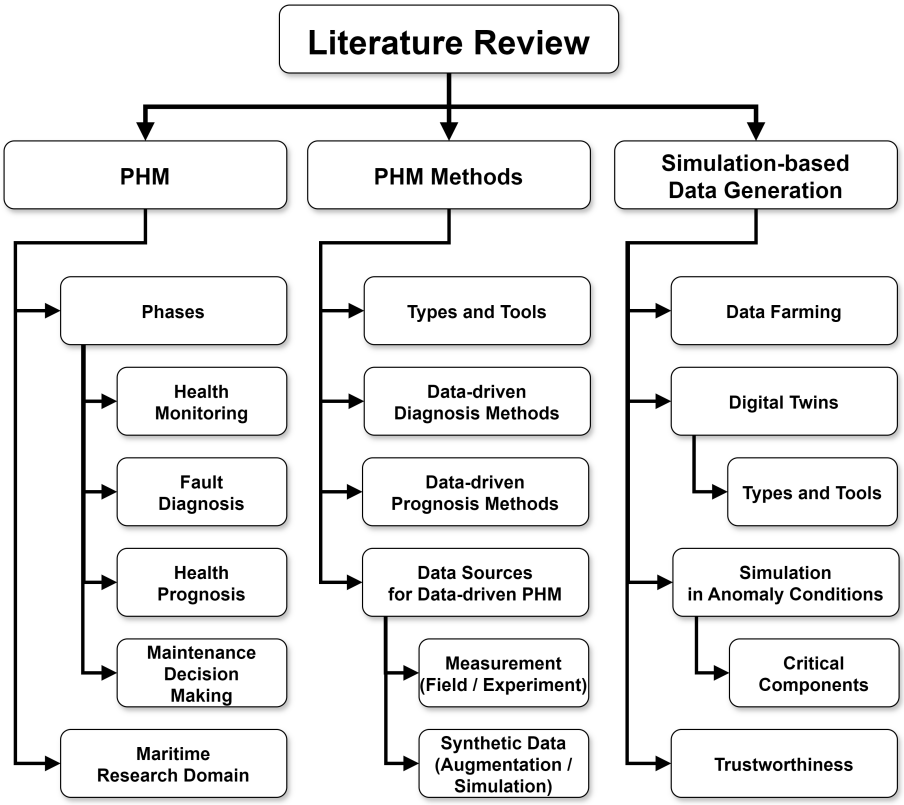


Figure 2.1: Literature review keywords

2.2 Prognostics and Health Management

The digitalisation of the maritime industry, coupled with the advancement in autonomous shipping technologies, is expected to produce considerable benefits, including improved efficiency, safety, and sustainability, as discussed in Chapter 1. In particular, the PHM system is one of the most promising digitalisation technologies within the maritime industry to support reliable operations without human intervention [59].

A PHM system estimates the current and future health of ship machinery [40], supporting optimal maintenance planning [41] and enhancing reliability with reduced downtime [42]. Early condition monitoring studies used data-driven models to extract anomaly features and build health boundaries to estimate the current machinery health condition by comparing new measurements with predefined boundaries [60]. Recently, PHM technologies have been improved and employed as a predictive decision support tool [9]. However, a wide operating envelope of ship machinery systems and stochastic degradation patterns are significant challenges for the implementation of data-driven PHM models in real-world operations.

2.2.1 PHM Phases

PHM systems include four phases, namely, health monitoring, fault diagnosis, health prognosis, and decision-making [40]. The health monitoring phase analyses sensory data at the current time slice to determine the system's health status. The fault diagnosis phase detects system anomalies, identifies anomaly types, and isolates anomaly locations. The health prognosis phase predicts the Remaining Useful Life (RUL) of the system. The decision-making phase involves maintenance and other operational decisions to manage the system's health condition.

Health monitoring models assess the system's health conditions continuously or periodically by analysing sensor measurements [61]. Tan et al. [62] employed six one-class classifiers to monitor the health status of a marine gas turbine propulsion system and tested them using simulation-generated labelled datasets. Vanem and Brandsæter [46] employed unsupervised clustering methods to monitor the health status of a marine diesel engine using unlabelled ship measurement datasets, however, validation

was not performed due to lack of labelled data. Oikonomou et al. [63] developed a real-time condition monitoring framework employing an outlier algorithm and testing its validity by using simulation-based degradation datasets for a ship propulsion system. Fu et al. [64] employed unsupervised algorithms to identify anomaly thresholds and monitor marine engine anomalies (filter blockage and cylinder leakage) using laboratory test measurements.

Fault diagnosis includes fault detection, isolation, and identification. Fault detection estimates the presence of faults and malfunctions, fault isolation determines the location of each fault, whereas fault identification estimates the severity and type of fault [65]. Zhan et al. [66] proposed a cascade framework based on one-class Support Vector Machines (SVM) to identify four fault types for the cylinder head of a marine propulsion engine by using vibration signals. Cheng et al. [67] employed a multi-class classifier with feature extraction methods to identify four faults and healthy operations for a marine vertical centrifugal pump by using vibration signals. Senemmar and Zhang [68] developed a deep learning-based fault diagnosis framework including fault detection, identification and isolation that was tested by using simulation-generated data of a ship power system. However, the data-driven models developed in these studies were tested in limited operating conditions, thus their performance cannot be validated when the systems are operated in different conditions.

Health prognosis estimates the health status of the component and/or system in future time slices, as well as calculates the system RUL [69]. Accurate RUL estimation is critical, as maintenance decision making relies on the calculated RUL [70]. Zhao et al. [71] proposed a remote fault prognostic and diagnostic expert system architecture for a marine gas turbine. Field measurements were transmitted to remote experts for real-time prognostic analysis. However, challenges pertaining to connectivity and availability of experts [29] undermine the effectiveness of this prognostic system. Tinga and Loendersloot [72] employed a physical degradation model for a marine diesel engine cylinder liner to predict RUL and facilitate predictive maintenance. Ma et al. [73] employed a multivariate Long Short-Term Memory (LSTM) to predict future degradation trends of an electro-hydrostatic actuator system. However, these health prognosis models, particularly when applied to longer

time horizon predictions, face the challenge of increase in prediction error, attributed to varying system operations and stochastic environmental conditions [74].

Decision making for health management deals with maintenance and operational actions, such as reconfiguration and changes in operating points, to extend the system RUL [75, 76]. Future smart and intelligent marine systems are expected to rely on prognostics-based decision-making for maintenance [77] to reduce operating costs [78]. Nguyen and Medjaher [41] proposed a dynamic predictive maintenance framework by integrating prognostics with maintenance decision-making and verified it for a turbofan engine. Do and Söffker [79] used prognosis results to optimise the operation and maintenance of wind turbines considering both the extension of the RUL and the increase in productivity. However, the system-level implementation of PHM in complex systems requires further development [80].

2.2.2 PHM Research in the Maritime Domain

In maritime research domain, numerous studies have tried to address the unique challenges associated with ship machinery operation and environment. Despite these efforts, ongoing challenges were recognised in the following studies.

Listou Ellefsen et al. [81] proposed a semi-supervised deep learning method to address the challenges associated with acquiring high-quality labelled data. This method improved accuracy with reduced labelled data, demonstrated using the public C-MAPSS dataset. However, the results also emphasised the essential importance of labelled data, indicating that the proportion of training labels significantly influences model error in both supervised and semi-supervised methods.

Han et al. [69] investigated marine engines in lab settings while accounting for maritime environmental conditions. The experiment applied two distinct operational profiles involving component faults and simulated load variations during autonomous voyages. The collected datasets facilitated real-time testing of a data-driven PHM model; however, artificial faults were rapidly induced, with experiments restricted to 20 minutes due to time limitations, highlighting data collection challenges in experimental contexts.

Velasco-Gallego and Lazakis [82] introduced an approach for RUL estimation tailored to marine machinery, aimed at supporting maintenance decision-making

processes. The study identified significant challenges, particularly the limited availability of anomaly data and the lack of methodologies for the practical application of the RUL estimation results. Data augmentation techniques, which expand the size of the data and generalise the model [83], were used to generate synthetic anomaly data; however, their trustworthiness was less considered. Furthermore, the generated dataset revealed that despite the application of augmentation methods, the operating envelope of the dataset remained unchanged, highlighting the constraints of these techniques.

Aizpurua et al. [84] presented a prognostic approach to estimate electric motor insulation degradation employing data-driven models to predict winding temperature using observed datasets (wind and vessel speed) and physics-based degradation models. The proposed approach was demonstrated by using shipboard measurements; however, the validity of the RUL model was confirmed only for limited operating conditions and a moderate degradation profile, whereas its testing in different operating conditions and considering stochastic degradation profiles was not performed.

2.3 Methods for PHM Models

PHM methods are typically classified into four categories: expert-based, model-based, data-driven, and hybrid methods. The expert-based method, which relies on expert knowledge and experience, involves transmitting field measurements to remote experts for real-time analysis [71], but it is vulnerable to connectivity and security challenges [29]. The model-based method employs a physics-based degradation model to predict future degradation trajectories. However, accurate predictions for complex systems are challenging because the performance of the model-based method is strongly correlated with the accuracy of the physics-based model [85]. The data-driven method relies on historical measurements to analyse degradation and predict future degradation trajectories [86]. This approach, however, demands extensive degradation data over prolonged periods, which is challenging to obtain. Insufficient data distort the results and reduce the efficacy of data-driven PHM models [87]. The hybrid method attempts to leverage the advantages by combining model-based and data-driven methods. However, limited studies have employed

the hybrid method due to the challenges in designing the appropriate integration mechanism [88].

Data-driven methods are prominently employed in recent research studies, supported by the advancements of sensor technologies and machine learning technologies [73]. The capability of these methods to capture intricate patterns in raw operational data [9] explains their increasing employment. The data-driven methods and approaches employed in pertinent research publications are comprehensively reviewed to determine applicable methods to the development of PHM models.

2.3.1 Data-driven Diagnosis Models

For fault detection models, Li et al. [89] employed SVM to detect engine oil leakage faults from experimentally acquired instantaneous angular speed data. Raptodimos and Lazakis [90] employed Artificial Neural Network and Self Organising Map (ANN-SOM) model to detect marine engine anomalies and demonstrate the model using shipboard measurement with a synthetic anomaly. Zhang et al. [91] employed Convolutional Neural Network (CNN) to detect a diesel engine misfire fault from experimentally acquired pressure and angular speed data. Vanem and Brandsæter [46] employed unsupervised methods including K-mean clustering, Gaussian Model, density-based clustering, SOM, and SVM to detect marine engine anomalies using shipboard measurement with a synthetic anomaly. Zhang et al. [92] employed binary adversarial autoencoder with Generative Adversarial Network (GAN) to detect marine engine faults through multivariate shipboard measurement data. The objective of the fault detection model is to classify health conditions as either healthy or faulty, commonly utilising binary classifiers with the unsupervised learning approach.

For fault isolation models, Lindahl et al. [93] employed an NN-based classification algorithm to detect and isolate grey water disposal system faults from electric power measurement data. Senemmar and Zhang [68] employed Deep Neural Network (DNN), Gated Recurrent Unit (GRU), and LSTM to detect, identify and isolate shipboard power system faults using simulation-generated voltage data. The fault isolation models frequently utilise NN-based multi-class classifiers. Determining location of faults in a multi-component system poses a significant challenge, as measurements are typically influenced by multiple components.

For fault identification models, Zhan et al. [66] employs multi-class SVM to identify cylinder cover faults using vibration signals of the engine. Basurko and Uriondo [45] employed ANN to identify multiple engine faults through multivariate sensory measurement data. Adams et al. [94] employed Random Forest (RF) and Classification Trees to identify multiple engine faults via experimentally acquired multivariate sensory data. Wang et al. [95] employed Manifold Learning and Isolation Forest to identify multiple engine faults using simulation-generated multivariate sensory data. Tan et al. [96] employed several multi-classification algorithms, including K-nearest Neighbour (KNN) and SVM, to identify multiple simultaneous faults with a publicly available simulation dataset. Cheng et al. [67] employed Weighted Kernel Principal Component Analysis and Particle Swarm Optimisation SVM to identify vertical centrifugal pump faults through experimentally acquired vibration data. Multi-class classifiers are frequently employed for fault identification models and demonstrated with data containing multiple faults.

2.3.2 Data-driven Prognosis Models

Three main approaches are generally utilised in the development of data-driven prognosis models: trend modelling with the Health Indicator (HI) [82, 97, 98], similarity-based curve matching [97, 99], and direct mapping [69, 74]. The trend modelling approach estimates the current health status and then predicts the future degradation trend for calculating RUL [100]. The trend modelling approach separates the prognosis process into the HI construction and HI forecast steps. The HI construction step transforms multivariate data into univariate HI values [101], while the HI forecast step employs the univariate HI to predict the HI of future time slices. The similarity-based curve matching approach builds a library of degradation trajectories and searches similar trajectories with the new observations from the library [100]. Simulations with a stochastic degradation model generate multiple trajectories that represent possible progressions of degradation [99]. However, the prediction error increases significantly when training trajectories are abundant or a testing trajectory has a short history [97]. The direct mapping approach involves training a prognosis model by linking input features to the output RUL using historical datasets, thereby enabling the straightforward calculation of RUL with

newly obtained input data [100]. Nevertheless, the error of the RUL estimation is substantially influenced by the quality of the training datasets and the appropriateness of the feature selection [101]. In the pertinent literature, the trend modelling approach is prevalent [100], due to its lower prediction error and good generalisation ability [101].

The HI construction step involves extracting the degradation status from the multivariate performance parameters, followed by the generation of statistical or quantitative HI [40]. The suitable method for developing the HI construction model varies according to the machinery types and data properties; however, three method groups are commonly employed in pertinent studies: statistical parameter-based, signal processing, and machine learning methods [102]. The statistical parameter-based methods are commonly used for the analysis of rotating machinery, extracting statistical parameters from signal trends. The signal processing methods are also commonly used for rotating machinery, specifically extract degradation information by accounting for cyclostationarity characteristics. Meanwhile, the machine learning methods are applicable to a broader range of objectives encompassing complex machinery systems and to diverse data types, including vibration signals and sensor measurements.

Among HI construction methods, machine learning-based methods are extensively used in pertinent studies [103]. Their popularity is supported by their strength in nonlinear data analysis, extensive applicability, and the independence for fault mechanism knowledge [102]. Yu et al. [101] constructed the HI of a turbofan engine and a milling machine by employing a bidirectional Recurrent Neural Network (RNN) based Autoencoder (AE). Nguyen and Medjaher [104] employed four common regression models including K-Neighbour Regression (KNR), Decision Tree Regression (DTR), Random Forest Regression (RFR) and Support Vector Regression (SVR) to construct the HI of turbofan engines and bearings. Koutroulis et al. [105] employed LSTM, Linear Regression, and Multi-Layer Perception (MLP) to construct the HI of turbofan engines. Huang et al. [106] proposed a HI construction method based on an AE and a LSTM to construct the HI of turbofan engines.

The HI forecast step predicts the future degradation trend using a constructed HI history and calculates the RUL based on a failure threshold [107]. Time-series

forecast models are frequently employed to predict HI in future time slices using time-series HI data. Wu et al. [108] employed the Autoregressive Integrated Moving Average (ARIMA) model to forecast the degradation trend of a rotating machine using vibration signals. Moura et al. [109] employed the SVR to predict the RUL of turbocharger failure in diesel engines and miles-to-failure for car engines using time series RUL data. Zhang et al. [110] employed the LSTM to predict lithium-ion battery degradation patterns. Ren et al. [111] employed DNN to predict the bearing RUL using vibration signals. Li et al. [112] employed CNN to predict the RUL of the aero-propulsion system using simulation-generated data.

The deterministic models employed in the above studies have demonstrated their effectiveness for RUL estimation. However, operational disturbances, including measurement errors and variations in operating conditions, can influence the performance of the HI forecast model in real-world applications [113]. To deal with these disturbances, uncertainty quantification becomes essential for the HI forecast step [84]. Nemani et al. [114] employed the ensemble approach to quantify errors and increase the accuracy of prediction models. Rivas et al. [115] employed the Bayesian approach with variational inference to estimate both the RUL and its uncertainty. Wei et al. [116] employed the Monte Carlo Dropout method to predict the RUL and its associated uncertainty, while mitigating heavy computation and addressing the model overfitting issue.

2.3.3 Data Sources for Data-driven PHM Models

The development of data-driven PHM models requires extensive datasets that represent a wide operating envelope of the machinery system, as data quality is of paramount importance for the performance of developed data-driven models [46]. However, a significant challenge persists among researchers in acquiring the appropriate datasets for the development of data-driven PHM models [44].

Shipboard datasets tend to contain errors, such as imprecision, bias, noise, human transcriptional errors, and missing data [40, 117]. Recently, quantum sensing technologies based on quantum effects have been investigated to address the limitations of conventional sensors. However, these technologies are still in the early stage of practice and require more time for commercial use [118].

Datasets for developing data-driven models in PHM systems are typically acquired from shipboard measurements, experimental data, or simulation results [107]. To support research on PHM systems, various experimental and simulation-based degradation datasets have been made publicly available; however, most of these datasets focus on specific tasks rather than considering the comprehensive health management task [119]. The required datasets include performance parameters, operating profiles, environmental conditions, and health indicators. The operating envelope of the datasets to train data-driven models significantly impacts the performance of these models when the actual operating conditions are significantly different compared to the training phase [120]. However, acquiring datasets corresponding to anomaly conditions is challenging, as most critical machinery systems on ships are effectively maintained throughout their lifetime [47].

Cheliotis et al. [121] employed measurements of engine exhaust gas temperature to develop a fault detection model for an ocean-going vessel propulsion engine. Velasco-Gallego and Lazakis [65] used measurement datasets from a diesel generator set to develop fault identification models based on five classification algorithms. In both studies, anomalies were injected into datasets by using the anomalous ratio, since the initial measurements only reflected healthy conditions. However, the reliability of these synthetic anomaly datasets is questionable, given that degradation patterns often deviate from a linear progression.

Han et al. [122] conducted a 10-day experiment using a marine engine on a research vessel with a clogged air filter to collect anomaly data and evaluate a proposed fault detection methodology. Wang et al. [123] experimented five typical marine engine faults and developed a performance prognostics framework. However, the experimental datasets did not represent the entire degradation phenomena and a wide operating envelope, as acquiring such datasets is associated with high costs and long timelines. The reliability of data-driven PHM models using datasets with a limited degradation process cannot be ensured for real-world operations that may encounter different degradation processes.

Coraddu et al. [47] employed simulation datasets derived by using a DT of a dual-fuel marine engine, experiencing air cooler degradation, to compare the accuracy and provide guidance on the effectiveness of several fault detection algorithms.

Altosole et al. [124] employed simulation-generated datasets for ten engine degradation scenarios to test an anomaly identification methodology, considering the impact of each anomaly on engine performance parameters. However, employing simplified coefficients to simulate each anomaly, without adequately accounting for the DT trustworthiness, introduced considerable uncertainty within the generated datasets.

According to the quantitative analysis from Ferreira and Gonçalves [125], the majority of PHM papers acquire datasets from the Prognostics Center of Excellence from the National Aeronautics and Aerospace Administration (NASA), which are publicly available simulation-generated datasets. However, reliable PHM models require datasets customised to specific machinery systems and operations. Measurement data often lacks comprehensiveness, and experimental data is often expensive. Consequently, there has been increased interest in simulation-generated anomaly datasets; however, the discrepancy in operating conditions between the development and application phases presents a substantial challenge to the effective implementation of data-driven PHM models [126].

2.4 Simulation-based Data Generation

To address the challenge in data acquisition, several researchers employed simulation-based data generation techniques, specifically the technique known as data farming. Data farming technique has been employed in the defence industry over the past decades to generate datasets using simulations [127]. Data farming includes modelling and simulation, high-performance computing, and statistical analysis [128]. Randomisation is commonly applied to the data farming to represent real-world uncertainty and fill the gaps between simulation outputs and real-world measurements [129]. The main applications of data farming are synthetic data generation for data-driven model development, as well as the analysis of generated data to identify critical factors and optimise system design [130].

Saxena et al. [44] recognised the scarcity of run-to-failure datasets and addressed this by generating degradation datasets with turbofan engine simulations. The generated datasets were originally intended for a PHM model development competition; however, they have been extensively used for training and testing PHM models after

the competition, primarily due to the persistent scarcity of run-to-failure datasets. Nikzadfar and Shamekhi [131] conducted simulations on 4,000 operating points of a diesel engine, generating datasets containing ten engine operating parameters. These datasets were then used to train a feed-forward neural network, as well as to investigate engine performance and emissions. Coraddu et al. [47] simulated a dual-fuel engine experiencing degradation in the charge air cooler to generate engine performance parameters, which were then utilised for comparing various fault detection algorithms. Bo et al. [132] simulated a diesel engine under anomaly conditions to generate engine performance parameters, which were subsequently used for training and testing fault diagnosis models.

The data farming technique improves the quantity and quality of data using DT and large-scale Design of Experiments (DOE) [130]. The DT is a tool for representing a selected physical system in a virtual environment [133], while the DOE is a systematic approach for exploring the system processes and gaining validity effectively and efficiently [134]. DT-based simulations replace expensive experimental campaigns [135] and supplement data for developing data-driven models [47].

2.4.1 Digital Twins

DT serves as a valuable tool for replicating the environment, physical processes, and embedded computations of a reference physical system [136]. To develop a DT, both measurement datasets and computational models of the reference system are required [133].

DT provides an alternative to costly physical experiments [137], but the development of complex system DT is challenging due to the presence of multiple sub-systems, their interactions, and their combined effects [138]. Furthermore, sensor malfunctions, including bias, drift, loss of accuracy, freezing, and calibration error, are prevalent in real-world operations [139]. Even sensors in good condition can transmit erroneous values due to sensor noise and thermal time constant, which denotes the required time to respond to 63.2% of the transition between the initial and final temperature states [140].

At the system level, DT requires the development of computational models for each component, as well as their integration [141]. To support the development of

complex system DT, co-simulation approach has been developed. This approach integrates various sub-modules into a master simulation platform, thereby reducing the integration efforts required [142]. Co-simulation enables comprehensively analysis of physical components (e.g. diesel engines, sensors, actuators) and cyber systems (e.g., controllers, alarm systems, operating panels), as well as communication networks within a single simulation [143]. This co-simulation approach has been applied to simulate complex maritime systems, including electric power management systems [142], navigation control systems [144], and network systems [143].

However, the co-simulation approach exhibits limitations under dynamic operating conditions, particularly in the presence of faults and degradation, as it simulates sub-module processes independently and exchanges their results at restricted communication points [145]. These communication discontinuities cause latency effects and module coupling errors that are non-existent in real-world operations [146]. Consequently, these discrepancies decrease the accuracy and reliability of the DT-based simulation and the generated datasets.

In the maritime research domain, DT has three typical modelling approaches: physics-based models (PM), data-driven models (DDM), and hybrid models (HM). PMs are based on first principles (e.g., laws of thermodynamics) and have the capability of adequately representing the machinery behaviours. Mean Value Engine Models (MVEM) and Zero-dimensional (0D) are the most widely used models for PMs due to their simplicity [147]. However, MVEM cannot simulate the variations in the in-cylinder process and 0D can only estimate simple parameters such as temperatures and pressures of the exhaust gas. One-dimensional (1D) and three-dimensional (3D) models are capable of simulating high-dimensional behaviours such as fluid flows and frictions. However, as the dimensions increase, DT becomes more complex and demands high computational effort [148]. DDMs do not require prior knowledge of the machinery systems; instead, they rely on an extensive amount of data. DDMs do not guarantee the causal relationship due to their black box characteristics but have strength in finding hidden or unknown patterns [149]. HMs combine PMs and DDMs to mitigate the drawbacks of each approach. Combining PMs and DDMs for HMs has three sub-approaches: PM-assisted DDM, DDM-assisted PM, and a hybrid approach with a strong and mutual interplay [149]. However, limited applications

are reported in the maritime research domain [150].

Tsitsilonis et al. [151] employed a physics-based marine engine model (0D) to generate performance parameters under both healthy conditions and prevailing conditions, facilitating the engine health assessment. Altosole et al. [124] employed a physics-based marine engine model (0D) to generate performance parameters considering ten degradation types, subsequently developing engine anomaly diagnosis models. Djeziri et al. [99] applied data-driven augmentation methods to investigate various potential pathways of transistor degradation, leading to the development of an offline RUL estimation model. Coraddu et al. [150] developed a marine engine hybrid model integrating MVEM and the Kernel Regularised Least Squares method to enhance the accuracy and interpretability of exhaust gas temperature prediction while minimising computational efforts.

2.4.2 Anomaly Simulation with DT

Developing degradation models and integrating them into the DT is essential for simulating faults and degradation processes. Degradation models can be developed based on either physical laws that describe degradation processes or measurement data that capture degradation phenomena [49].

Most degradation processes can be classified into three general models based on their shapes: linear, convex, and concave [152]. The general path models are widely used because of their simplicity; however, sometimes these models are not appropriate due to their oversimplification [49]. Degradation from major failure mechanisms (e.g. corrosion, fatigue, wear, and creep) can be effectively modelled using generalised meta-models, including Arrhenius, Archard, Paris' law, Eyring, and Inverse Power models [49, 153].

An appropriate stochastic process model depends on the specific characteristics of the degradation phenomena, broadly categorised into monotonic and non-monotonic degradation types. Monotonic degradation (e.g. wear and cumulative damage) generally uses the Gamma process and the Inverse Gaussian process, whereas non-monotonic degradation (e.g. material fatigue and corrosion) generally uses the Wiener process and Poisson process [153].

Chin and Katta [154] applied the global-chemistry-kinetics model to predict

fuel injector clogging degradation, subsequently validating the prediction through simulation data. [155] employed Archard's model to analyse cam wear degradation in valve trains, considering mixed lubrication, and validated the wear depth model with experiment-based measurements. Lewis and Dwyer-Joyce [156] developed a degradation law model for valve recession, deriving coefficients through laboratory measurements. Tung and Huang [157] utilised laboratory experiment-based measurements to develop a three-body abrasive wear model for predicting the piston ring and cylinder liner wear degradation. Eker et al. [158] developed a physics-based filter-clogging model using laboratory experiment-based measurements

Regarding critical equipment in ship machinery systems, the primary sources of failures within the vessel's system are the main engine and the propulsion mechanism [159]. Diakaki et al. [160] identified the fuel oil valve, the fuel oil pump, and the exhaust valve as the engine components with the highest failure rates. Similarly, Nahim et al. [139] identified the fuel supply system, the water system, and the valve system as the most frequently failing systems. Prioritisation should be given to frequently failing components and systems when incorporating degradation models within DT.

2.4.3 Trustworthiness of DT

Although DT-based simulations were employed in several research publications to address the limited data availability, the trustworthiness of their DT was only partially addressed. To develop a reliable data-driven PHM model using simulation-based datasets, the trustworthiness of the DT must be ensured comprehensively in advance.

Trustworthiness is a broad term, hence different users can interpret it differently [161]. Connett and O'Halloran [162] reviewed the trustworthiness definition from four perspectives: design, behavioural, physical, and anomaly perspectives. The design perspective implies that the system fulfils the set critical requirements. The behavioural perspective is related to confidence in the similarity between expectations and the system behaviour. The physical perspective implies that the trust pertains to the intended system functionality. The anomaly perspective implies that the system maintains its functionality despite the presence of disturbances/anomalies. Trustworthiness cannot be addressed in a single step; continuous assessments are

required during the system design, development, and operation phases. Schneider et al. [163] argue that a trustworthy system should maintain its trustworthiness regardless of constraints throughout the system's lifetime. The trustworthiness assessment of the developed DT is not a one-time occurrence, but it is performed iteratively as the DT undergoes variations.

The requirements for ensuring the trustworthiness of DT differ depending on the particular industry sector. For city automation, Wang and Burdon [164] proposed the following three elements to ensure DT trustworthiness: ability (functional quality), integrity (conformity to industrial standards), and benevolence (information privacy). For manufacturing automation, Babiceanu and Seker [165] considered dependability and cybersecurity to ensure the trustworthiness of the cyber-physical system. For systems and software engineering, de la Vara et al. [166] employed system architecture, multi-concern dependability, interoperability, and cross/intra-domain reuse to ensure the trustworthiness of cyber-physical systems in the development of the open-source platform. However, the trustworthiness of DT within the maritime research domain, particularly in simulation-based data generation, has not been explored.

2.5 Research Gaps

The literature review provides a comprehensive understanding of the methods and tools pertinent to the development of trustworthy DT and data-driven PHM models, specifically for marine engine applications. However, the following research gaps have been identified, requiring additional research:

1. While data-driven methods for developing data-driven sub-models have been widely researched in pertinent studies, the processes for selecting and integrating these methods for specific applications have rarely been reported. Lack of systematic approaches for selecting and integrating data-driven methods cannot guarantee the effectiveness of develop PHM models.

2. The performance of PHM models has been improved in recent studies using advanced data-driven methods; however, these model demonstrations predominantly use similar operating conditions for both the testing and training phases. Lack of a framework to retain the performance of the PHM model regardless

of the varying operating conditions raises concerns about the trustworthiness of such data-driven PHM models.

3. Multiple studies have reported the use of simulation datasets employing DT to substitute measurement data under anomaly conditions; however, the degradation modelling process was not considered. Inefficient approach for developing DT under anomaly conditions results in improper data generation and subsequent failure in training data-driven models.

4. In several studies, the DT is evaluated by comparing simulation results with a limited set of measurement data; however, this method only validates the DT under limited operating conditions. Ensuring the trustworthiness of the DT cannot be accomplished through a single step and requires a comprehensive assessment. The lack of a framework to ensure the trustworthiness of the DT results in inaccurate simulation data generation, thereby raising concerns about the credibility of PHM models that employ these simulation-generated datasets.

5. The operating environment of ship machinery systems is inherently dynamic and uncertain; however, several studies test their PHM models within moderate degradation scenarios. Additionally, the measurement data of a complex machinery system capture the coupled behaviour of multiple components and their interactions; however, this aspect is frequently overlooked in pertinent studies. Lack of PHM method tests in complex degradation scenarios, including simultaneous degradation in multiple components and stochastic degradation patterns, remains a concern about the performance of PHM models in real-world applications.

6. Most studies in the PHM research domain primarily address accurate health monitoring, diagnosis, and prognosis. However, there is a noticeable gap in the integration of this health information into the decision making phase. Lack of a framework to apply the health information derived from PHM models in the decision making phase leads to inefficient PHM system applications.

7. The quality and quantity of data are of paramount importance for the perfor-

mance of data-driven PHM models; however, no method has been reported in pertinent studies to assess the suitability of data for the development of PHM models. Lack of methods to identify the need for data generation using DT-based simulation results in either inaccurate data-driven PHM models.

8. Fault diagnosis models typically employs classification algorithms. For classifiers that utilise decision boundaries, small differences near the class boundary can result in incorrect classification. To address the dynamic operating environment of ship machinery systems, sensitivity analysis is required for fault diagnosis models. Lack of methods to identify decision sensitivity undermines the trustworthiness of the fault diagnosis model in real-world operations.

9. The shipboard measurement data contain various types of sensing errors, including noise and missing data. The trustworthy PHM system should provide consistent results despite these sensing errors. Lack of methods to address sensing errors impedes the implementation of PHM systems in real-world operations.

This study addresses five of the nine research gaps identified from the extensive literature review, concentrating on advancing the PHM system. Novel frameworks will be developed to address these gaps, with the digital tools in the following chapters. The unaddressed gaps, still vital to improving the efficiency of the PHM system, are expected to be investigated in future research.

Chapter 3

Research Approach and Methodology

3.1 Research Approach

To address the aim of the research, eight objectives are investigated, including literature review, reference system & anomalies, physics-based DT development, trustworthiness assurance framework, simulation-based data generation, data-driven PHM model development, data-driven PHM model management framework, and digital tools & frameworks demonstration. Figure 3.1 illustrates each objective with their connections.

A comprehensive literature review (Obj. 1) systematically collects information on the reference physical system along with its anomalies (Obj. 2), while also identifying state-of-the-art methods for the development of physics-based DT (Obj. 3) and data-driven PHM model (Obj. 6). A physics-based DT (Obj. 3) is developed considering various operating profiles of the reference system (Obj. 2), with its trustworthiness being ensured through a trustworthiness assurance framework (Obj. 4). Trustworthiness-assured DT generates simulation datasets encompassing an extended operating envelope (Obj. 5), thus facilitating the development of the data-driven PHM model (Obj. 6) and the model management framework (Obj. 7), as well as their demonstration (Obj. 8).

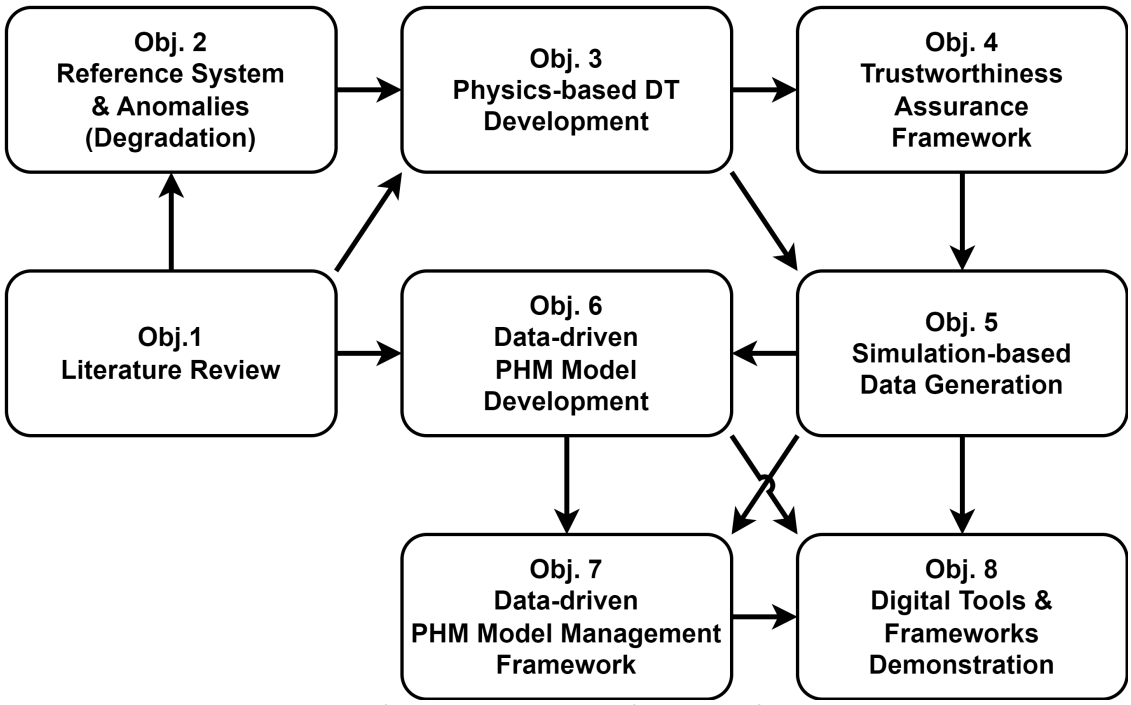


Figure 3.1: Research approach

3.2 Research Methodology

Figure 3.2 illustrates an integrated research methodology to achieve the research aim and objectives. Reference system data along with operating envelope and anomaly summary table serve as essential inputs for the methodology. The resulting outputs comprise a trustworthy DT, simulation-generated datasets, and PHM model.

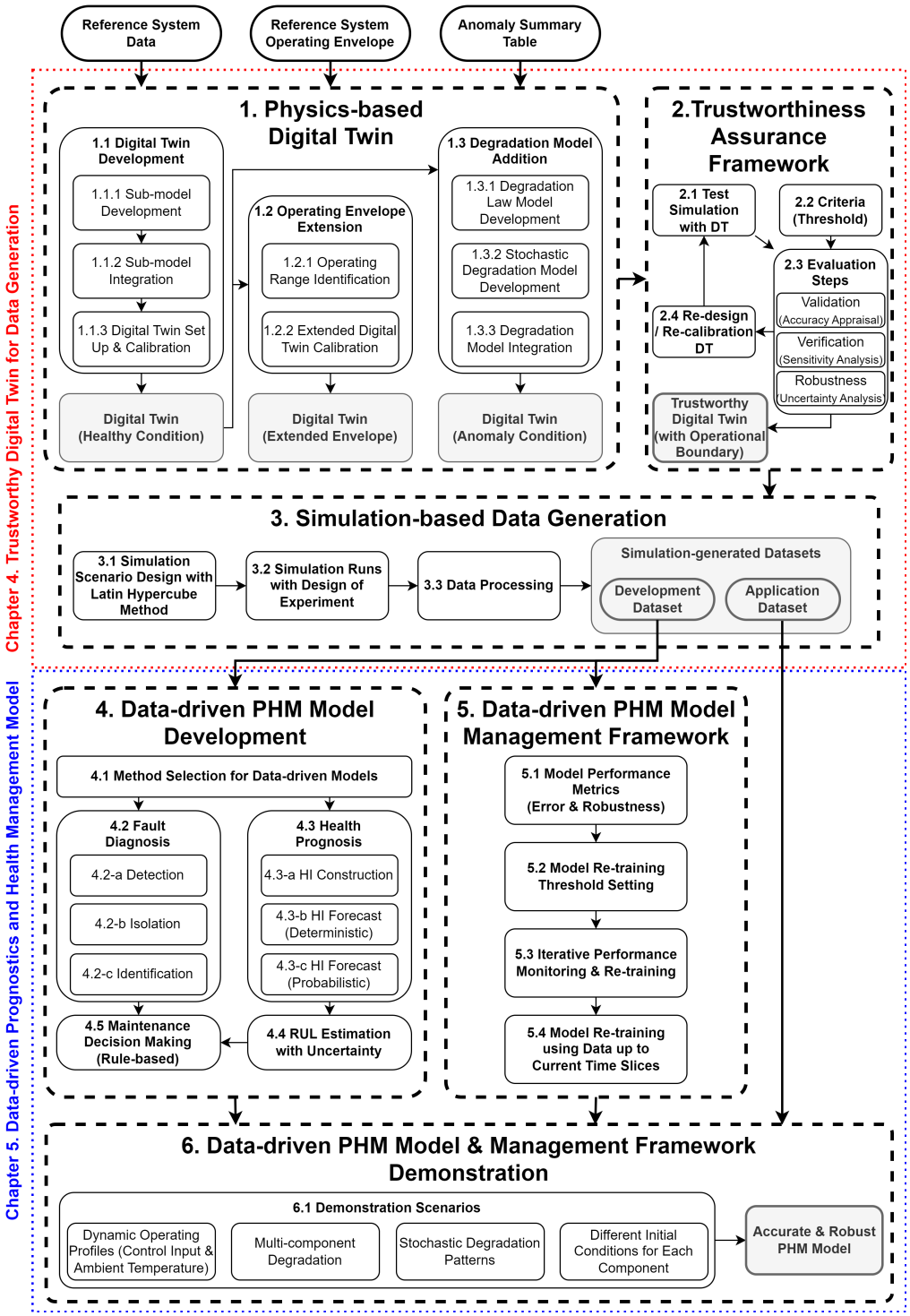


Figure 3.2: Research methodology

The first phase focuses on the development of a DT based on physics-based models. The physics-based model is employed for the development of DT due to its capability to accurately simulate the intricate operations of the engine [150]. Engine sub-models are developed, integrated, and calibrated to simulate a DT of the reference system under healthy conditions. The operating envelope of the DT is expanded considering

the operating range of the reference system, and the extended DT is re-calibrated to simulate the DT in the extended operating conditions. Component degradation models are added to the developed DT by employing the degradation law model and stochastic degradation model. The DT with degradation model is integrated and re-calibrated to simulate the DT under anomaly conditions.

Phase 2 develops a framework for assuring the trustworthiness of the developed DT. This phase is critical within this research methodology, as the trustworthiness of DT is often overlooked in related studies. Test simulations are executed to comprehensively evaluate the trustworthiness of the DT by employing three evaluation steps (validation, verification, and robustness) coupled with pre-set thresholds. The validation step compares the DT-generated performance parameters against available measured reference data. The accuracy appraisal with a pre-set error threshold ensures that the DT is accurate in the tested operating conditions. The verification step performs sensitivity analysis to investigate engine performance parameter trade-offs considering input parameters changes. These trade-offs are compared to the reference ones published in the pertinent literature to ensure that the DT behaviours align with the pertinent literature. The robustness step performs uncertainty analysis, quantifying the uncertainty ratio of input and output parameters. The uncertainty ratio less than one ensures that the DT does not contribute to uncertainty propagation. If any criteria are not satisfied, the DT necessitates re-design or re-calibration to ensure its trustworthiness. The satisfaction of all criteria confirms that the DT is trustworthy within the operational boundary considered during the trustworthiness evaluation.

In Phase 3, the trustworthy DT generates simulation datasets by executing several simulation runs. The simulation scenarios are designed using the Latin hypercube method to effectively represent a wide operating envelope. The design of experiment method is employed to systematically simulate multiple scenarios, subsequently processing the simulation datasets to generate development and application datasets, which facilitate the development and demonstration of the data-driven PHM model, respectively.

Phase 4 develops a PHM model employing data-driven methods due to their ability to capture complex degradation patterns without necessitating expert knowledge [9].

Algorithms identified through the literature review are systematically compared using development datasets that include various operating profiles reflecting the unique characteristics of real-world ship operations. With the selected algorithms, fault diagnosis models and health prognosis models are developed. The fault diagnosis model includes three sub-tasks (detection, isolation, and identification) to diagnose fault information with anomaly components and locations. The health prognosis model includes three sub-tasks (HI construction, HI forecast with deterministic approach, and HI forecast with probabilistic approach) to calculate the RUL of the anomaly component with uncertainty. Maintenance decisions are made based on the results of fault diagnosis and prognosis models to determine an effective maintenance schedule.

In Phase 5, a framework for data-driven model management is developed to evaluate the performance of the health prognosis model and re-train it, thereby maintaining its continued effectiveness. The decision of model re-training is made when its performance falls below pre-defined thresholds. The management process operates as an iterative process throughout the lifetime of the reference system, and the re-training utilises data acquired up to current time slices. This framework is essential for developing a trustworthy PHM model that addresses the inherent stochasticity in degradation patterns [167]. Selecting appropriate performance metrics and re-training threshold is crucial for efficient model management.

In Phase 6, the developed data-driven PHM model and its management framework are demonstrated using application datasets derived from Phase 3. The demonstration scenarios examine the challenges for the implementation of the data-driven PHM model, including dynamic operating conditions induced by variations in control inputs and ambient temperatures, multi-component degradation processes, stochastic degradation patterns, and different initial conditions for each component. The results demonstrate the error and robustness of the PHM model within the operating envelope tested. The details of Phases 1, 2, and 3 are described in Section 4, while those of Phases 4, 5 and 6 are described in Section 5.

Chapter 4

Trustworthy Digital Twin for Data Generation

4.1 Physics-based Digital Twin

The physics-based engine DT is developed by using GT-SUITE v2022 software, which provides a 0D/1D simulation interface with libraries for the system components and sub-models [168]. The DT is developed based on the previous studies [169, 170]; however, in this study, the DT is extended by the inclusion of the modelling of sensors and their dynamics for the engine performance parameter measurements. The sensor sub-model monitors and collects virtual signals from the simulations representing the sensors deployed in engine operations. The layout of the modelled engine components, along with the sensors and monitoring system block, is illustrated in Figure 4.1.

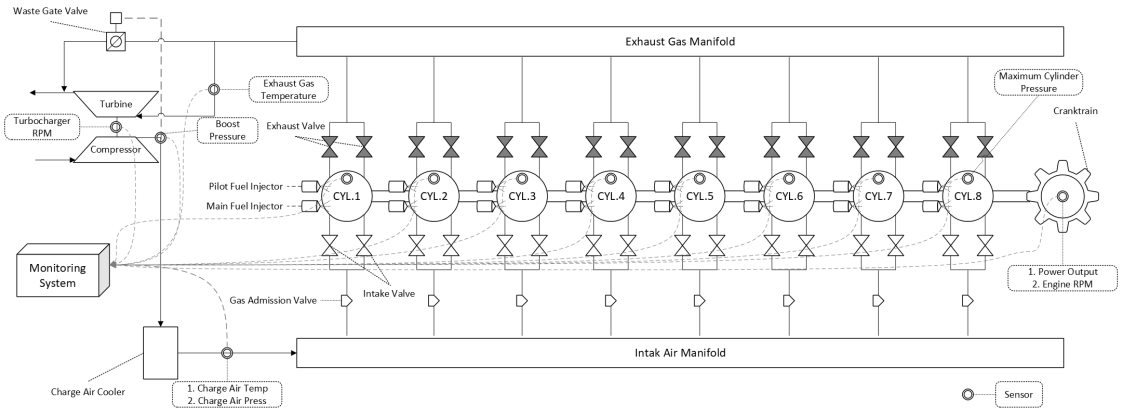


Figure 4.1: Marine engine components and sensors modelled in the developed DT

This study considers the four-stroke, non-reversible, turbocharged, and intercooled Wärtsilä 8L50DF engine for developing the engine model, which consists of several sub-system models for representing the engine components. The Woschni equation [171] is employed for representing the in-cylinder heat transfer process, whereas the single Wiebe function [172] is used for modelling the combustion process. The fresh air and exhaust gas flow sub-models are based on one-dimensional approaches, which consider the fluid momentum, mass, and energy conservation equations. The engine control sub-system models for fuel supply systems and wastegate valve operations employ the Proportional–Integral–Derivative (PID) controllers.

In addition to the engine component sub-models, the DT incorporates a sensor sub-model. Sensors play a critical role in marine engine operations, as they measure physical quantities such as torque, pressure, temperature, and fluid flow. However, sensor values are significantly affected by the surrounding environment [173], leading to inevitable sensing errors at the current technology level. Sensor faults commonly include issues such as bias, drift, loss of accuracy, freezing, and calibration error [139]; however, even healthy sensors have erroneous values because of sensor noise and thermal time constant.

This study employs an additive noise model to introduce random error values into simulation results, simulating dataset irregularities akin to sensor noise effects on measurements. The additive noise model simplifies and represents the sensing errors such as offset, drift, spikes, and calibration errors. The additive noise model is developed based on Equation 4.1 [174], and the random error values are generated by a Gaussian random process.

$$y_{meas}(t) = y_{act}(t) + x(t) \tag{4.1}$$

where $y_{meas}(t)$ denotes sensor measurements at time t, $y_{act}(t)$ is actual physical quantities at time t, and $x(t)$ is random error at time t.

4.1.1 Operating Envelope Extension

The typically accessible data from the engine shop tests are collected under controlled conditions; however, the engine is operated in a wide operating envelope in real-

world operations. Data farming significantly enhances the dataset by extending the operational envelope.

This study considers two operating parameters to extend the operating envelope: ambient temperature variation and engine load variation. The operating range of ambient temperature is set to 15–45°C, according to the engine manufacturer's guidelines [175]. During simulations, ambient temperature values are randomly selected using a probability density function derived from 3-year engine operation records. Engine load percentages are also randomly selected from 70%, 75%, 80%, and 85%, which are the most efficient operating range for the investigated marine engines [176]. To avoid frequent load variations, the selected engine load values are not changed for a randomly selected load duration within the range of 60–600 hours.

4.1.2 Component Degradation Model

The data farming can also generate datasets under anomaly conditions, which are infrequent event and challenging to obtain from shipboard measurement data. DT-based simulations under anomaly conditions eliminate the problems of expensive testing cost, experiment preparation, engine component damage, spare parts waste, and pollution. However, a thorough understanding of engine anomalies is required to appropriately simulate the degradation process.

Degradation models can be classified into the following two categories: Degradation Law Models (DLM) and Stochastic Process Models (SPM). DLM employs basic laws for describing the degradation using deterministic mathematical equations. SPM reconstructs the degradation pathway using measurement data [153]. However, complex degradation processes cannot be adequately represented by a single mechanism [153]. This study integrates these two degradation model types to develop a comprehensive degradation sub-model. The laws model are employed to characterise the generic behaviour of the engine components degradation observed over a long-term period. In contrast, the stochastic model characterises short-term variations during the operations, which represent the temporal dynamics of engine component degradation.

Table 4.1: Anomaly summary table format

Anomaly Identification Process with FMECA										**Anomaly Analysis for Simulation Design**				
FMECA Output										DT Input parameters				DT Output parameters
Component	Function	Failure Mode	Failure Causes	Failure Effects	Detection Method	O	S	D	RPN	Name	Manufacturer limits	Input Range	Simulation Steps	Symbol

Anomaly Summary Table

An anomaly summary table collects the necessary information to model the anomaly components. The anomaly table is developed by employing identification and analysis processes as presented in Table 4.1.

The identification process employs the Failure Mode, Effects, and Criticality Analysis (FMECA) method to identify the failure modes, failure causes, failure effects, as well as the criticality of these failures [177]. The anomaly criticality is evaluated by the Risk Priority Number (RPN), which is derived by the multiplication of occurrence level (O), severity level (S), and detectability level (D), according to the following equation [178]:

$$RPN = OSD \tag{4.2}$$

In the analysis process, the following parameters are determined: anomaly control parameters (DT input), dependent parameters (DT output), manufacturer's limits, ranges of the pertinent input parameters, and steps; the latter is derived from the manufacturer's limits, ranging from healthy to severe anomaly conditions.

Valve Recession Model

This study considers the wear degradation of the engine cylinder valves, which is one of the most frequent anomalies in marine engines [139]. The valve recession model reported in Lewis and Dwyer-Joyce [156] is employed, which was developed based on laboratory experiments for the wear of the automotive engine valves. The wear

volume is calculated according to Equation (4.3).

$$V = \left(\frac{k\overline{P}N\delta}{h} + KNe^{n}\right)\left(\frac{A_i}{A}\right)^{j} \tag{4.3}$$

where V denotes the wear volume, k is the sliding wear coefficient, $\overline{P}$ is the average contact force at the interface, N is the number of cycles, δ is the slip at the interface, h is the seat hardness, K is the impact wear constant, e is the energy on impact, n is the impact wear constant, A_i is the valve's initial contact area, A is the contact area after N cycles, and j is the wear constant.

The valve recession is calculated from the wear volume according to Equation (4.4).

$$r = \left(\sqrt{\frac{V}{\pi R_i cos\theta_s sin\theta_s} + w_i^2} - w_i\right) sin\theta_s \tag{4.4}$$

where r denotes the valve recession, R_i is the initial seat insert radius, θ_s is the seat insert seating face angle, and w_i is the initial seat insert seating face width.

It must be noted that the effect of sliding wear is not considered herein, as it is negligible for valves with high closing velocities [179], such as cylinder valves of large marine engines.

Stochastic Process Model

The degradation parameters for valve recession, obtained from DLM, are adjusted with the SPM. The SPM captures the stochasticity of the degradation and represents various disturbances, including other component anomalies and the spare parts quality. However, acquiring abundant degradation data is a significant challenge. To represent the stochasticity of the degradation, this study employs the gamma process, as its irreversible and continuous characteristics [49, 153] closely correspond to wear degradation patterns. The probability density function (PDF) of the gamma distribution is defined according to Equation(4.5) [180].

$$Ga(x|v,u) = \frac{u^{v}}{\Gamma(v)} x^{v-1} exp\{-ux\} I_{(0,\infty)}(x) \tag{4.5}$$

where $Ga(x|v,u)$ denotes a random variable x following the gamma distribution with shape parameter v and scale parameter u, $\Gamma(a) = \int_0^\infty z^{a-1}e^{-z}dz$ for $a > 0$, and $I_A(x) = 1$ for $x \in A$ and $I_A(x) = 0$ for $x \notin A$.

The valve recession at each time interval is calculated by employing Equation (4.4). The stochasticity to valve recession is introduced considering Equation (4.5), leading to the estimation of total valve wear, which is calculated according to Equation (4.6).

$$W(t_n) = \sum_{t=0}^{t_n} r\ x(t) \tag{4.6}$$

where W denotes the total valve wear (mm), t_n is the running time after n time intervals, r is valve recession (mm) at each time interval, whereas $x(t)$ is a multiplier, which takes random values according to Equation (4.5).

The maximum valve lift is subsequently calculated by employing Equation (4.7).

$$L_{max}(t_n) = L_{max}(0) - W(t_n) \tag{4.7}$$

where L_{max} denotes the maximum valve lift (mm) at t_n, $L_{max}(0)$ is the maximum valve lift (mm) at healthy conditions, and $W(t_n)$ is the total valve wear (mm) at t_n.

The multiplier x at each time step is sampled using a random number generator from the Python NumPy library. The random number generator employs the shape and scale parameters to define the PDF of gamma process, and the size parameter to define the number of samples. This study employs gamma distributions with two different stochasticities, both having the x mean value equal to 1 to retain the basic degradation laws. At each time step, the sampled random variables are used as multiplication factors to adjust the valve recession calculated by the DLM. The adjusted valve recession is accumulated over time to calculate total valve wear at a specific running time. This total valve wear then modifies the valve lift profile by limiting the maximum valve closing.

4.1.3 Digital Twin Integration & Calibration

By using the GT-SUITE software, the thermodynamic model, the degradation model, and the sensor models are integrated to develop a physics-based DT, which is

subsequently used to generate extensive envelope datasets. Using the running time, initial conditions, and parameters of the Gamma PDF as input, the degradation model calculates the maximum valve lift at each time interval (step). This is fed as input to the thermodynamic model, which subsequently simulates the engine performance parameters.

The sensor model receives the performance parameters and applies the time delay and noise of the considered sensor, thereby generating measurable engine performance parameters as output. The digital twin constituents, interconnections, and information flow are illustrated in Figure 4.2.

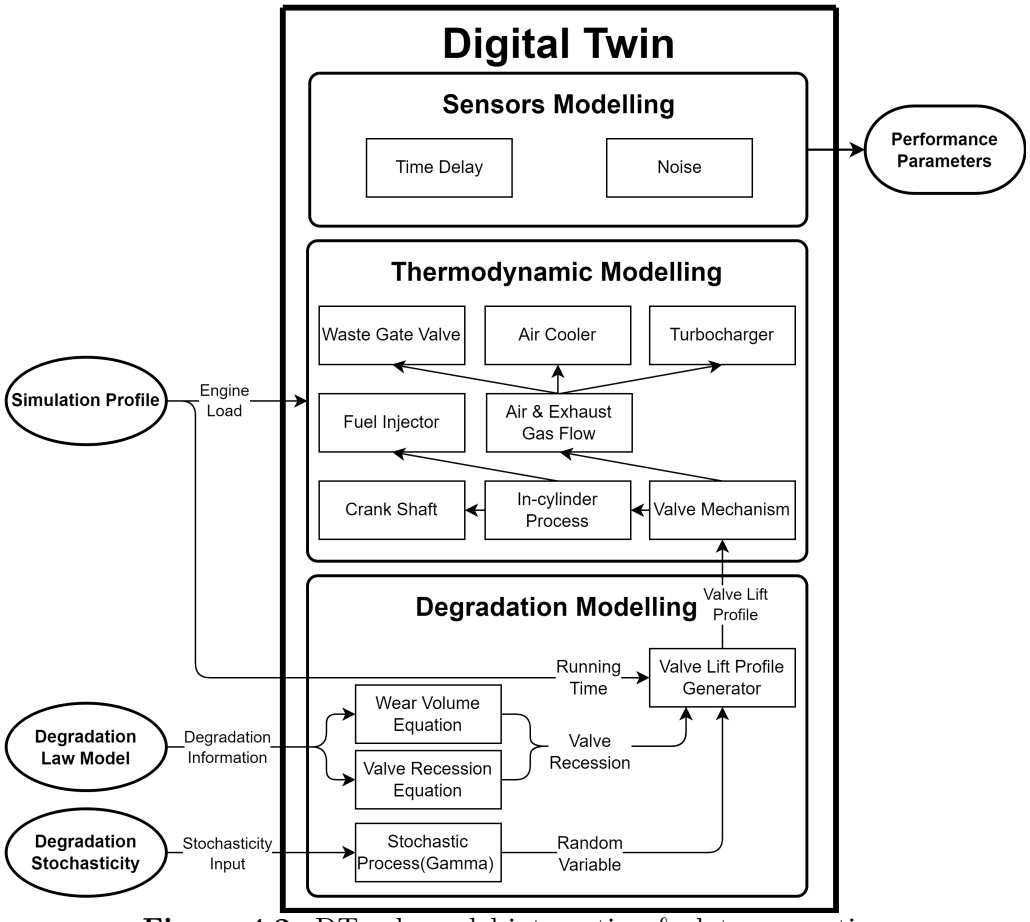


Figure 4.2: DT sub-model integration & data generation

To simulate the marine engine DT within the extended envelope, the DOE [134] with the Latin Hypercube sampling method is employed to reduce the number of samples, whilst retaining a sufficient representation of each input variable [181].

The DT sub-models are calibrated by employing an optimisation method, which uses the reference performance parameters and calculates the optimal values for the sub-model constants to achieve the lowest value of an objective function [151, 182].

GT-SUITE software provides the Advanced Direct Optimiser (ADO) that employs the Non-dominated Sorting Genetic Algorithm (NSGA-III) [183]. In this study, the ADO optimises turbocharger turbine mass flow and efficiency coefficients at 75% load using an objective function (weighted sum of errors) associated with three performance parameters (turbocharger shaft speed and turbine inlet and outlet temperature). An example for calibrating the turbocharger turbine model is shown in Table 4.2.

Table 4.2: Example of using the advanced direct optimiser to calibrate the turbocharger turbine model

Factors			**Responses**			
Parameter	Turbine Mass Flow coefficient [−]	Turbine Efficiency [−]	Parameter	TC Shaft Speed $[rev/m]$	Exhaust Gas Temperature Upstream Turbine $[K]$	Exhaust Gas Temperature Downstream Turbine $[K]$
Search Range	$0.8 - 0.99$	$1.0 - 1.12$	Target Value	16690	749	594
Optimal Value	0.9354	1.0753				

4.2 Trustworthiness Assurance Framework

A framework for assuring the trustworthiness of physics-based DTs is illustrated in Figure 4.3. It consists of three phases (listed in numerical order), whereas each phase includes several steps.

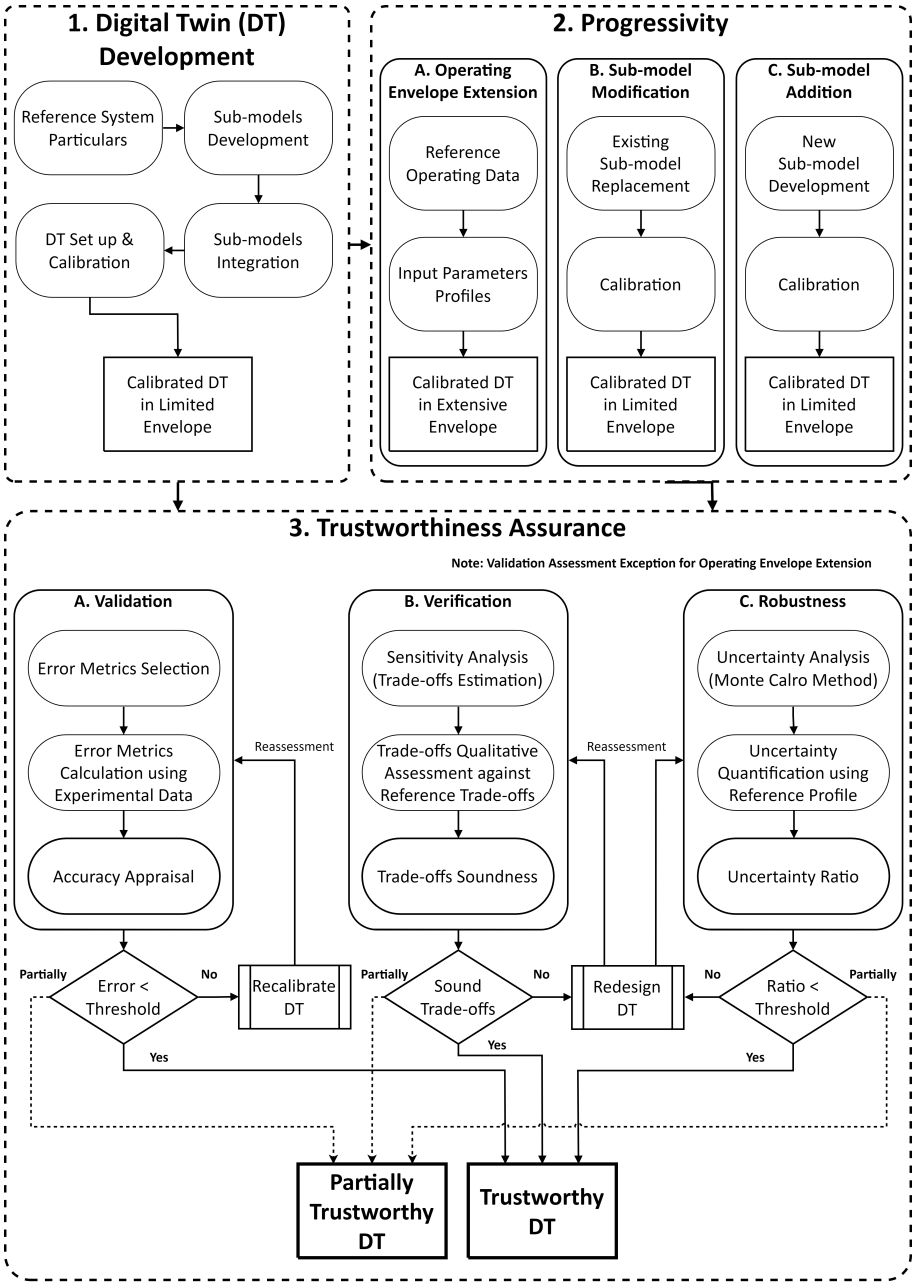


Figure 4.3: Systematic framework for assuring the trustworthiness of physics-based DT

The first phase focuses on the sub-model developments and their integration to develop the DT. The DT is subsequently calibrated by adjusting the sub-model coefficients/constants to achieve acceptable accuracy.

Phase 2 deals with the progressivity (otherwise called non-regressivity) of the DT. This phase includes three options: (a) extending the operating envelope, (b) modifying/replacing existing sub-models, and (c) adding sub-models. Option (a) enables simulation of engine anomaly conditions or/and varying ambient conditions. The ranges of input operating parameters are set considering the acquired reference

operating data. Option (b) deals with the modification or replacement of the initially employed sub-models. Option (c) caters for new sub-model additions, such as component degradation sub-models or/and auxiliary system sub-models. Options (b) and (c) require re-calibration of the DT to achieve acceptable accuracy.

In Phase 3, the trustworthiness of the DT is assured through three steps: validation, verification, and robustness. The validation step employs error metrics to calculate the accuracy of the DT outputs against experimental data. If the estimated errors exceed predefined thresholds, the DT re-calibration and reassessment are required. The verification step estimates the performance and emissions parameters trade-offs, and assesses their soundness by qualitatively comparing to the reference trade-offs. When the estimated trade-off is not sound, the DT redesign is required, followed by repeating the verification step. The robustness step calculates the uncertainty ratio between the DT output and input parameters, using input parameter reference profiles. When the estimated uncertainty ratio exceeds predefined thresholds, the DT redesign and repetition of the robustness step are required. When the DT re-calibration and redesign activities do not result in improved trustworthiness, it is advisable to stop the trustworthiness assurance phase and return to the DT development phase.

Completing these three steps assures the DT trustworthiness for the considered operating envelope. In cases where the validation, verification, and robustness criteria are partially fulfilled (e.g., only for specific performance parameters), partial assurance of the trustworthiness should be indicated. In cases where the DT version (Phase 1) fulfils all step criteria, the operating envelope extension option (Phase 2-A) does not require repetition of the validation step (Phase 3-A).

4.2.1 Trustworthiness Steps

This study develops a comprehensive method for assuring the trustworthiness of the marine engine DT, considering the following three steps: validation, verification, and robustness. Each step partially ensures the trustworthiness of the DT due to limited data and information; however, the integration of these steps provides comprehensive assurance. Pre-set decision criteria for each step are decided based on the DT requirements, which are described in the following subsections.

Validation

The validation step depends on available measured data, which are expected to be limited. For marine engines, the engine shop tests (or factory acceptance tests) and ship trials provide the engine performance parameters under healthy conditions at limited operating points. Shipboard data acquisition systems are a recent trend in the maritime industry; however, their use is limited. The limited data restrict the validation, and thereby the validation ensures the trustworthiness of the developed DT only within the limited operating envelope where the measured dataset are available. The validation of the DT is based on the comparison between the predicted performance parameters and available measured datasets. The criteria employed for the validation step involve the percentage errors between the predicted and measured parameters, as calculated by the following equation:

$$E = \frac{OP - RD}{RD} 100 \tag{4.8}$$

where E denotes the percentage error, OP is the output parameter, and RD is measured value of the same parameter.

The acceptance limits for the error of performance parameters (pressure, temperature, rotational speed, and fuel flow) are based on the guidelines from the ISO standards applicable to engine testing measurements [184]. If the estimated percentage error is smaller than the guideline's allowance, the DT is considered validated. For cases where validation of only specific parameters is achieved, the DT is considered partially validated.

Verification

The verification step deals with identifying the DT behaviour considering several engine operating conditions. However, access to extensive measurements under specific operating conditions is limited. Instead, the evaluation is based on consistency with expected trade-offs from the pertinent literature. The verification step is based on sensitivity analysis to evaluate the influence of the input parameter changes on the output parameter variations [185]. The Spearman rank correlation coefficient [186, 187] is employed to quantify the strength and direction of the relationship

(trade-off) between the input and output parameters. The Spearman rank correlation examines sample rankings rather than their values, unlike the Pearson correlation, facilitating the analysis of trendability and trade-off. This coefficient is calculated by Equation (4.9), which results in a range from -1 (strong reciprocal correlation) to $+1$ (strong direct correlation) with 0 representing negligible correlation.

$$r_s = 1 - \frac{6\Sigma \left(R\left(X_i\right) - R\left(Y_i\right)\right)^2}{N\left(N^2 - 1\right)} \tag{4.9}$$

where r_s denotes the Spearman's coefficient, X is the input parameter values, Y is the output parameter values, N is the number of samples, and R denotes the ranked variables.

This study takes into account varying operating conditions for ambient temperature as well as the anomaly conditions. The Latin hypercube method samples the combinations of these three parameters. The DOE tool of the GT SUITE is employed to perform the simulations and predict the trade-offs of performance parameters for each combination of input parameters. If the estimated trade-offs for the DT performance parameters match the trade-offs from the pertinent literature, the DT verification is assured.

Robustness

The last step of the trustworthiness assurance deals with the evaluation of the DT's robustness, which is based on the predicted performance parameter uncertainty considering anomaly conditions and varying environmental conditions (ambient temperature in this study).

Uncertainty analysis is employed to measure the probability of unexpected event occurrences [188]. This study employs the variance for quantifying the uncertainty [189] and uses the Monte Carlo (MC) method [190] to estimate the variance of the DT output performance parameters. The input (anomaly parameters and ambient temperature) and the output (predicted performance parameters) are both normalised within the range of [0, 1] (to eliminate scale differences between parameters) by employing the min-max scaling method according to the following Equation (4.10)

[191]:

$$X' = \left(\frac{X - X_{min}}{X_{max} - X_{min}} \right) (M - m) + m \tag{4.10}$$

where X' denotes the normalised data ranging within 0–1, X denotes the original data, X_{min} is the minimum value of the original data, X_{max} is the maximum value of the original data, M is the pre-defined maximum range value (1), and m is the pre-defined minimum range value (0).

The variances of the input and output parameters are calculated according to the following equation:

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)}{n - 1} \tag{4.11}$$

where S^2 denotes the variance, n is the sample number, X_i is the value of the ith sample, and $\bar{X}$ is the mean value of the samples.

The uncertainty ratio between the input and output parameters is calculated using the following Equation (4.12):

$$R_m = \frac{S_m^2}{S_{in}^2} \tag{4.12}$$

where R_m denotes the uncertainty ratio, S_m^2 is the variance of the mth predicted performance parameter, and S_{in}^2 is the variance of the input parameter.

The acceptance criteria for the DT robustness are established based on a requirement that the DT should not introduce additional uncertainty or propagate the uncertainty (from input to output parameters). This requirement implies that the uncertainty of the normalised output parameters should not exceed the uncertainty of the normalised input parameters. Hence, uncertainty ratio values below 1 indicate that DT robustness is assured.

From the results of these three steps, the calculated metrics are summarised in a tabular format, as presented in Table 4.3. The trustworthiness table provides an intuitive assessment, this supporting decision making.

Table 4.3: Trustworthiness assurance table

Steps	Acceptance Criteria	Trustworthiness Checks
Validation	Acceptable Errors	Pass/Fail
Verification	Trade-offs Soundness	Pass/Fail
Robustness	Uncertainty ratio	Pass/Fail

4.3 Simulation-based Data Generation

A framework for generating needed datasets is illustrated in Figure 4.4. The framework employs the developed trustworthy DT framework and consists of four stages.

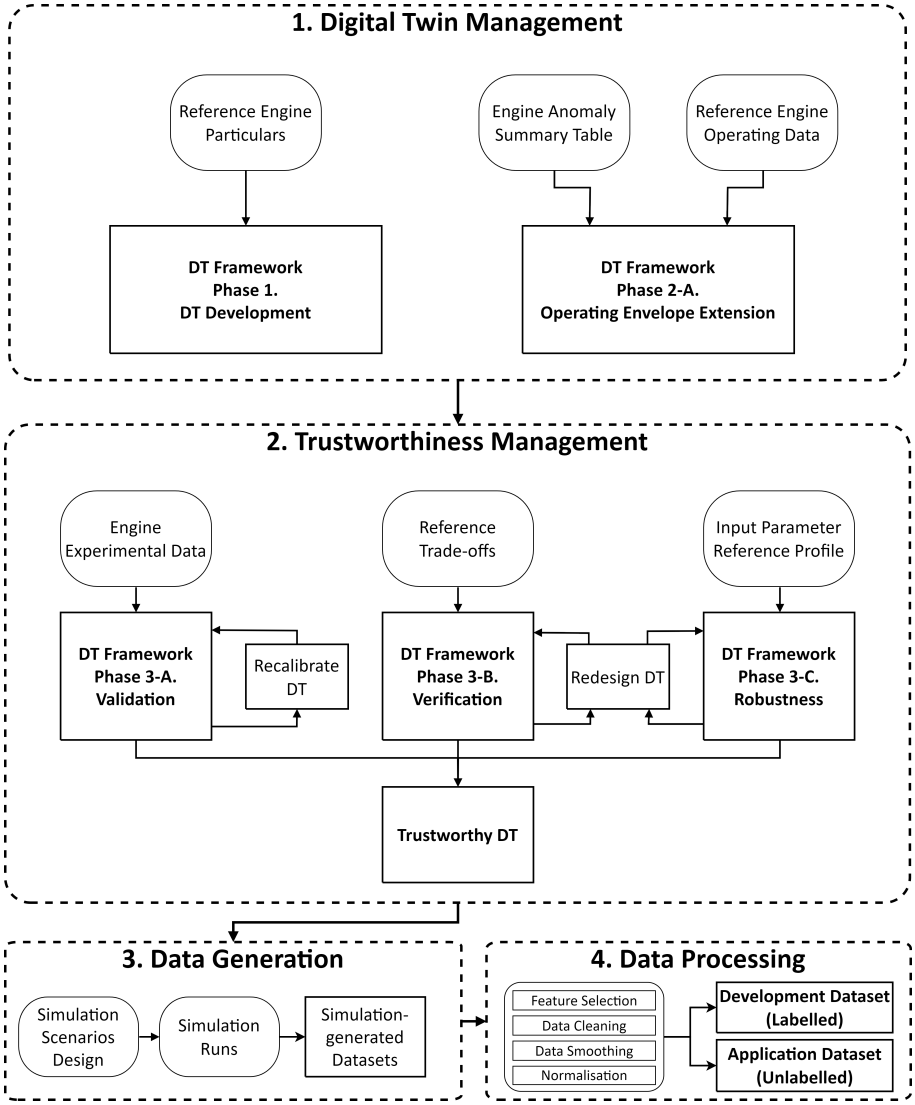


Figure 4.4: Systematic framework for generating needed datasets for anomaly diagnosis model

Stage 1 deals with digital twin management. The reference engine particulars

are employed, and the investigated engine DT is developed according to Phase 1 of the framework (physical model-based DT development). The engine particulars include geometrical information, turbocharger component performance maps, and measured engine performance parameters. To extend the engine operating envelope, which accommodates anomaly conditions and a wide range of ambient conditions, a reference engine anomaly summary table and reference engine operating data are employed. The anomaly summary table provides critical components for investigating anomalies and outlines their functions within the physical model. This information is provided as input to Phase 2-A of the framework (progressivity).

Stage 2 deals with DT trustworthiness management according to Phase 3 of the framework (DT trustworthiness assurance). Engine experimental data are employed for the validation step, reference trade-offs for the engine performance parameters are used for the verification step, whereas an input parameter reference profile is employed for the robustness step.

Stage 3 focuses on data generation. Simulation scenarios are designed corresponding to the application of the generated datasets, considering the engine operating envelope, anomalies, and environmental conditions. Simulation runs are performed to generate the engine performance parameters for these scenarios. The simulation results are integrated to generate datasets.

Stage 4 deals with the data processing of the generated simulation datasets. The data processing includes feature selection, data cleaning, data smoothing, and normalisation. The processed datasets are then organised into a development dataset that includes labels and an application dataset that isolates health labels.

4.3.1 Data Generation

To explore a wide engine operating envelope effectively, this study employs DOE [192]. The GT-SUITE DOE tool [168] with the Latin hypercube method is used to define the number of samples required to adequately represent the operating envelope [181].

The anomaly threshold is identified by examining engine cylinder valve leakages within the 0–1.0 mm range using DOE. The variation of performance parameters is noticeable between 0.2 mm and 0.3 mm. For the classification task, the generated

anomaly datasets include three anomaly steps: acceptable level (0.1 mm), weak leakage (0.3 mm), and severe leakage (0.5 mm). For the regression task, the generated time-series datasets include continuous anomaly propagation from health condition of 0 mm valve wear at 0 running hours to failure condition of 0.3 mm valve wear at 4,000 running hours.

4.3.2 Data Processing

Recent advancements in sensor and information technology have significantly facilitated the acquisition of ship machinery system measurements [193]. However, conditioning of the acquired datasets is required to remove bias and noise, as well as identify human transcriptional errors and missing values [6, 117]. Additionally, challenges pertaining to big data, multiple sources, synchronisation, and diverse data structures must be addressed. To effectively manage the extensive and diverse ship machinery data, an appropriate data management is required. However, the data management system is left outside the scope of this study, to focus on the critical functions of developing and managing data-driven models. This study only considers data storage to collect datasets from multiple sources and data processing to remove missing values, duplicates, outliers, irrelevant features, and inconsistent dependencies.

The correlation between engine performance parameters and HI of simulation datasets is evaluated for feature selection, employing the Spearman's coefficient computed via Equation (4.9). Features exhibiting absolute Spearman's coefficient values exceeding 0.9 are selected for the PHM model development.

The Savitzky-Golay (savgol) smoothing filter [194] is employed to reduce stochasticity in datasets while preserving the shape of the original signal, unlike moving average methods. The savgol filter, typically used in signal processing, employs a polynomial function to fit the signal within each time window, thereby removing noise and outliers, imputing missing values, and identifying underlying trends [87]. This study employed a polynomial order of 3 to accommodate the data non-linearities.

The data processing also employs the Min-Max normalisation [195] method to remove inconsistent dependencies between performance parameters (features) by fitting all data values within the same range (m to M) according to Equation (4.10).

Chapter 5

Data-driven Prognostics and Health Management Model

5.1 PHM Model Development

Fault diagnosis models aim to identify and understand the current health class of the reference system [196]. Data-driven fault diagnosis typically employs classification algorithms to determine health classes. This study employs supervised learning methods with simulation-generated labelled training datasets, as these methods surpass semi-supervised or unsupervised learning methods when extensive labelled data is available. The diagnosis model includes sub-models for anomaly detection, identification, and isolation. The detection models employ binary classifiers to determine the system condition as either healthy or abnormal. The isolation models employ multi-class classifiers to determine the anomaly location (cylinder). The identification models employ multi-class classifiers to determine the anomaly types (sub-systems or components). For maintenance decision making, the identification results are more important than the isolation results [197].

The health prognosis model employs the trend modelling method, which exhibits a low prediction error and a good generalisability [97]. The health prognosis model includes one HI construction sub-model and multiple HI forecast sub-models for each component. The HI construction sub-model estimates the HIs for each component by using performance parameters as input. The HI construction model employs multi-input multi-output regression algorithms to construct the HI values of each

component from multivariate data. Each HI forecast sub-model employs the HI pertaining to a specific component, derived from the output of the HI construction sub-model, to predict the HI in future time slices for the same component. The HI forecast model employs time series forecast algorithms to predict the HI for future time slices based on the HI values of the past time slices. In the HI forecast models, probabilistic methods are also employed to quantify the uncertainty of the forecast result. Subsequently, the RUL is estimated using the predicted HI and a predefined end-of-life (EoL) point [107]. Maintenance decisions are made based on the estimated RUL and HI.

To select the most effective methods for the sub-models, applicable methods determined from literature review are explored and compared in the following chapters.

5.1.1 Fault Diagnosis Model

The performance of applicable methods are compared using the Classification Learner in the MATLAB toolbox, and a fault diagnosis model with a selected method is developed in Python for seamless integration with the PHM model. This study considers the following six algorithms and their derivatives to determine the most effective algorithms for marine engine diagnostics: Logistic Regression, KNN, SVM, Naive Bayes Classifier, and MLP.The one-versus-one method [198] is employed for multi-classification of anomaly identification and isolation. The dataset with labels is divided into training and testing sets, ensuring the testing set contains an equal number of samples per class.

Logistic Regression is a powerful model for linear and binary classification tasks due to its easy implementation [199]. The logistic regression uses the logistic sigmoid function (Equation 5.1), where z is the net input. The input z is the linear combination of weights w_i and input features x_i, representing the logit function of the conditional probability that a particular example belongs to class 1 as shown in Equation (5.2). The logit function (Equation 5.3) is the natural log of odds $(\frac{p}{(1-p)})$ which means the odds in favour of a particular event, where p stands for the

probability of a class.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{5.1}$$

$$z = w^T x = \sum_{i=0}^{m} w_i x_i = logit(p(y = 1|x)) \tag{5.2}$$

$$logit(p) = log \frac{p}{(1 - p)} \tag{5.3}$$

KNN is a simple non-parametric machine learning model using a local interpolation approach. The approach approximates a new value with its K-nearest neighbours by majority voting for classification tasks and averaging for regression tasks [200]. This study selects the KNN because of its simple prediction process and low computational cost with low-dimensional data. During the hyperparameter tuning, the values of 1, 10, and 100 were tested for the number of neighbours. Consequently, the value of 10 was selected.

SVM employs decision hyperplanes, which have minimum overall error [201] and estimates classes of new data with the trained hyperplanes [202]. However, the kernel function, which transforms nonlinear data into linearly separable data, is additionally required since the SVM was originally designed for linear problems [96]. The Radial Basis Function (RBF) kernel [203] is employed herein due to its simplicity in hyperparameter tuning, including only one hyperparameter (σ), according to Equations (5.4) and (5.5).

$$\kappa\left(x_1, x_2\right) = \exp\left(-\frac{\left\|x_1 - x_2\right\|^2}{2\sigma^2}\right) \tag{5.4}$$

where κ denotes the kernel function of two-dimensional inputs (x_1, x_2) and σ is the width of the kernel.

$$F(x) = \sum_{i=1}^{n} w_i \kappa\left(x, x_{\mathrm{i}}\right) \tag{5.5}$$

where F denotes the multidimensional kernel function, n is the number of dimensions, and w is the weight of dimensions.

Naive Bayes Classifier is based on Bayes' Theorem and the conditional independence assumption (Equation 5.6) [204]. The naive Bayes model trains the joint probability distribution of input and output parameters based on the conditional probability distribution. The trained model estimates the output label c_j of a new input x using the biggest posterior probability (Equation 5.8) by Beyes' Theorem (Equation 5.9), where K is the number of classes.

$$P(X = x|Y = c_j) = P(X^{(1)} = x^{(1)}, ..., X^{(n)} = x^{(n)}|Y = c_j) \tag{5.6}$$

$$= \Pi_{l=1}^{n} P(X^{(l)} = x^{(l)}) j = 1, 2, ..., K \tag{5.7}$$

$$y = \underset{x}{\text{argmax}} P(Y = c_j)\Pi_l P(X^{(l)} = x^{(l)}|Y = c_j) \tag{5.8}$$

$$P(Y = c_j|X = x) = \frac{P(X = x|Y = c_j)P(Y = c_j)}{\Sigma_j P(X = x|Y = c_j)P(Y = c_J)} \tag{5.9}$$

MLP is one of the most frequently used deep learning models with multiple hidden layers [205]. MLP consists of multiple neurones and estimates the output by calculating a weighted sum of the neurone outputs [206]. This study selects the MLP due to its universal approximation capability [207]. During the hyperparameter tuning, the number of hidden layers was searched among the values 1, 2, and 3, as well as number of neurons was searched among the values 10, 25, and 100. As a result, the MLP with 2 hidden layers and 10 neurons for each layer was selected. Training MLP involves adjusting the weights of the connections between neurones to minimise classification or regression errors [208].

Evaluation Criteria

The performance of the anomaly diagnosis models is characterised by employing the confusion matrix, whereas the model accuracy is calculated according to Equation

(5.10). The confusion matrix [209] illustrates the distribution of true and false predictions over all classes in the tabular format shown in Table 5.1.

$$A = \frac{TN + TP}{TN + TP + FN + FP} 100 \tag{5.10}$$

where A denotes the accuracy percentage, TN is the number of true negative samples, TP is the number of true positive samples, FN is the number of false negative samples, and FP is the number of false positive samples.

Table 5.1: Example of a multi-classification confusion matrix for class 2

		Predicted Class		
		Class 1	Class 2	Class 3
True Class	Class 1	TN	FP	FN
	Class 2	FN	TP	FN
	Class 3	FN	FP	TN

5.1.2 Health Prognosis Model

The objectives of the health prognosis model are to estimate the HI and RUL of the degradation components. HI (for each component) represents the health status of the component in a specific time slice. HI ranges from 1 (healthy) to 0 (failure) and is calculated by Equation (5.11).

$$HI = 1 - \frac{H_c - H_h}{H_f - H_h} \tag{5.11}$$

Where H_c is health-related parameters at the current time slice, H_h is health-related parameters in healthy conditions, and H_f is health-related parameters in failure conditions.

The RUL of each engine component is calculated by Equation(5.12) considering the time horizon (in running hours) until the foretasted HI value reaches a user-defined threshold.

$$RUL = RH_t - RH_c \tag{5.12}$$

Where RH_t is the running hour at the threshold point, and RH_c is the running hour at the current time slice.

The parameters used to calculate HI and RUL are illustrated in Figure 5.1.

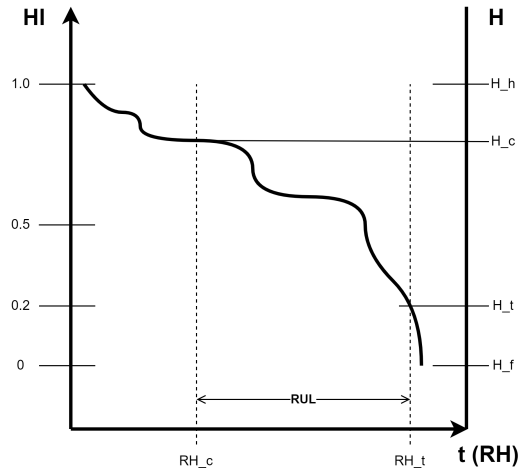


Figure 5.1: Parameters for calculating HI and RUL

HI Construction Sub-model

The HI construction model employs multi-input multi-output regression algorithms to construct the HI values of each component from multivariate data. Regression models investigate how variations in input parameters affect output parameters. This study considers the following six algorithms: Linear Regression (LR), KNN, RF, SVR, LSTM, and MLP.The algorithms are selected based on those commonly employed in the literature. The selected algorithms are employed as basic architectures, facilitating an effective comparative assessment. The algorithms, excluding those described in the above fault diagnosis section, are described in detail below.

LR fits the regression model to a linear line [210]. In this study, the LR is selected due to its simple structure and low computational cost. Training of LR is finding the optimal weights and bias values for the input features to minimise the sum of squared errors between actual HI and constructed HI. To avoid overfitting, the Ridge regularisation method is applied to the LR model [206].

RF is an ensemble learning method consisting of multiple decision trees. The RF estimates regression results by combining the probabilities from each decision path [211]. In this study, the RF is selected due to its favourable error and generalisation ability, attributed to its randomness [212]. Training of RF involves constructing multiple regression trees using a different bootstrap sample of the training data. The constructed regression trees estimate individual outputs, and the RF estimates a regression output by aggregating the results from each tree [213].

SVR is a machine learning model with a generalisation approach. The SVR disregards errors within specified margins [214]. The SVR is selected for its robustness to noisy data and strong generalisation ability. In this study, the RBF kernel [215] is used to extend the linear SVR for non-linear regression problems. Training of the SVR involves identifying an optimal hyperplane that minimises the regression error [109].

LSTM is a type of RNN, which includes a series of recurrent units [216]. The recurrent units include input, forget, and output gates to retain hidden information over time [106]. LSTM is selected for its ability to process long-term sequential data by maintaining important information and discarding short-term disturbances. Training of the LSTM involves adjusting the biases and weights of LSTM cells to minimise regression errors [217].

HI Forecast Sub-model

The HI forecast model employs time series forecast algorithms to predict the HI for future time slices based on the HI values of the past time slices. This study tested the following four deterministic algorithms: ARIMA, LSTM, CNN, and MLP.The basic architectures of these algorithms are employed. The algorithms, excluding those described in the above Fault Diagnosis and HI construction section, are described in detail below.

ARIMA is a statistical model based on the autoregressive concept, which hypothesises that the present value of a time series is dependent on past values [87]. ARIMA is selected as a representative of traditional forecast methods. The ARIMA combines

the autoregressive process and the moving average process to predict near-future data by using time series data [218].

CNN is a specialised deep learning model composed of multiple convolution layers and pooling layers [111]. In the convolution layer, convolutional kernels slide the input data space to extract low-dimensional feature maps. The pooling layer reduces feature map dimensions, and the fully connected layer summarises the features to predict HI forecast results [74]. The CNN is selected due to its parameter sharing ability, which reduces the number of trainable parameters and enhances generalisation ability [219]. Training of CNN involves adjusting the weights in the kernel to minimise the regression error [217].

Probabilistic Methods for HI Forecast

The probabilistic method estimates the probability distribution of the model output and quantifies the uncertainty of the HI forecast results. This study estimates the probability distribution by assuming that RULs follow a normal distribution. The wide shape of the distribution indicates that the method has high uncertainty; conversely, the narrow shape of the distribution indicates that the method has low uncertainty.

For probabilistic HI forecast, this study considers the ensemble method, the Bayesian variational inference method, and the Monte Carlo dropout method. The considered probabilistic methods are illustrated employing MLP in Figure 5.2.

Ensemble Method employs multiple parallel models and generates an output by combining multiple prediction results [220]. Randomly initialised models forecast HI trends using a shared training dataset [221]. The diverse results from individual models represent potential HI trajectories of future time slices and provide the probability of the forecast results [222]. The ensemble method demonstrates superior generalisation ability and achieves improved performance relative to a single model [223].

Bayesian Variational Inference employs weights as probability distributions instead of fixed values and generates output as probability distributions, particularly

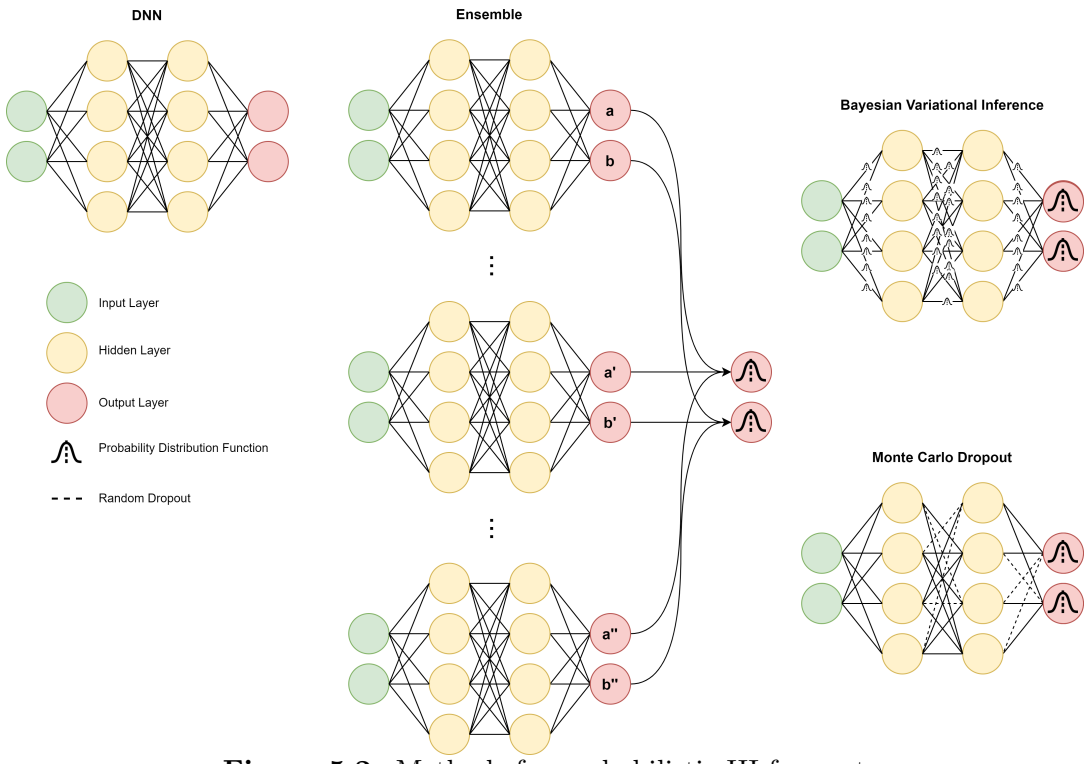


Figure 5.2: Methods for probabilistic HI forecast

in neural network architectures [115]. This probabilistic model is useful for quantifying the uncertainty of the HI forecast. The probability distribution of the results (output variables) is calculated by Equation 5.13.

$$p(HI|t,D) = \int p(HI|t,\theta)p(\theta|D)d\theta \tag{5.13}$$

Where, $p(HI|t,D)$ denotes the probability distribution of the output variables HI given a new input t and the training dataset D, $p(HI|t,\theta)$ is the likelihood function given the BNN model parameter θ that represents the BNN model weights and biases, and $p(\theta|D)$ is the posterior distribution given the training dataset D.

To approximate the posterior distribution of the output HI for future time slices, this study employs the variational inference method [224]. The method employs the variational distribution ($q_\phi(\theta)$) and trains the distribution parameter ϕ to make the distribution as close as possible to the actual posterior distribution $p(\theta, D)$. The Kullback-Leibler divergence (KL-divergence) method is employed to measure the closeness between the distributions, according to which the loss $Loss$ is calculated

by Equation (5.14).

$$Loss = D_{KL}(q_\phi \parallel P) = \int_\theta q_\phi(\theta') \log\left(\frac{q_\phi(\theta')}{P(\theta'|D)}\right) d\theta' \tag{5.14}$$

Where $D_{KL}(q_\phi \parallel P)$ denotes the differences between the variational distribution and the posterior distribution.

To avoid the need for the posterior distribution $p(\theta', D)$ in Equation (5.14), the Evidence Lower Bound (ELBO) method is used, which employs Equation(5.15).

$$\int_\theta q_\phi(\theta') \log\left(\frac{P(\theta', D)}{q_\phi(\theta')}\right) d\theta' = \log(P(D)) - D_{KL}(q_\phi \parallel P) \approx -Loss \tag{5.15}$$

As $\log(P(D))$ is solely a function of the prior distribution, the maximisation of ELBO is equal to the minimisation of KL-divergence.

To maximise ELBO, the stochastic variational inference (SVI) method [225] is employed, which applies the stochastic optimisation method to approximate the posterior distribution. The BNN training is based on the process reported in Jospin et al. [224], the main steps of which are listed in the pseudo-code shown in Algorithm 1.

Monte Carlo Dropout is a simplification technique to implement the Bayesian approach to deep learning algorithms [224]. The Monte Carlo dropout method applies stochastic dropout to the hidden layers of existing neural network models to randomise the results [226]. The randomly generated outputs approximate the sampled outputs from a posterior predictive distribution [227].

Hyperparameters Setting

For the prognosis model, hyperparameters of the employed methods are tuned using a grid search with a five-fold cross-validation approach [228], which tests all possible combinations of hyperparameters [229], to find the best combination. For HI construction, regressor-based methods like LR, KNN, RF, and SVR utilise the default settings from the Python 'sklearn' library, whereas deep learning methods like LSTM, and MLPs utilises the Python 'keras' library. Hyperparameters, search ranges, and selected options for deep learning methods are summarised in Table 5.2.

Algorithm 1 Development process of BNN

Input:
HI history with time t,
BNN hyperparameters (number of layers and nodes),
Type of distribution for BNN parameters θ (bias and weights)
Number of Epochs N_{epochs}, Samples $N_{samples}$, and Epochs for early stopping N_{es}
Output:
Distribution parameters ϕ (mean and standard deviation) for the BNN parameters θ

Define the BNN structure
Define a distribution $q_\phi(\theta)$) with distribution parameters ϕ
Initialise the distribution parameters ϕ for bias and weights θ
for epoch = 1 to N_{epochs} **do**
  Shuffle the HI history dataset
  for each mini-batch (t_{batch}, HI_{batch}) from the HI history dataset **do**
    for i = 1 to $N_{samples}$ **do**
      Sample $\theta_i \sim q_\phi(\theta)$ to predict $\hat{HI}_{batch}$ for θ_i
    end for
    Use Equation(5.15) to calculate Loss (L)
    Calculate gradients of the Loss (L) w.r.t. ϕ
    Use Adam optimiser with the Loss (L) gradients to update the distribution parameters ϕ
    if Loss (L) is not improved for N_{es} successive epochs **then**
      Terminate training loop
    end if
  end for
end for
return Distribution parameters ϕ

Table 5.2: Hyperparameter search for HI construction

Method	Hyperparameter	search range
LSTM	Number of Neurons	[16, 32, **64**]
	Batch Size	[16, **32**, 64]
	Epochs	[30, **50**]
MLP	Number of Neurons (1)	[16, 32, **64**]
	Number of Neurons (2)	[16, 32, **64**]
	Batch Size	[16, **32**, 64]
	Epochs	[30, **50**]

***Bold**: Best combination of the hyperparameters

Table 5.3: Hyperparameter search for HI forecast

Method	Hyperparameter	search range
ARIMA	Number of Lags (p)	[0, 1, 2, 3]
	Differencing Order (d)	[0, 1, 2, 3]
	Moving Average Window (q)	[0, 1, 2, 3]
LSTM	Number of Neurons	[16, 32, **64**]
	Batch Size	[**16**, 32, 64]
	Epochs	[30, **50**]
CNN	Number of Neurons	[16, 32, **64**]
	Batch Size	[**16**, 32, 64]
	Epochs	[30, **50**]
	Kernel Size	[1, **2**, 4]
MLP	Number of Neurons (1)	[16, 32, **64**]
	Number of Neurons (2)	[16, **32**, 64]
	Batch Size	[16, **32**, 64]
	Epochs	[30, **50**]

***Bold**: Best combination of the hyperparameters

For HI forecast, hyperparameters, search ranges, and selected options for ARIMA model and deep learning methods like LSTM, CNN, MLP, and BNN are summarised in Table 5.3. The ReLU (rectified linear unit) serves as the activation function in deep learning methods, which is a piecewise linear function [217]. Unspecified hyperparameters use default values in Python libraries.

In the hyperparameter search, the maximum number of neurons for deep learning methods is limited to 64 to avoid extended searching time. Although the selected hyperparameters may not be optimal, the employed methods achieved performance above 0.9 on the R^2 score within the limited search range, making the methods suitable for comparing performances.

Evalutaion Criteria for HI construction

To evaluate the effectiveness of the selected methods, this study measures computational cost along with model performance metrics. The run time of each method is measured from the start of training to the completion of output generation. The performance metrics employed for the evaluation of each sub-model are described in detail below.

The selection of appropriate metrics to evaluate the performance of the HI construction sub-model is varied based on the type of data. When actual HI data is unavailable, the performance of the HI construction sub-model can be evaluated through monotonicity, trendability, and prognosability [230]. If RUL information is available, the correlation with RUL is useful to evaluate the performance of the HI construction sub-model [104]. If actual HI data is fully accessible, the error metrics, including but not limited to the coefficient of determination, serve as effective measures for assessing the performance of the HI construction sub-model.

Monotonicity assesses that the constructed HI has a consistent trade-off to capture the global degradation trend [230]. Since the degradation process is irreversible, a perfect score of -1 indicates a continuous decrease throughout the period. In this study, the Signum formula is used to measure the monotonicity (Mon) as follows:

$$Mon = \frac{1}{M}\sum_{j=1}^{M}\left|\sum_{k=1}^{N_j-1}\frac{\operatorname{sgn}\left(x_j(k+1)-x_j(k)\right)}{N_j-1}\right| \tag{5.16}$$

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{5.17}$$

where x_j denotes the measurement vector on the j^{th} sub-system, M is the number of sub-systems, and N_j is the number of measurements on the j^{th} sub-system.

Trendability assesses whether the constructed HI trend follows a consistent pattern throughout the operating period [230]. Pearson's correlation function is employed to measure the correlation between the constructed HI and the operating time. The score is in the range from zero (negligible correlation) to one (strong correlation).

The formula for measuring the trendability (Trend) is presented as follows:

$$Trend = \min \left| \rho \left(t, x_j \right) \right|, \quad j = 1, 2, \ldots, M \tag{5.18}$$

$$\rho \left(t, x_j \right) = \frac{\operatorname{cov} \left(t, x_j \right)}{\sigma_t \sigma_{x_j}} \tag{5.19}$$

where ρ denotes the Pearson's correlation function, t is the operation time, M is the number of sub-systems, cov is the covariance, σ_t and σ_{x_j} are the standard deviations of t and x_j, respectively.

Prognosability assesses the variability of the constructed HI during degradation, comparing it to the range between its initial and end values [74]. The score range starts from zero, with a perfect score reaching one. The formula for measuring the prognosability (Prog) is presented as follows:

$$prog = exp\left(- \frac{std(x_{HI})}{mean|x_{HI}(ini) - x_{HI}(end)|} \right) \tag{5.20}$$

where x_{HI} denotes the vector of constructed HI, $x_{HI}(ini)$ is the initial value of the HI vector, and $x_{HI}(end)$ is the end value of the HI vector.

Correlation with RUL evaluates the association between the constructed HI variation and the actual trend of RUL over time [104]. In this study, Spearman's Correlation, calculated by Equation 4.9, is used. The score is in the range from -1 (strong reciprocal correlation) to +1 (strong direct correlation) with 0 (negligible correlation).

Prediction Error Metric evaluates the correctness of the constructed HI values by comparing them to the actual HI values. This study employs the coefficient of determination (R-squared, R^2) due to its high interpretability. A perfect score is represented as one, while a negative value means that the direction of the regression line does not align with the actual trend. The formula for calculating the coefficient of determination is presented as follows [231]:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{m} (\bar{Y}_i - Y_i)^2} \tag{5.21}$$

Where m is the number of samples, Y_i is the i^{th} actual value, $\hat{Y}$ is the i^{th} predicted value, and $\bar{Y}$ is the mean value of actual values.

Evalutaion Criteria for HI Forecast

The performance of the HI forecast is evaluated using various metrics corresponding to the function of the model [107]. When the forecast model focuses on the RUL estimation, the performance of the forecast model can be measured by the RUL estimation error and the RUL margin of error. When the forecast model focuses on the HI variation, the performance of the forecast model can be measured by the overall prediction error and average variability of the regression results. In this study, the same error metric is employed for both HI construction and HI forecast, resulting in its singular description in the HI construction section.

RUL error measures the difference between the actual RUL and the estimated RUL. This study measures the difference directly as follows:

$$RUL_{err} = RH_{EoL} - \hat{RH}_{EoL} \tag{5.22}$$

Where RUL_{err} denotes the RUL error, RH_{EoL} is the run hour at the end of life, and $\hat{RH}_{EoL}$ is the forecasted run hour at the end of life.

RUL Margin of Error(MoE) is employed to quantify the uncertainty of the RUL estimation. Small MoE implies that the RUL estimation results have less uncertainty, whereas large MoE implies that the RUL estimation results have high uncertainty. The RUL MoE is associated with the uncertainty of the HI forecast models at the end of life point. In this study, a Z-score of 1.96 is employed with a 95% confidence level, as presented in the following equation [232]:

$$MoE = 1.96\frac{\sigma}{\sqrt{n}} \tag{5.23}$$

Where MoE denotes the margin of error, whilst σ and n denote the standard deviation and number of the output variables, respectively.

Variability is employed to quantify the uncertainty of the HI forecast models throughout the prediction period. This study calculates the variability by averaging the standard deviations. The perfect score of the overall average variability is 0, and the metric is formulated as follows [107]:

$$OAV = \frac{1}{M}\Sigma_{m=1}^{M} std(m) \tag{5.24}$$

Where OAV denotes the overall average variability, M is the number of models, and std is the standard deviation.

5.1.3 Maintenance Decision Making

The maintenance decision-making system uses the estimated health class from the diagnosis model and the calculated RUL from the prognosis model. The calculated RUL includes the deterministic result of the RH at failure and the probabilistic result of MoE. Decision-making utilises the lower bound of a 95% Confidence Interval (95% CI) that integrates the deterministic and probabilistic results. The confidence interval of 95% indicates that the system failure is expected between the lower and upper bounds with the probability of 95%. The lower bound is utilised instead of the mean or the upper bound to provide a safety margin for the prognostic uncertainty.

The maintenance decision making is based on user-defined criteria (rule-based). In this study, two statements are used, considering the estimated HIs and RULs for each component.

1. The engine needs maintenance if the RUL for at least one component is expected to be less than the Estimated Time of Arrival (ETA) to the next port, taking into account the time required for arranging and preparing maintenance activities. This criterion implies that the component is predicted to experience failure during the upcoming voyage, before the ship reaches the next port.

2. The engine needs maintenance if the HIs of all components reach the pre-defined warning threshold. The warning threshold corresponds to the manufacturer's maintenance recommendation. This criterion implies that the health conditions of all components are degraded to a level below the recommended health condition by the manufacturer for maintenance.

The maintenance decision criteria used are described as pseudocode in Algorithm 2.

Algorithm 2 Pseudocode for maintenance decision-making

Input: HI, RUL, and ETA
Output: Maintenance Decision

```
if RUL for one (or more) engine components is less than the ETA to the next port then
    Maintenance is required
else if HIs for all components are less than the warning threshold then
    Maintenance is required
else
    Continue operation without maintenance
end if
```

5.2 Framework for Managing PHM Model Management

A pre-trained (one-time developed) PHM model cannot address the diverse and dynamic degradation patterns during ship operations. Model management is needed to maintain the effectiveness of the developed PHM model. A range of real-world degradation scenarios, both diverse and dynamic, can be managed through iterative monitoring and updates.

5.2.1 Overall Management

The development of the PHM model employs three main digital tools: the DT, the HI construction sub-model, and the HI forecast sub-model. These tools are managed throughout the physical system's lifetime as illustrated in Figure 5.3.

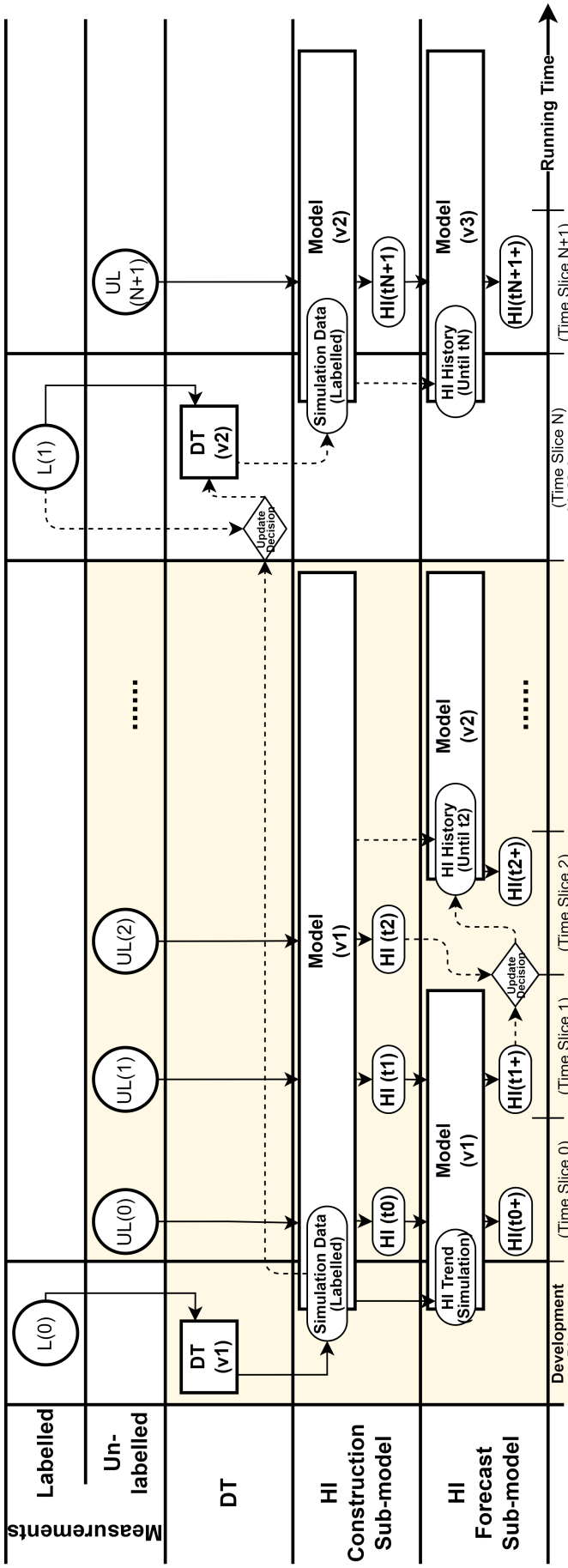


Figure 5.3: Use and management of the digital twin, as well as the HI construction and HI forecast sub-models throughout the lifetime of the physical system

In practical scenarios, the health conditions of the engine components cannot be evaluated while in use and can only be measured during engine overhaul. This study develops a data-driven PHM model management framework considering the availability of actual measurements. The framework employs only performance parameters that are accessible during engine operation.

Datasets $L(0)$ and $L(1)$ denote labelled datasets based on measurements of health conditions and performance parameters. $L(0)$ represents the datasets based on measurements during the engine trials (shop and sea) performed in healthy conditions. $L(1)$ represents the datasets based on measurements during maintenance, when the engine components exhibit degradation or faults. Labelled dataset $L(0)$ is employed to calibrate the DT, which subsequently is used to generate labelled simulation datasets for the HI construction sub-model training.

When $L(1)$ is available (after maintenance activities), the trustworthiness of DT can be evaluated and DT re-calibration can be performed if required. In the latter case, the re-calibrated DT is employed to generate new labelled simulation data, which can be used to re-train the HI construction sub-model. The re-trained HI construction sub-model is then employed to provide the HI at past time slices (using the respective performance parameters as input), which, in turn, can be used to re-train the HI forecast sub-models.

Dataset $UL(0)$, $UL(1)$, $UL(2)$, and $UL(N+1)$ denote unlabelled datasets based on measurements of performance parameters during the engine lifetime in consecutive time slices. The unlabelled data cannot be used to evaluate the DT and HI construction sub-model performance. However, datasets UL can be provided (as input) to the HI construction sub-model to estimate the HI at each time slice (e.g., $HI(t0)$, $HI(t1)$, etc.), which along with the HI at the past time slices are provided as input to the HI forecast sub-models to predict HI at future time slices (e.g., $HI(t0+)$, $HI(t1+)$, etc.). The HI and RUL predictions are then employed to assess the HI forecast sub-model re-training criteria that include accuracy and robustness metrics.

The error between the forecasted HI (e.g., $HI(1t+)$) and the constructed HI (e.g., $HI(t2)$) is calculated and used as the error metric. The uncertainty ratio of the current RUL distribution ($RUL(1)$) and the initial RUL distribution ($RUL(0)$) is used as the robustness metric. If the forecast error and the uncertainty ratio exceed

their pre-set retraining thresholds, the HI forecast sub-model is re-trained as follows.

Unlabelled datasets up to the current time slice (e.g., $UL(0) + UL(1) + UL(2)$) are provided as input to the HI construction model, which generated HI datasets of the past time slices(e.g., $HI(t0) + HI(t1) + HI(t2)$). The latter is used to re-train the HI forecast model. The maximum quantity of re-training data is limited to the maintenance interval of the considered component to avoid excessive computation for the model re-training.

This research outlines the overall management strategy, with emphasis on the HI forecast sub-model management. Management of the HI forecast sub-model is critical for retaining the PHM system performance, particularly because labelled datasets $L(t)$ are not available during the system operations. This study defines the research boundary as the management of the HI forecast sub-model. The boundary is highlighted in Figure 5.3 and the details of the management of the HI forecast sub-model are illustrated in Figure 5.4. The outside of the research boundary is expected in future studies.

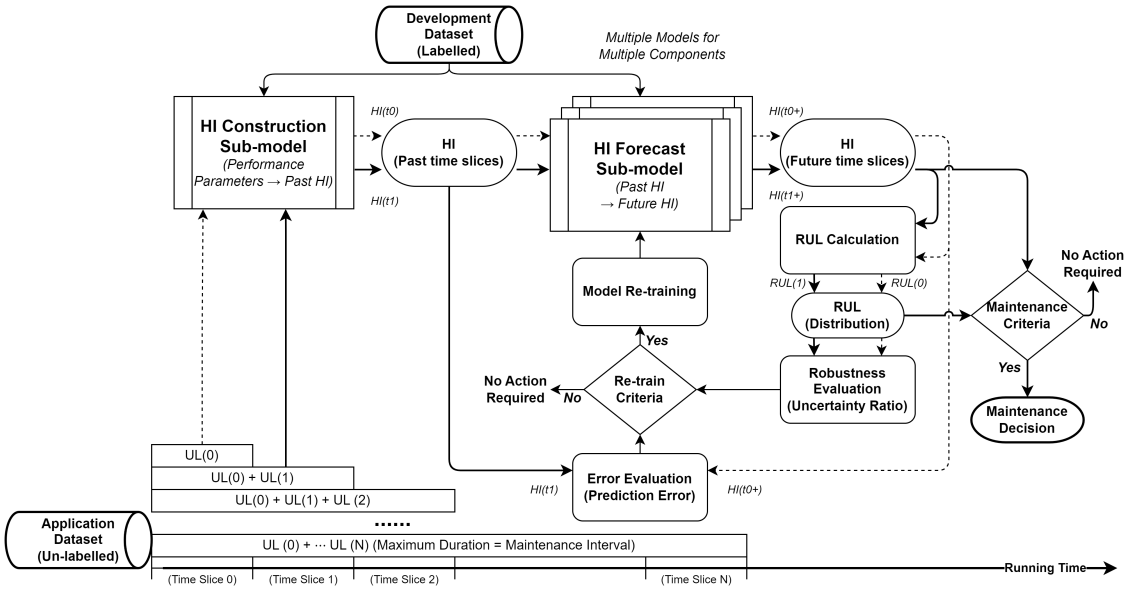


Figure 5.4: Flow diagram of HI construction data-driven sub-model HI forecast sub-model management

5.2.2 Re-training Criteria

The data-driven model management re-trains the HI forecast sub-models based on evaluations of error and robustness metrics. Two criteria are chosen based on the trustworthiness of the data-driven PHM model and its applicability to autonomous

ships. Autonomous shipping requires a precise and highly confident RUL estimation to effectively schedule maintenance activities during the limited duration of port stays.

The error of the data-driven model is evaluated using the coefficient of determination (R^2) according to Equation (5.21). The re-training based on the error criteria ensures that the performance of the data-driven model remains above the specified threshold. The minimum performance requirements vary depending on the purpose of the model; however, an R^2 value of at least 0.5 indicates statistical significance [233].

The uncertainty of the prediction is quantified as the margin of error with 95% confidence calculated according to Equation (5.23) is employed to represent the uncertainty metric of the output variables. The robustness metric for the data-driven model is evaluated using a predefined uncertainty threshold or the uncertainty ratio calculated according to Equation (5.25).

$$R_u = \frac{MoE_c}{MoE_f} \tag{5.25}$$

where R_u denotes the uncertainty ratio, MoE_c is the margin of Error for the RUL distribution at the current time slice, and MoE_f is the margin of error for the RUL distributions at the first time slice.

The uncertainty associated with the estimation of the RUL is expected to decrease over time, due to the shortening of the forecast time horizon as the system approaches failure [234]. The uncertainty of a robust PHM model is expected to have the highest uncertainty in the first time slices, which employs small training datasets but has a longer forecast time horizon. This study uses the uncertainty ratio value of 1 as the robustness threshold for re-training the model.

Chapter 6

Reference System and Case Studies

6.1 Reference System

The reference physical system in this study is a marine engine along with the sub-systems and components for its monitoring, control, and safety functions. The engine is part of a generator set used in electric propulsion systems for generating electric power. The reference system is the Wärtsilä 8L50DF engine, which is a four-stroke, turbocharged, and intercooled dual-fuel engine [175], offering flexibility, as it can operate in diesel or gas modes [235]. The main particulars of this engine are listed in Table 6.1, whereas the engine's sub-systems and components which may face anomalies during the engine's lifetime are listed in Table 6.2.

Table 6.1: Reference engine system technical specification

Engine Type	8L50DF
Maximum Continuous Rating Power	7800 kW
Nominal Engine Speed	514 rev/m
Cylinder Bore	500 mm
Stroke	580 mm
Number of Cylinders	8
Turbocharger	1 TPL 76

Table 6.2: Reference dual-fuel generator engine along with sub-systems and components

Combustion System	Air Supply System	Fuel Supply System	Lubricating Oil System	Others
Fuel Injector	Turbocharger	Fuel Pump	LO Pump	Cooling Water System
Gas Admission Valve	Waste Gate Valve	Fuel Filter	LO Cooler	Safety & Monitoring System
Intake Valve	Air Cooler	Fuel Rack & Governor	LO Filter	Compressed Air System
Exhaust Valve	Air Filter			(Control & Starting)

6.1.1 Engine Anomalies

This study employs seat ring wear degradation in exhaust valves and intake valves. The valve wear degradation is an unavoidable degradation during engine operations and require frequent maintenance. The summary table of the anomalies is provided in Table 6.3.

The wear of the engine cylinder valve seat causes a reduction in the clearance of the valve stem, as shown in Figure 6.1 [236]. If this reduction exceeds the manufacturer's allowance, the valve does not fully close and causes air/gas leakages, leading to engine efficiency deterioration. To simulate these anomalies, the valve lift profile and the valve lash were determined based on the manufacturer's allowances for the valve stem clearance (0.147–0.199 mm). Hence, the following four severity levels were considered: No Leakage, Acceptable Clearance, Weak Leakage, and Severe Leakage. The engine performance parameters with the greater impact from these anomalies as well as their trade-offs were identified based on the pertinent literature [237, 238].

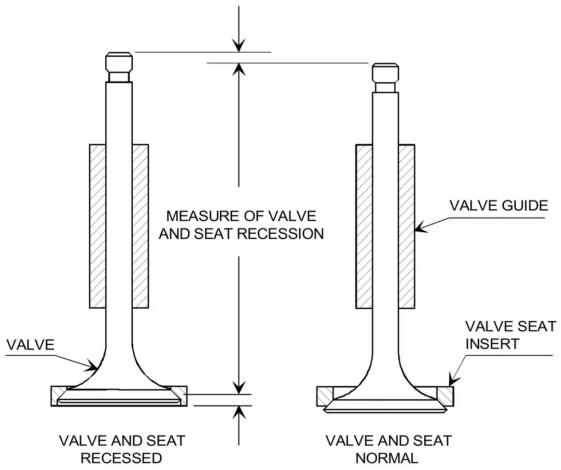


Figure 6.1: Valve Recession

Table 6.3: Anomaly summary table – Intake valve and exhaust valve

Component	Function	Failure Mode	Failure Causes	Failure Effects	Detection Method	O	S	D	RPN	Simulation Input				Simulation Output
										Name	Manufacturer Limit	Input Range	Simulation Steps	Symbol
Intake Valve	Supply intake air into a cylinder	The valve is not fully closed	Valve seat wear	Intake air leakage → Engine efficiency deterioration	Manual measurement of clearance	4	5	6	120	Valve remained lift and lash [mm]	Normal clearance: 1.0 mm (cold) Wear limit: 0.147–0.199 mm	Valve remained lift: 0–0.75 mm Valve lash: 0.25–1.0 mm	(4 steps of clearance) 0 mm—No leakage 0.1 mm—Acceptable clearance/No leakage 0.3 mm—Slight leakage 0.5 mm—Severe leakage	N_{TC} T_{Exh} p_{max} p_{CA} T_{CA} FOC
Exhaust Valve	Releases burned gases from a cylinder	The valve is not fully closed	Valve seat wear	Exhaust gas leakage → Engine efficiency deterioration	Manual measurement of clearance	4	5	6	120	Valve remained lift and lash [mm]	Normal clearance: 1.5 mm (cold) Wear limit: 0.147–0.199 mm	Valve remained lift: 0–1.0 mm Valve lash: 0.5–1.5 mm	(4 steps of clearance) 0 mm—No leakage 0.1 mm—Acceptable clearance/No leakage 0.3 mm—Slight leakage 0.5 mm—Severe leakage	N_{TC} T_{Exh} p_{max} p_{CA} T_{CA} FOC

6.2 Case Studies

The application of digital tools and frameworks within the research methodology is demonstrated in Chapter 7, utilising the aforementioned reference engine system and its anomalies. The case studies contain three demonstrations as shown in Table 6.4: Case 1 demonstrates trustworthiness assurance framework, Case 2 determines appropriate methods for each sub-task in the PHM model using sub-system (EV only) degradation datasets, and Case 3 demonstrates the integrated PHM model using multi-system (EV and IV) simultaneous degradation datasets. In Case 2, three sub-tasks for the fault diagnosis model, the health prognosis model, and the PHM model management framework are independently examined in Case 2–a, Case 2–b, and Case 2–c.

The employed datasets are listed in Table 6.5. Datasets for Case 1 and 2-a are structured data for classification tasks, containing performance parameters and the corresponding health classes. Datasets for Case 2-b, 2-c, and 3 are time series data, which include performance parameters, health indicators, and system running time for regression tasks.

Table 6.4: Case study summary table

Case	Material	Scope	Details	Methodology Phase
1	DT trustworthiness assurance framework	Demonstrating the developed framework for assuring trustworthiness of DT	Three steps: Validation, Verification, and Robustness	Phase 1 2 3
2–a	Diagnosis algorithms comparison	Investigating applicable algorithms for diagnosis model under various operating conditions	Logistic regression; KNN; Decision Tree; SVM; Naïve Bayes Classifier; MLP	Phase 4
2–b	Prognosis algorithms comparison	Investigating applicable algorithms for prognosis model under various operating conditions	LR; KNN; RF; SVR; LSTM; MLP; ARIMA; CNN	Phase 4
2–c	PHM model management framework	Demonstrating developed framework for managing the trustworthiness of PHM model	Two criteria: Accuracy and Robustness	Phase 5
3	Integrated PHM model	Demonstrating integrated PHM within complex scenario (Simultaneous degradation in multiple component)	Multi-components and simultaneous degradation	Phase 6

Table 6.5: List of employed dataset in case studies

Case Study	Scope	Health Condition	Load	Anomaly Type	Anomaly Severity
1	Trustworthiness(Validation)	Healthy	25; 50; 75; 100	–	–
1	Trustworthiness(Verification)	Healthy; $Leak_{EV}$; $Leak_{IV}$	25; 50; 75; 100	Single	0.1; 0.3; 0.5
1	Trustworthiness(Robustness)	Healthy; $Leak_{EV}$; $Leak_{IV}$	25; 50; 75; 100	Single	500 Severities
2–a	Fault Detection	Healthy; $Leak_{EV}$; $Leak_{IV}$	25; 50; 75; 100	Single	Binary
2–a	Fault Identification	$Leak_{EV}$ & $Leak_{IV}$	25; 50; 75; 100	Single	42 Severities
2–a	Fault Isolation	$Leak_{EV}$	25; 50; 75; 100	Multi-Location	Binary
2–a	Fault Isolation	$Leak_{IV}$	25; 50; 75; 100	Multi-Location	Binary
2–b	Health Prognosis – HI Construction Model Training	$Leak_{EV}$	70–85	DOE (Latin Hypercube 6561)	Degradation
2–b	Health Prognosis – HI Forecast Model Training	$Leak_{EV}$	70–85	Weak Stochasticity and Same Condition	Degradation
2–b, c	Health Prognosis – Model Testing	$Leak_{EV}$	70–85	Weak Stochasticity and Same Condition	Degradation
2–b, c	Health Prognosis – Model Testing	$Leak_{EV}$	70–85	Strong Stochasticity and Same Condition	Degradation
2–b, c	Health Prognosis – Model Testing	$Leak_{EV}$	70–85	Weak Stochasticity and Different Condition	Degradation
2–b, c	Health Prognosis – Model Testing	$Leak_{EV}$	70–85	Strong Stochasticity and Different Condition	Degradation
3	Integrated PHM model - Model Training	$Leak_{EV}$ & $Leak_{IV}$	70–85	DOE (Latin Hypercube 6720)	Degradation
3	Integrated PHM model - Model Training	$Leak_{EV}$ & $Leak_{IV}$	70–85	Monotonic + Same Condition	Degradation
3	Integrated PHM model - Model Testing	$Leak_{EV}$ & $Leak_{IV}$	70–85	Strong Stochasticity and Different Condition	Degradation

6.2.1 Assumptions

The following assumptions are employed:

- The considered marine engine exhibits a single anomaly type (valve recession), but multiple degradation components (exhaust and intake valves of each engine cylinder). The effects from the other engine components anomalies are taken into account via the considered stochastic degradation pattern. The stochasticity renders the degradation trajectories irregular; however, the degradation trajectories do not neglect the reference degradation pattern in a long time interval.

- Unmanned machinery operations were considered, which do not include human interventions during voyages, such as actual health status measurement or maintenance activities. Consequently, the actual HI from the simulation datasets is isolated from the data-driven model development and management processes. Due to the unavailability of the actual HI, accuracy evaluation and re-training for the HI construction sub-model are not performed. Moreover, the HI forecast sub-models employ the estimated HI history produced by the HI construction sub-model, rather than the actual HI.

- Voyage duration of the considered marine engine was established as one month, whereas the interval for data-driven model management was set at one week, as shown in Figure 6.2. During the operations, the engine changed its load randomly within the considered range; however, it was assumed that the engine would run continuously without starting or stopping.

- Lastly, the employed HI values and thresholds are assumed to demonstrate the key functions of the developed methodology. These values can be adjusted in future studies corresponding to the actual conditions of the considered physical system.

Figure 6.2: Timeline for the assumed voyages and management intervals

Chapter 7

Results and Discussion

7.1 Case Study 1 - Trustworthy Digital Twin for Data Generation

The DT of reference engine system was developed by the integration of thermodynamic model, degradation model, and sensor model, followed by a subsequent calibration process. The developed DT facilitates simulations under both healthy and abnormal conditions to generate corresponding simulation datasets. The trustworthiness of the DT was evaluated through the proposed trustworthiness assurance framework.

7.1.1 Trustworthiness Assurance

The DT trustworthiness assurance was based on the outcomes of the validation, verification, and robustness checks, which are summarised in the trustworthiness decision table.

In the validation step, simulated performance parameters were compared with the respective reference engine shop trial data, as shown in Figure 7.1. The estimated percentage errors for six performance parameters are presented in Table 7.1. The maximum absolute percentage error was identified as 3% for the fuel oil consumption at 100% load. However, all predicted parameters (at all loads) satisfied the acceptance criteria (allowable errors), which appear in the right column of Table 7.1. Hence, it is deduced that the DT passed the validation check.

In the verification step, the trade-offs of the DT-generated performance parameters

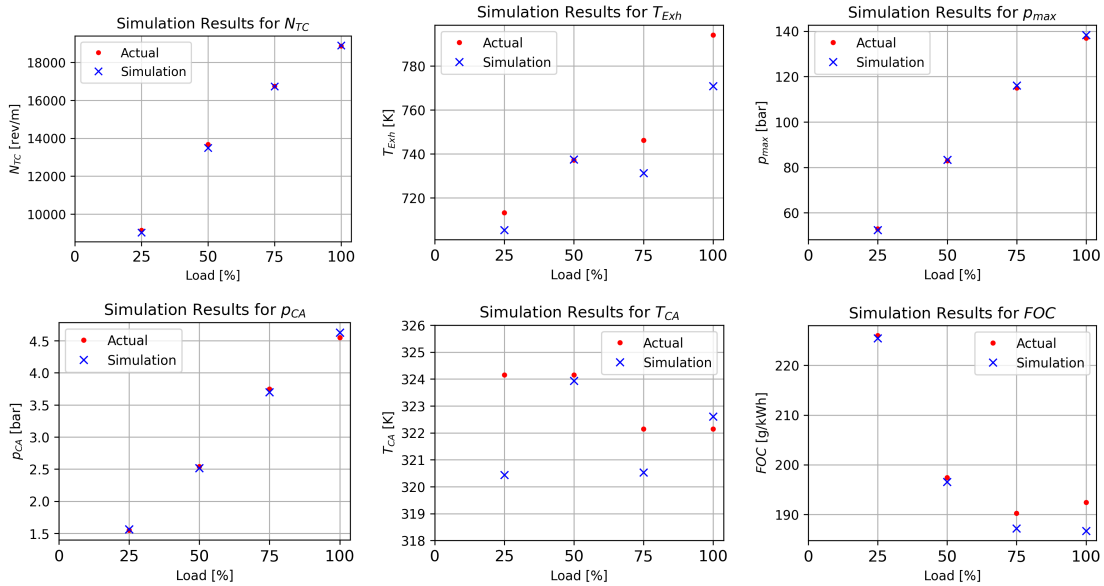


Figure 7.1: Case 1 – Comparison between simulation results and actual measurements

were compared with the trade-offs reported in pertinent experimental studies. Table 7.2 summarised the sensitivity analysis results, which were used to determine the trade-offs in the performance parameters caused by variations in the input parameters. These trade-offs were then compared against those reported in relevant experimental studies, as presented in Table 7.3 for the ambient temperature variations and Table 7.4 for the anomalies, respectively. The ambient temperature trade-offs were compared with the trade-offs reported by Whitehouse et al. [239], Serrano et al. [240], MAN [241], whereas the anomaly conditions trade-offs were compared with those reported by Kowalski [237]. From these tables, it is deduced that the predicted trade-offs correspond to the reported ones. Therefore, the developed DT passed the verification test.

In the robustness step, 524 operating points were employed to perform an uncertainty analysis of the predicted engine performance parameters. The uncertainty ratios are presented in Table 7.5. The DT in the various ambient conditions and anomaly conditions satisfied the robustness acceptance criteria as the uncertainty ratios of all the predicted performance parameters are below 1. Hence, it is deduced that the developed DT satisfies the robustness acceptance criteria. The results of the three steps are summarised in Table 7.6, from which it is confirmed that the developed DT is considered trustworthy.

Table 7.1: Validation results – percentage errors

Load [%]	**100**	**75**	**50**	**25**	**Acceptable Error**
N_{TC}	0.2	0.0	−1.3	−1.2	± 2
T_{Exh}	−2.9	−2.0	0.0	−1.1	±5 (±25 K)
p_{max}	1.0	0.9	0.4	−1.1	±5
p_{CA}	1.7	−1.4	−1.0	0.9	±2
T_{CA}	0.1	−0.5	−0.1	−1.1	±1.2 (±4 K)
FOC	−3.0	−1.6	−0.5	−0.3	±3

Table 7.2: Sensitivity analysis results

Input Parameters	**Spearman's Coefficient [–]**					
	N_{TC}	T_{Exh}	p_{max}	p_{CA}	T_{CA}	FOC
T_{Amb}	0.19	0.84	−0.23	−0.58	0.58	0.35
$Leak_{EV}$	0.92	0.15	−0.31	0.61	0.76	0.43
$Leak_{IV}$	0.36	0.56	−0.95	0.40	0.37	0.85

Table 7.3: Verification results – ambient temperature variation trade-offs

Output Parameters	**Low Temperature**		**High Temperature**		**Reference**
	DT	**Reference**	**DT**	**Reference**	
N_{TC}	↓	↓	↑	↑	[240, 241]
T_{Exh}	↓	↓	↑	↑	[241]
p_{max}	↑	↑	↓	↓	[241]
p_{CA}	↑	↑	↓	↓	[241]
T_{CA}	↓	↓	↑	↑	[239]
FOC	↓	↓	↑	↑	[241]

Table 7.4: Verification results – influence of anomaly trade-offs

Output Parameters	$Leak_{EV}$		$Leak_{IV}$		Reference
	DT	Reference	DT	Reference	
N_{TC}	↑	↑	↑	↑	[237]
T_{Exh}	↑	↑	↑	↑	
p_{max}	↓	↓	↓	↓	
p_{CA}	↑	↑	↑	↑	
T_{CA}	↑	↑	↑	↑	
FOC	↑	↑	↑	↑	

Table 7.5: Robustness results – uncertainty ratio

Input Parameters	Load [%]	Uncertainty Ratio [–]					
		N_{TC}	T_{Exh}	p_{max}	p_{CA}	T_{CA}	FOC
T_{Amb}	100	0.001	0.162	0.005	0.006	0.622	0.007
	75	0.001	0.151	0.004	0.006	0.605	0.008
	50	0.000	0.211	0.003	0.005	0.427	0.010
	25	0.000	0.125	0.000	0.001	0.729	0.009
$Leak_{EV}$	100	0.004	0.055	0.054	0.009	0.573	0.028
	75	0.005	0.025	0.030	0.010	0.750	0.024
	50	0.004	0.034	0.014	0.004	0.374	0.032
	25	0.002	0.060	0.005	0.000	0.082	0.045
$Leak_{IV}$	100	0.001	0.207	0.009	0.127	0.071	0.028
	75	0.001	0.156	0.004	0.139	0.068	0.024
	50	0.000	0.084	0.001	0.179	0.105	0.032
	25	0.000	0.036	0.000	0.073	0.129	0.045

Table 7.6: Trustworthiness decision results

Steps	**Acceptance Criteria**	**Trustworthiness Checks**	
		Environment Conditions	**Anomaly Conditions**
Validation	Acceptable Errors	Pass	
Verification	Trade-off Soundness	Pass	Pass
Robustness	Uncertainty ratio	Pass	Pass

7.1.2 Data Generation and Application Test

Following the assurance of the developed DT trustworthiness, 79,980 data samples were generated within the considered operating conditions. Data samples depicted distinct combinations of operating conditions, integrating input parameters such as exhaust valve leaks, intake valve leaks, engine loads, and ambient temperatures. The generated dataset contained several engine performance parameters (N_{TC}, T_{Exh}, T_{CA}, p_{CA}, FOC, p_{max} for each cylinder). The generated datasets were randomly split for training and testing data-driven models.

The anomaly diagnosis model was developed using SVM. Figure 7.2 illustrates confusion matrices with classification accuracies and errors related to detection, identification, and isolation sub-tasks. The anomaly detection model exhibits an accuracy of 98.8% with 2,000 test datasets. The anomaly identification model exhibits an accuracy of 97.6% with 3,000 test datasets. The anomaly isolation model exhibits an accuracy of 90.1% for determining the location (cylinder number) of the exhaust valve leakage and an accuracy of 91.8% for determining the location (cylinder number) of the intake valve leakage with 2,400 test datasets.

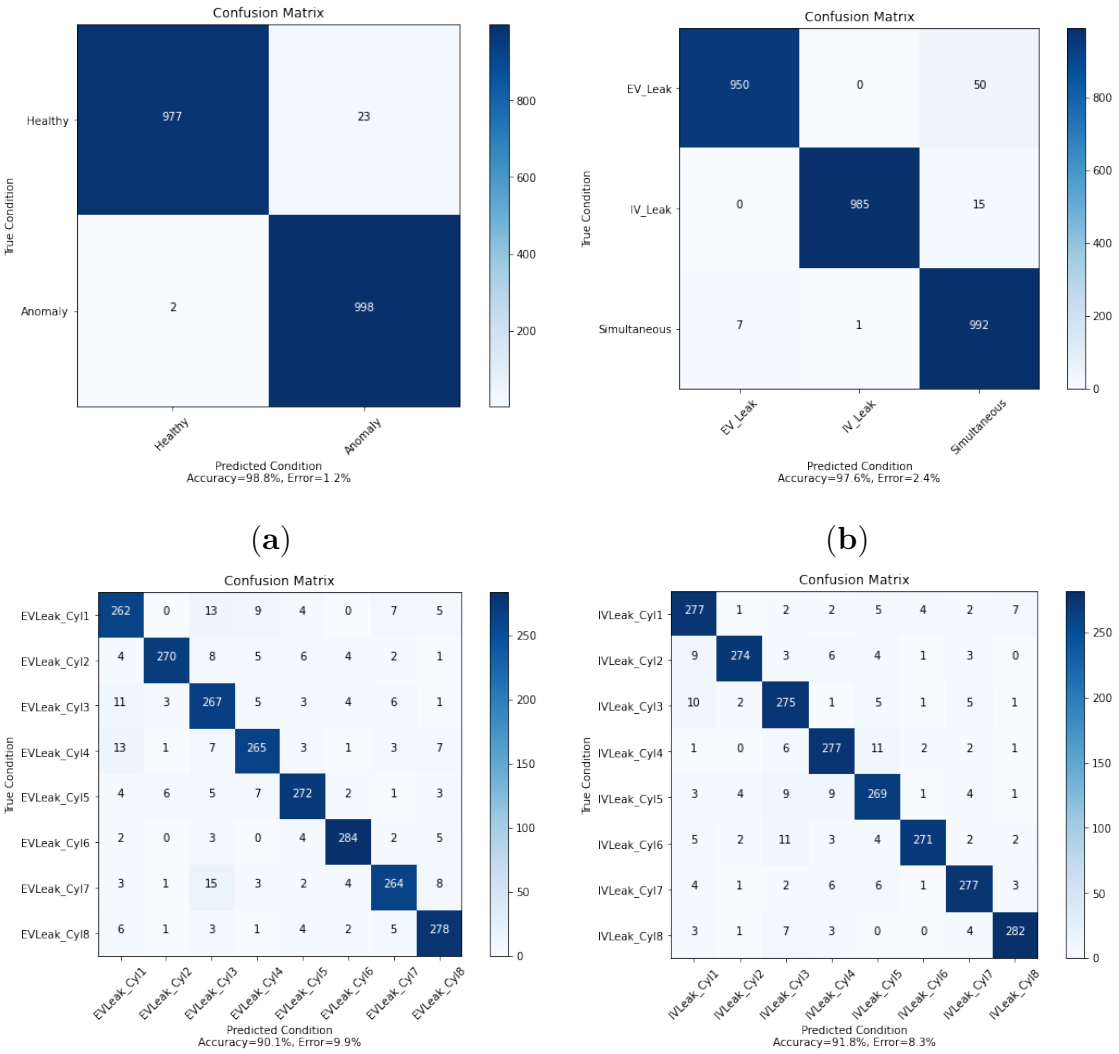


Figure 7.2: Case 1 – Anomaly diagnosis results (SVM) (**a**) Anomaly detection (**b**) Anomaly identification (**c**) Anomaly isolation ($Leak_{EV}$) (**d**) Anomaly isolation ($Leak_{IV}$)

The application test results indicate that the generated datasets are suitable for developing data-driven models. This is supported by the fact that the accuracy of the anomaly diagnosis models closely matched the values reported in the pertinent literature [66, 242], which are in the range of 91.7–99.5% employing actual measurement data for binary and multi-classification with SVM.

7.2 Case Study 2 - Data-driven Methods for PHM Model Development

The commonly used methods in pertinent PHM studies were tested using simulation-generated datasets. The employed datasets represent diverse operating profiles, including combinations of either similar or different initial conditions for each component and high or low stochastic degradation trajectories. In addition, the PHM model management framework was tested to confirm its ability to maintain the trustworthiness of the PHM model within stochastic degradation profiles.

7.2.1 2–a - Fault Diagnosis

The five classification algorithms were employed to detect, identify, and isolate the engine exhaust and intake valve faults. For testing the algorithms, same number of samples for each class was extracted. The confusion matrices for each sub-task were presented in Figure 7.3, 7.4, 7.5, and 7.6. The accuracy of the algorithms for each sub-task was summarised in Table 7.7.

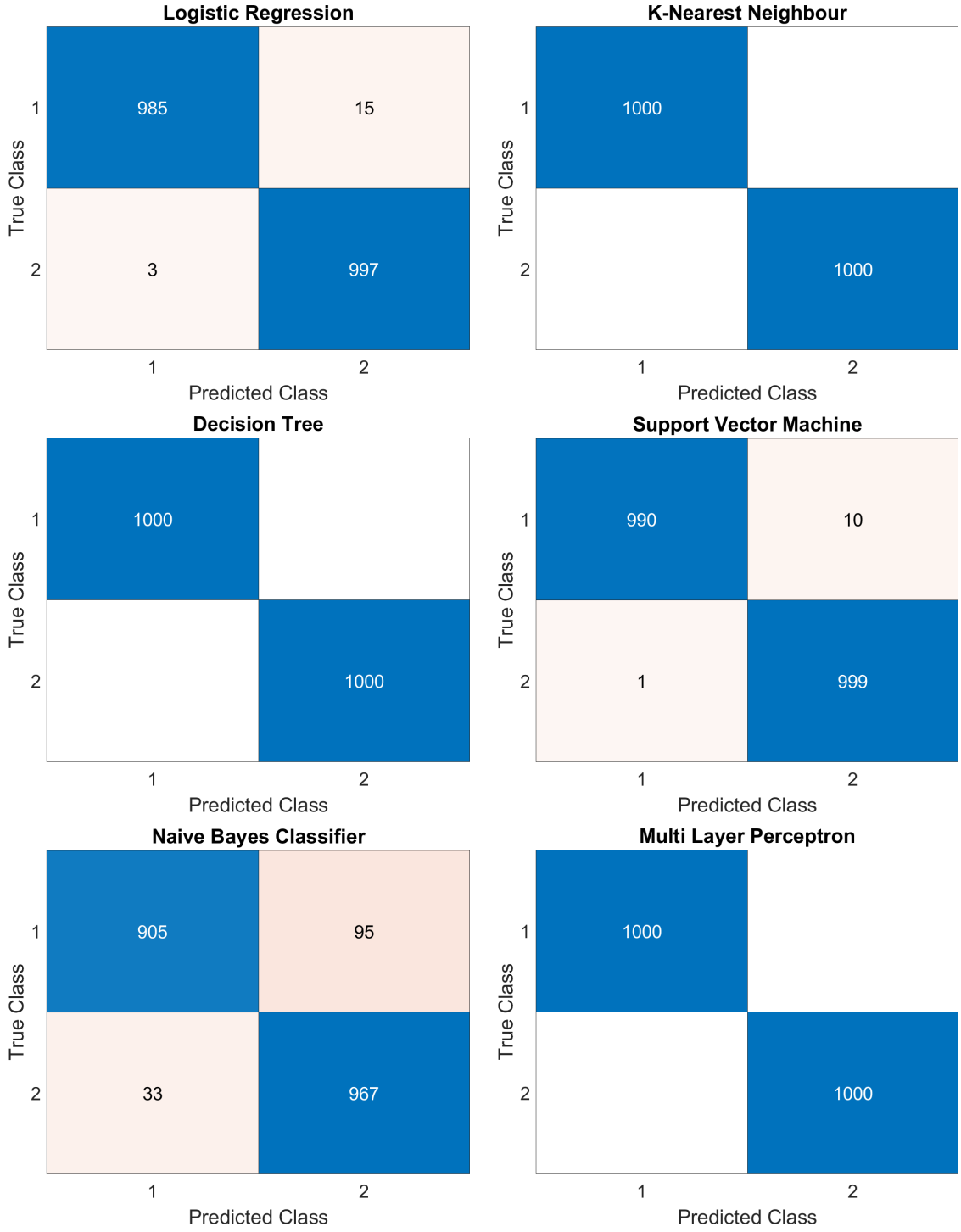


Figure 7.3: Fault detection result [1]

[1]Class 1: Healthy, Class 2: Anomaly

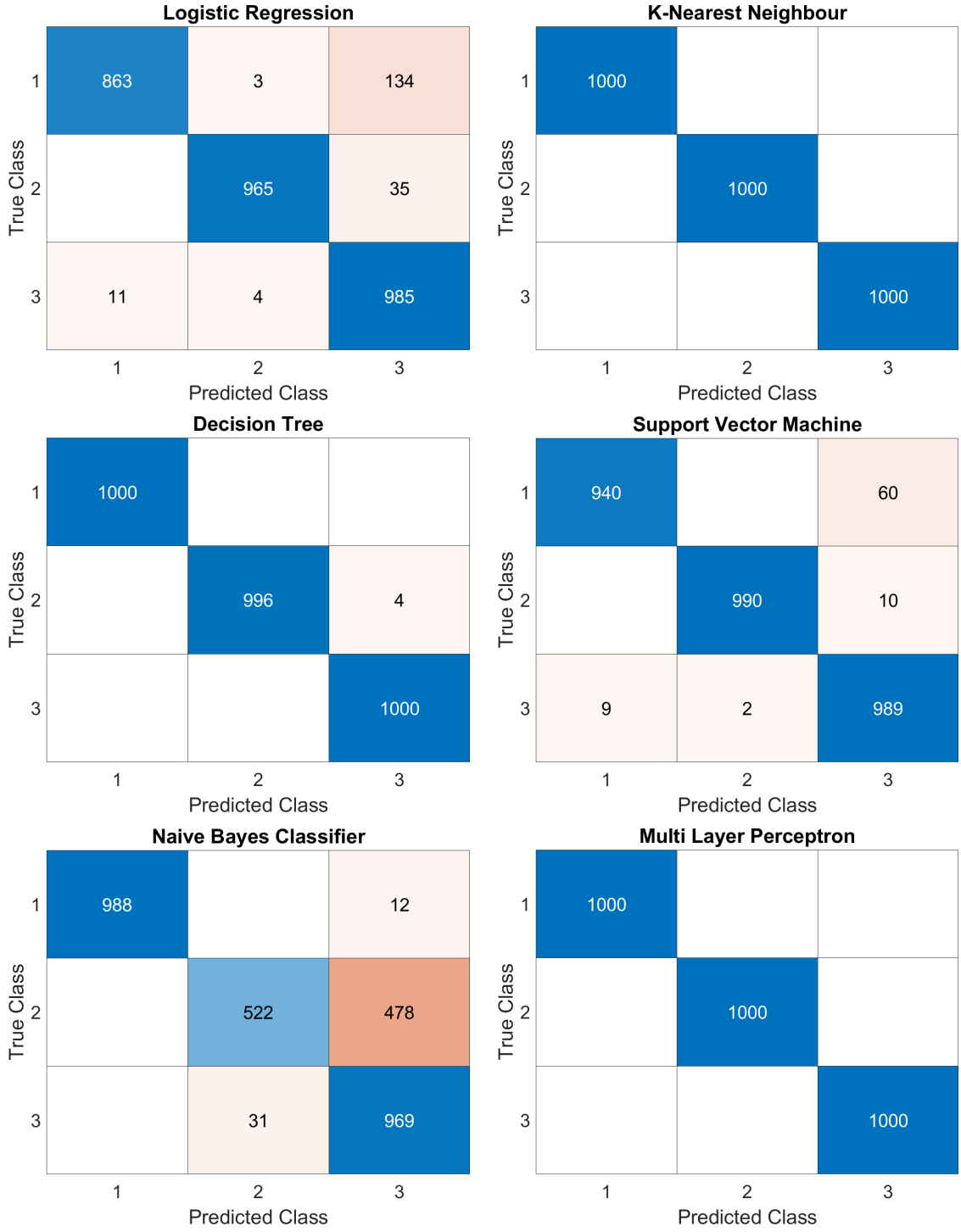


Figure 7.4: Fault identification result [2]

[2]Class 1: $Leak_{EV}$, Class 2: $Leak_{IV}$, Class 3: Simultaneous anomalies

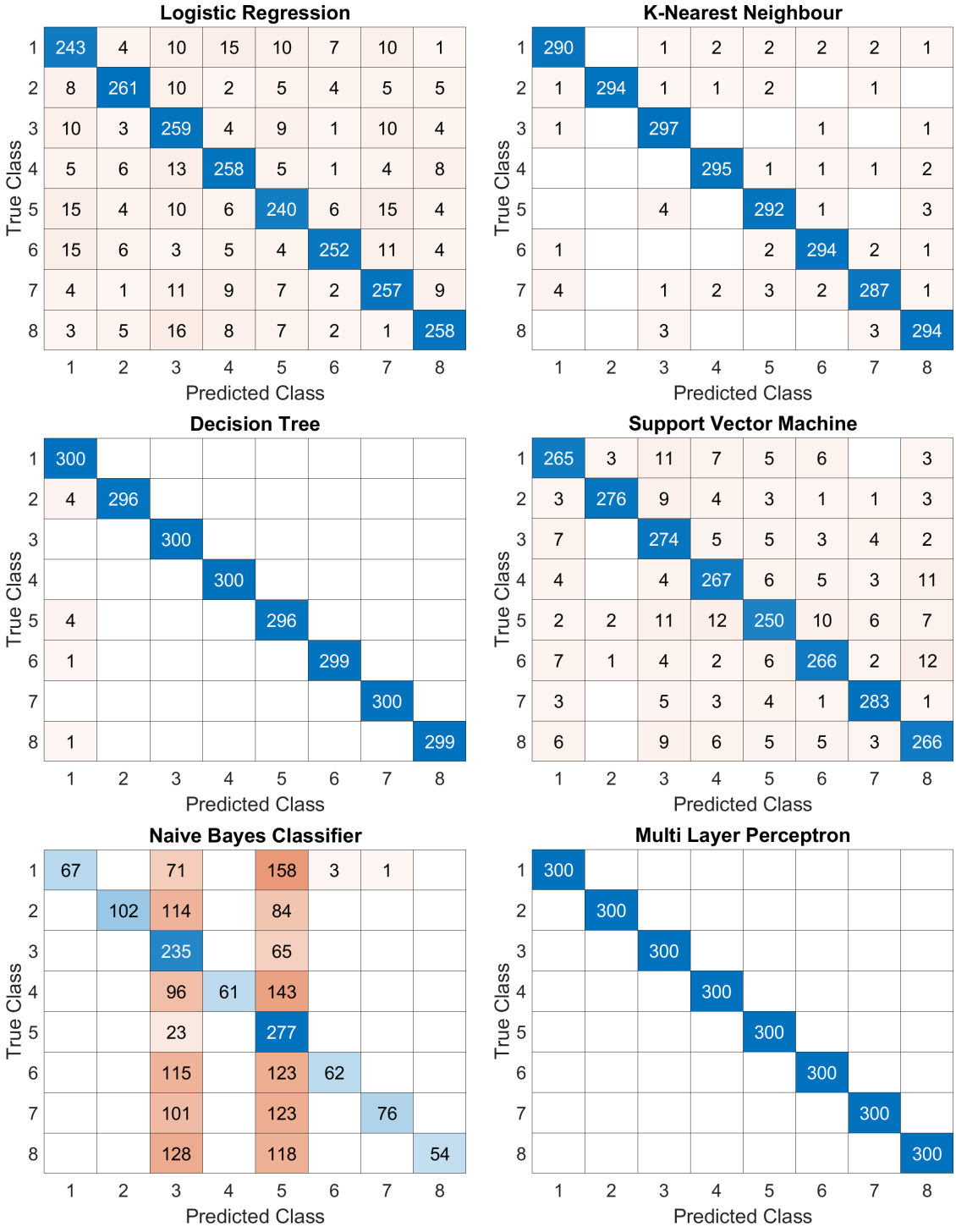


Figure 7.5: Fault isolation result – Exhaust valve [3]

[3]Class 1: Leakage in Cylinder 1, Class 2: Leakage in Cylinder 2, Class 3: Leakage in Cylinder 3, Class 4: Leakage in Cylinder 4, Class 5: Leakage in Cylinder 5, Class 6: Leakage in Cylinder 6, Class 7: Leakage in Cylinder 7, Class 8: Leakage in Cylinder 8

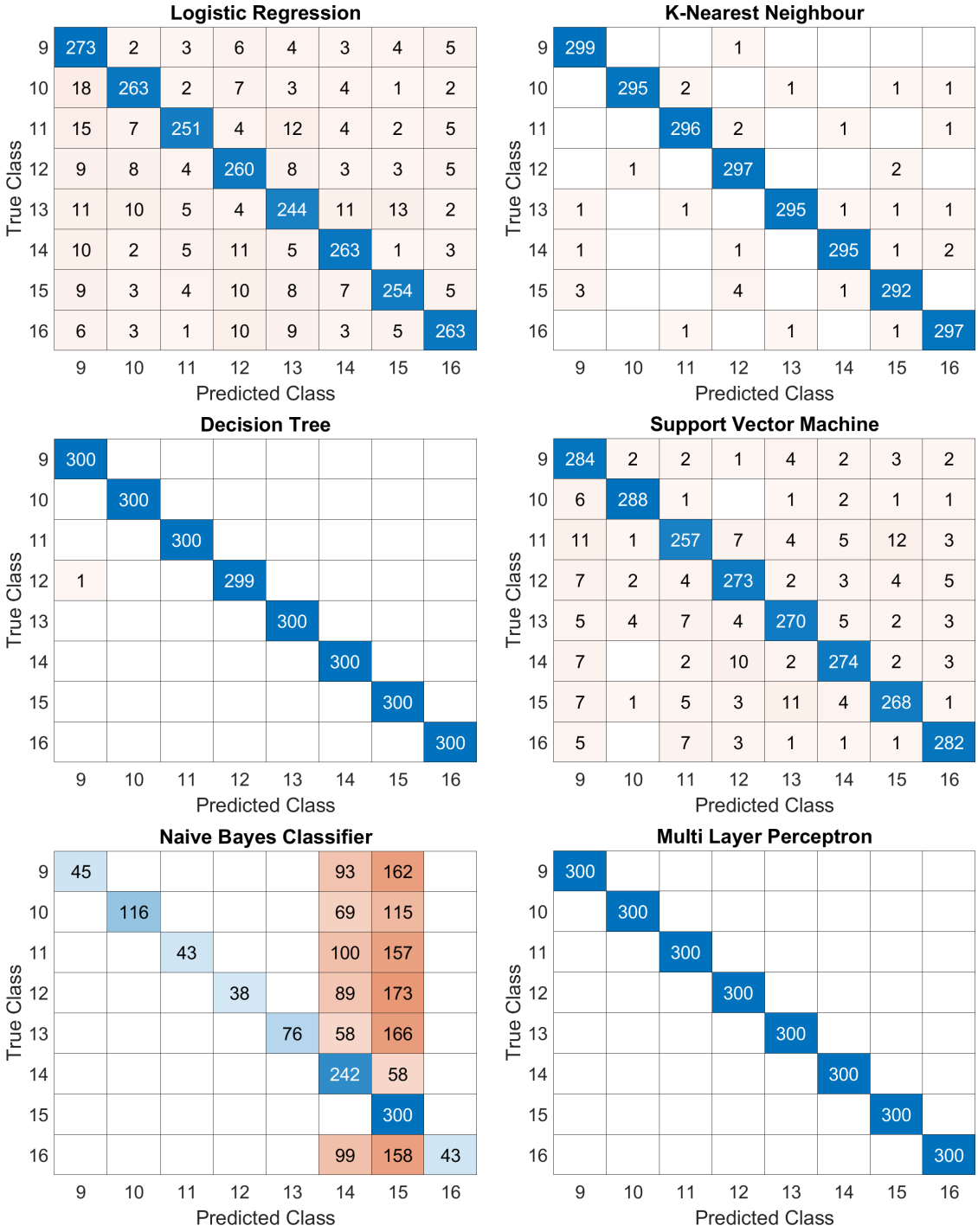


Figure 7.6: Fault isolation result – Intake valve [4]

[4]Class 1: Leakage in Cylinder 1, Class 2: Leakage in Cylinder 2, Class 3: Leakage in Cylinder 3, Class 4: Leakage in Cylinder 4, Class 5: Leakage in Cylinder 5, Class 6: Leakage in Cylinder 6, Class 7: Leakage in Cylinder 7, Class 8: Leakage in Cylinder 8

Table 7.7: Fault diagnosis result table – classification accuracy with testing datasets

	Detection	Identification	Isolation (EV)	Isolation (IV)
LR	99.1%	96%	84.5%	86.3%
KNN	100%	100%	97.6%	98.6%
DT	100%	100%	99.6%	99.9%
SVM	99.5%	97.8%	89.5%	91.5%
NB	93.6%	86.7%	38.9%	37.6%
MLP	100%	100%	100%	100%

KNN, DT, and MLP had no classification error in both detection and identification sub-tasks. For isolation sub-tasks, MLP consistently maintained a perfect score, while KNN performed slightly lower in exhaust valve leakage (97.6%) and intake valve leakage (98.6%). DT also had slightly lower accuracy in both anomalies, as 99.6% in exhaust valve leakage and 99.9% in intake valve leakage. SVM and LR also performed well with accuracy above 80% in all sub-tasks, however, NB showed the lowest accuracy across all sub-tasks, particularly in isolation sub-tasks.

7.2.2 2–b – Health Prognosis

Methods for HI Construction

The performance metrics of the selected HI construction methods are summarised in Table 7.8. The HI construction result plots for each dataset are visualised in Figure A.1,A.2,A.3, and A.4.

KNN and SVR exhibited shorter run times than other methods, whereas LSTM and MLP exhibited the longest run time in most datasets. In the monotonicity criterion, LR had the highest values for most datasets excluding Dataset 4, and KNN had the lowest values for all datasets. In the trendability criterion, SVR and RF had the highest value for most datasets, and LSTM had the lowest value for most datasets excluding Dataset 4. In the prognosability criterion, SVR had the highest value for Datasets 3 and 4, and LSTM had the lowest value for most datasets excluding Dataset 1. In the error criterion, MLP achieved the highest R^2 for all

Table 7.8: Performances of the selected HI construction methods

Dataset	Method	Run Time [s]	Monotonicity [–]	Trendability [–]	Prognosability [–]	R^2 [–]
Dataset 1	LR	11.6	**0.570**	**0.998**	0.973	0.952
	KNN	**6.3**	*0.527*	**0.998**	*0.962*	*0.894*
	RF	19.8	0.548	**0.998**	0.965	0.947
	SVR	7.6	0.540	**0.998**	0.973	0.938
	LSTM	*103.6*	0.548	*0.986*	**0.976**	0.988
	MLP	52.8	0.558	0.997	0.967	**0.993**
Dataset 2	LR	13.4	**0.558**	0.994	0.932	0.931
	KNN	11.7	*0.490*	0.997	0.960	*0.676*
	RF	24.3	0.513	**0.998**	**0.969**	0.808
	SVR	**5.7**	0.533	**0.998**	0.967	0.766
	LSTM	46.9	0.537	*0.956*	*0.835*	0.986
	MLP	*51.2*	0.542	0.985	0.835	**0.993**
Dataset 3	LR	7.3	**0.556**	0.993	0.964	0.958
	KNN	**6.3**	*0.497*	**0.998**	0.959	0.812
	RF	19.5	0.554	**0.998**	0.964	0.869
	SVR	7.0	0.544	0.997	**0.973**	*0.766*
	LSTM	*55.2*	0.536	*0.967*	*0.957*	0.987
	MLP	52.8	0.544	0.987	0.956	**0.995**
Dataset 4	LR	12.6	**0.547**	0.981	0.924	0.845
	KNN	9.5	*0.505*	0.998	0.958	*0.728*
	RF	22.7	0.553	0.998	0.966	0.781
	SVR	**5.7**	0.523	**0.999**	**0.973**	0.751
	LSTM	*63.4*	0.533	*0.956*	*0.875*	0.984
	MLP	37.1	0.539	0.961	0.868	**0.991**

***Bold**: Best result for each criterion

**Italic* : Worst result for each criterion

datasets. KNN had the worst error for most datasets excluding Dataset 3. The error criterion results exhibit a similar trade-off with the monotonicity criterion.

In the monotonicity criterion, LR had the highest average value but was below 0.6, indicating that over 40% of constructed HI trends are unrealistic for degradation processes. Most models scored above 0.9 in the trendability and prognosability criteria. SVR achieved the highest average values of 0.998 in trendability and 0.971 in prognosability but did not attain the highest error metric, with 0.805 on R^2 value, which is lower than that of deep learning methods. This inconsistency implies that trendability and prognosability criteria did not adequately evaluate the correctness of the HI construction methods.

Although deep learning methods had longer run times, they exhibited higher R^2 values. MLP achieved the highest average R^2 value of 0.993. The best-performing methods varied depending on the employed evaluation criteria, however, it was noteworthy that the most accurate method (MLP) did not attain the highest score in the other performance metrics.

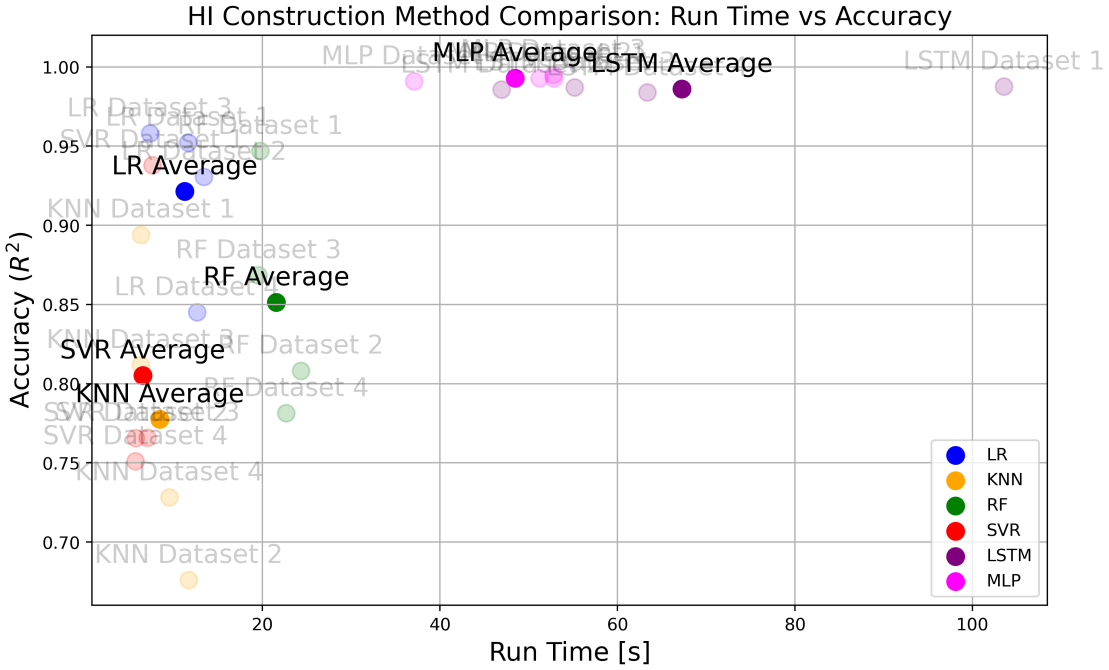


Figure 7.7: Case 2-b – Method comparison result for HI construction

Figure 7.7 visualised the run time and error of the selected methods to identify the strengths and weaknesses of each method. The top right corner is the ideal result with the minimum computational cost and the maximum R^2 value. The methods were grouped into two groups: the neural network group and the regressor group.

The neural network group had high R^2 values but relatively high computational cost, whereas the regressor group had low computational cost but relatively low R^2 values. If error is prioritised, MLP is recommended because of its highest R^2 value and relatively low computational cost. Conversely, if computational cost is prioritised, LR is recommended for its fast computation and relatively high R^2 value.

Effect of Degradation Characteristics The different initial conditions decreased the R^2 value of the HI construction methods. SVR showed the highest R^2 value drop of 0.093 on average, while deep learning methods were not significantly affected. Performance changes in the other evaluation criteria were less than 0.03 in all methods. The different initial conditions do not significantly impact the performance of the HI construction model.

The increasing stochasticity of the degradation trajectories increased error of the machine learning methods (LR, KNN, RF, and SVR), while it did not affect the error of the deep learning methods (LSTM and MLP). KNN was the most affected method, and its R^2 value dropped 0.151 on average. However, the other evaluation criteria did not show any significant reduction. In the prognosability criterion, deep learning methods decreased more than 0.1 value. These results indicate that increasing stochasticity leads to increased error in HI construction models, making it more challenging to develop RUL estimation models.

Methods for HI Forecast

Deterministic Methods The average performance of the deterministic methods was examined through 30 runs per method. The performance metrics of the selected HI forecast methods are summarised in Table 7.9, and the result plots are presented in Figure A.5,A.6,A.7, and A.8.

ARIMA showed the cheapest run time for forecasting HI, and the multiple runs produced consistent results. However, R^2 value for ARIMA was negative for all datasets, which means that the forecasted HI trajectories were not appropriate for RUL estimation. CNN and LSTM exhibited the most expensive run times for all datasets. These were more than 1.5 times longer than the run time of MLP. In the RUL estimation error criterion, LSTM had the smallest error, and CNN had the

Table 7.9: Performances of the selected HI forecast method (deterministic)

Dataset	Method	Run Time [s]	RUL Error [h]	RUL MoE [h]	R^2 [–]	Variability [–]
Dataset 1	ARIMA	204	10698.8	0.0	-1.377	0.000
	LSTM	3672	**1790.8**	**6.7**	0.954	**0.004**
	CNN	*3779*	*1882.2*	*43.4*	*0.917*	*0.018*
	MLP	**2102**	1803.0	9.4	**0.956**	**0.004**
Dataset 2	ARIMA	162	12641.3	0.0	-1.934	0.000
	LSTM	*3478*	**1853.1**	**7.0**	0.833	**0.004**
	CNN	3348	*2066.2*	*43.5*	**0.834**	*0.017*
	MLP	**1842**	1881.6	21.2	*0.817*	0.011
Dataset 3	ARIMA	166	10986.0	0.0	-0.892	0.000
	LSTM	*3187*	**1522.9**	**5.8**	*0.977*	**0.003**
	CNN	2875	*1580.5*	*31.6*	**0.981**	*0.016*
	MLP	**1695**	1537.6	7.4	0.980	0.004
Dataset 4	ARIMA	178	8902.5	0.0	-1.308	0.000
	LSTM	3319	**1521.4**	**5.6**	**0.896**	**0.003**
	CNN	*3599*	*1635.1*	*37.5*	*0.807*	*0.018*
	MLP	**2015**	1563.5	17.3	0.882	0.010

***Bold**: Best result for each criterion

**Italic* : Worst result for each criterion

largest error for all datasets excluding ARIMA. These results align with the RUL estimation MoE criterion. LSTM had the smallest MoE, whereas CNN had the largest MoE. In the prediction error criterion, the method with the highest R^2 varied for each dataset, however, all methods, excluding ARIMA, achieved R^2 greater than 0.8 for all datasets. On average, LSTM exhibited the highest R^2 value, whereas CNN had the lowest R^2 value. LSTM exhibited the lowest variability, whereas CNN had the highest variability.

ARIMA is the fastest method to predict the HI trend of future time slices, however, it is not appropriate for prognostics due to its imprecision. MLP exhibited a lower computational cost compared to the other deep learning methods, while the R^2 value was not significantly different from the other methods, excluding ARIMA. LSTM exhibited the highest average R^2 value of 0.915.

In Figure 7.8, three groups were identified: ARIMA, MLP, and the others. ARIMA was excluded from recommendations due to its failure to make accurate predictions. MLP exhibited lower computational costs across all datasets. LSTM demonstrated higher average R^2 value than the other methods. If computational cost is prioritised, MLP is recommended for its fast computation and relatively high R^2 value. Conversely, if error is prioritised, LSTM is recommended for their remarkable R^2 value. However, HI forecasting with LSTM requires high computational cost.

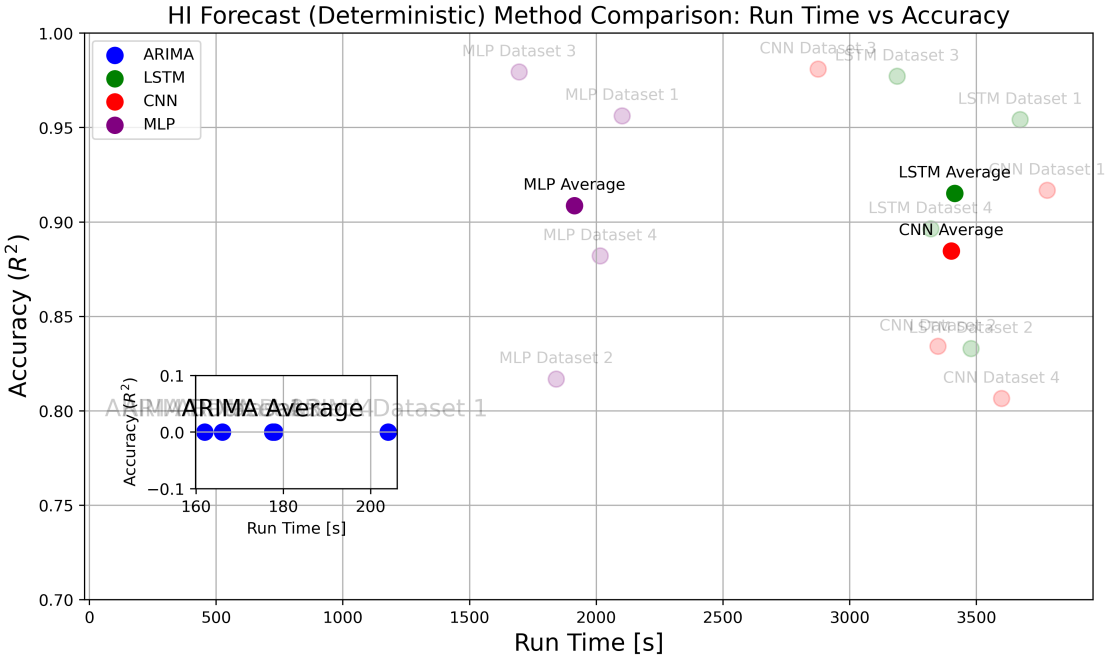


Figure 7.8: Case 2-b – Method comparison result for HI forecast (deterministic)

Table 7.10: Performances of the selected HI forecast methods (probabilistic)

Method	Dataset	Run Time [sec]	RUL Error [h]	RUL MoE [h]	R^2 [–]	Variability [–]
Ensemble	Dataset 1	3751	1812.7	9.2	0.955	0.005
	Dataset 2	3118	1897.6	16.6	0.831	0.009
	Dataset 3	2887	1540.3	8.9	0.980	0.005
	Dataset 4	2862	1556.1	22.8	0.887	0.015
Variational Inference	Dataset 1	781	1785.2	32.5	0.929	0.036
	Dataset 2	825	1797.7	52.6	0.809	0.032
	Dataset 3	752	1543.9	47.5	0.975	0.029
	Dataset 4	744	1552.9	45.8	0.862	0.033
Dropout	Dataset 1	182	1931.2	12.9	0.902	0.011
	Dataset 2	139	1929.4	17.5	0.823	0.010
	Dataset 3	172	1561.9	23.1	0.963	0.009
	Dataset 4	142	1579.0	21.2	0.873	0.010

Probabilistic Methods The performance metrics of the selected probabilistic methods are summarised in Table 7.10.

In the run time criterion, the dropout method had the smallest run time, whereas the ensemble method had the heaviest run time. Despite the heaviest run time, the ensemble method had the highest R^2 value and the lowest MoE and Variability for both RUL estimation and HI prediction. However, the RUL errors and the R^2 values of the three probabilistic methods did not have significant differences on average. The difference between the method with the largest RUL error (dropout) and the method with the smallest RUL error (variational inference) was 80.5 hours, which is less than 5% of the estimated RUL values. The R^2 values were 0.913 for the ensemble method,

0.894 for the variational inference method, and 0.890 for the dropout method. The variational inference method exhibited the largest values in the uncertainty criteria of MoE and variability.

In the comparison of probabilistic methods, the RUL estimation and prediction errors of the considered methods were not significantly different. In the RUL error criterion, the variational inference method had the smallest error of 1669.9 hours, and the dropout method had the largest error of 1750.4 hours on average. In the prediction error criterion, the ensemble method had the highest score of 0.913 and the dropout method had the lowest score of 0.890 on average. On the other hand, the uncertainty-related criteria showed the distinguishable differences between the considered methods. The variational inference had 44.6 hours in the MoE criterion and 0.033 in the variability criterion on average. These values were more than two times higher than the other methods. Lastly, the run times of the considered methods showed remarkable differences between the considered methods. The dropout method runs 158.75 seconds, the variational inference method runs 775.5 seconds, and the ensemble method runs 3154.5 seconds for generating 30 different prediction outputs on average.

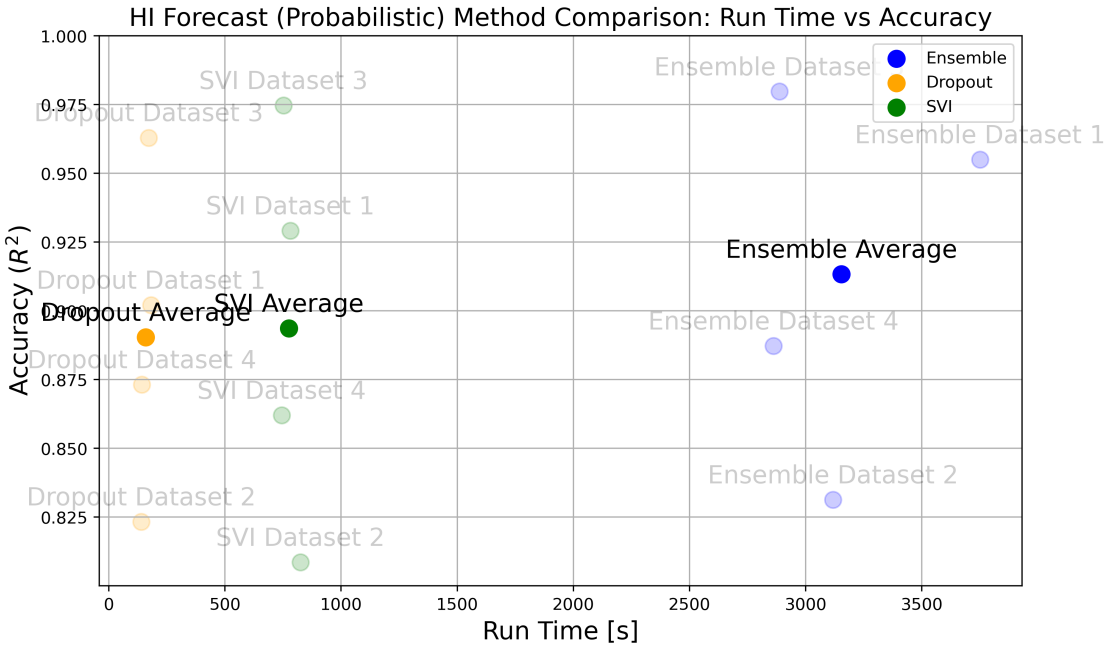


Figure 7.9: Case 2-b – Method comparison result for HI forecast (probabilistic)

In Figure 7.9, two groups are identified: Ensemble method and the others. Ensemble method attained the highest average R^2 value, however, it is not recommended

due to its exceptionally high computational cost. Considering both prediction error and computational cost, dropout method is recommended for its fast computation and acceptable prediction error for HI forecast.

Effect of Degradation Characteristics All methods did not show significant changes in prediction error when exposed to different initial condition profiles, with the R^2 score increasing by a maximum of 0.044. ARIMA showed the most remarkable performance improvement; however, the performance was still negative for the HI forecast sub-task. The uncertainty of the HI forecast and RUL estimation was reduced in the deep learning methods. The initial conditions have negligible influence on the forecast model, as it independently predicts each cylinder's HI trend.

In all methods, increasing stochasticity led to reductions in prediction error by at least 0.1 in the R^2 score, as well as increases in RUL errors. However, the changes were not significant, with most increments being up to 119 hours, as observed in the CNN. The uncertainty of the HI forecast and RUL estimation increased with the high stochasticity, except for LSTM. Simple deep learning models like MLP have a relatively large impact, whereas specialised deep learning models like LSTM and CNN have a relatively small impact. The run times of the methods increased and decreased depending on the methods, but the changes were not significant.

Methods Recommendation and Integration

The comparative assessment results indicate that each method has its strengths and weaknesses, and the most effective method varies depending on the circumstances and requirements. In this research, the recommended methods were integrated considering both prediction error and computational cost. For both HI construction and forecast sub-models, MLP is recommended due to its high R^2 value and relatively low computational cost. The integrated method is used to develop an RUL estimation model, which is applicable within PHM systems for marine engines.

The performance metrics of the PHM model developed that integrates the MLP and the dropout method are summarised in Table 7.11, and the result plots for each dataset are visualised in Figure A.9,A.10,A.11, and A.12.

The run time for the entire RUL estimation process took 220 seconds on aver-

Table 7.11: PHM model performances with recommended methods

Dataset	Run Time [sec]	C-Err[a] [–]	RUL Error [h]	RUL MoE [h]	F-Err[b] [–]	Variability [–]
Dataset1	209	0.969	1755.7	7.6	0.913	0.008
Dataset2	218	0.954	1982.6	7.0	0.777	0.008
Dataset3	229	0.969	1686.7	8.3	0.900	0.007
Dataset4	223	0.947	1680.4	7.7	0.782	0.007

[a] HI Construction Sub-model Error (R^2)
[b] HI Forecast Sub-model Error (R^2)

age.The combined RUL estimation model exhibited an average R^2 of 0.960 for the HI construction sub-task and 0.843 for the HI forecast sub-task.

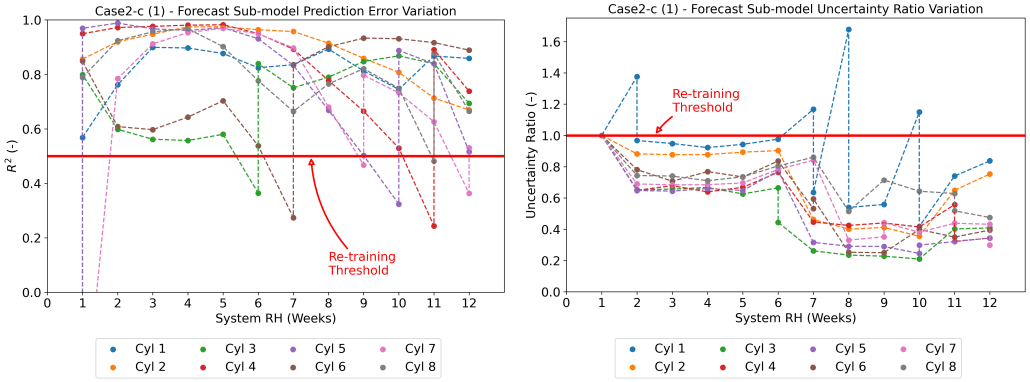
7.2.3 2–c – PHM Model Management with Maintenance Decision Making

In Case 2-c, the PHM model management and maintenance decision-making processes were demonstrated using the health prognosis sub-model with MLP for the HI construction sub-task and BNN for the HI forecast sub-task. The prognosis sub-model was trained using development datasets with less stochastic degradation patterns and the same initial conditions, then tested using application datasets with more stochastic degradation patterns. Two application datasets, with same and different initial conditions respectively, were employed to investigate the impact of the initial condition for investigating the impact of initial conditions on the performance of the PHM model management and maintenance decision-making processes.

Same initial conditions

Figure 7.10 presents the prediction error and uncertainty ratio variations of the HI forecast sub-models in Case 2-c (1). It is inferred that the performances of the HI forecast sub-models were managed above the pre-defined re-training threshold throughout the 12 weeks of operation apart from the first week. The HI forecast sub-models were re-trained seven times based on the error criteria and four times based on the robustness criteria during the 12 weeks of operation. In the first week, the HI forecast sub-models for Cylinders 3, 5, and 7 exhibited negative R^2 values and required re-training. Despite re-training in the first week, the sub-model for Cylinder 7 did not achieve the minimum required R^2 value. However, the R^2 value

was improved and retained above the minimum requirement in the subsequent weeks. The lowest R^2 value of 0.24 was observed for Cylinder 4 at Week 11; however, it increased to 0.89 after re-training.



(a) Prediction error variations **(b)** Robustness variations

Figure 7.10: Case 2-c (1)– HI forecast sub-models performance variations

The HI forecast sub-model results and re-training decision for Cylinder 6 are summarised in Table 7.12, whereas the HI variation trends for Cylinder 6 are presented in Figure 7.11. The data-driven model management process updated the HI forecast sub-model for Cylinder 6 once in week 7 during 12 weeks of operation. A maintenance decision was made at Week 12 (2,184 hours) as the HI of Cylinder 6 was forecasted to cross the failure threshold at 2,701 hours. The RUL at the time of decision-making was 517 hours, which was smaller than the estimated time of arrival (ETA) to the next port. The margin of error with 95% confidence was reduced from 5 hours in Week 1 to 2 hours in Week 12.

Table 7.12: Case 2-c (1) – HI forecast sub-model results and management for Cylinder 6

Results	Unit	W1[a]	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12
Re-training	–	No	No	No	No	No	No	Yes	No	No	No	No	No
R^2	–	0.85	0.61	0.60	0.64	0.70	0.54	0.84	0.90	0.93	0.93	0.92	0.89
Failure RH [b]	hours	3,024	3,034	3,033	3,052	3,035	3,034	2,701	2,701	2,700	2,701	2,701	2,701
RUL	hours	2,688	2,530	2,361	2,212	2,027	1,858	1,357	1,189	1,020	853	685	517
RUL Error	hours	– 282	– 292	– 291	– 310	– 293	– 292	41	41	42	41	41	41
MoE [c]	hours	5	4	4	4	4	4	3	1	1	2	2	2

[a] Week Number
[b] Actual Failure RH: 2,742 (hours)
[c] Margin of Error

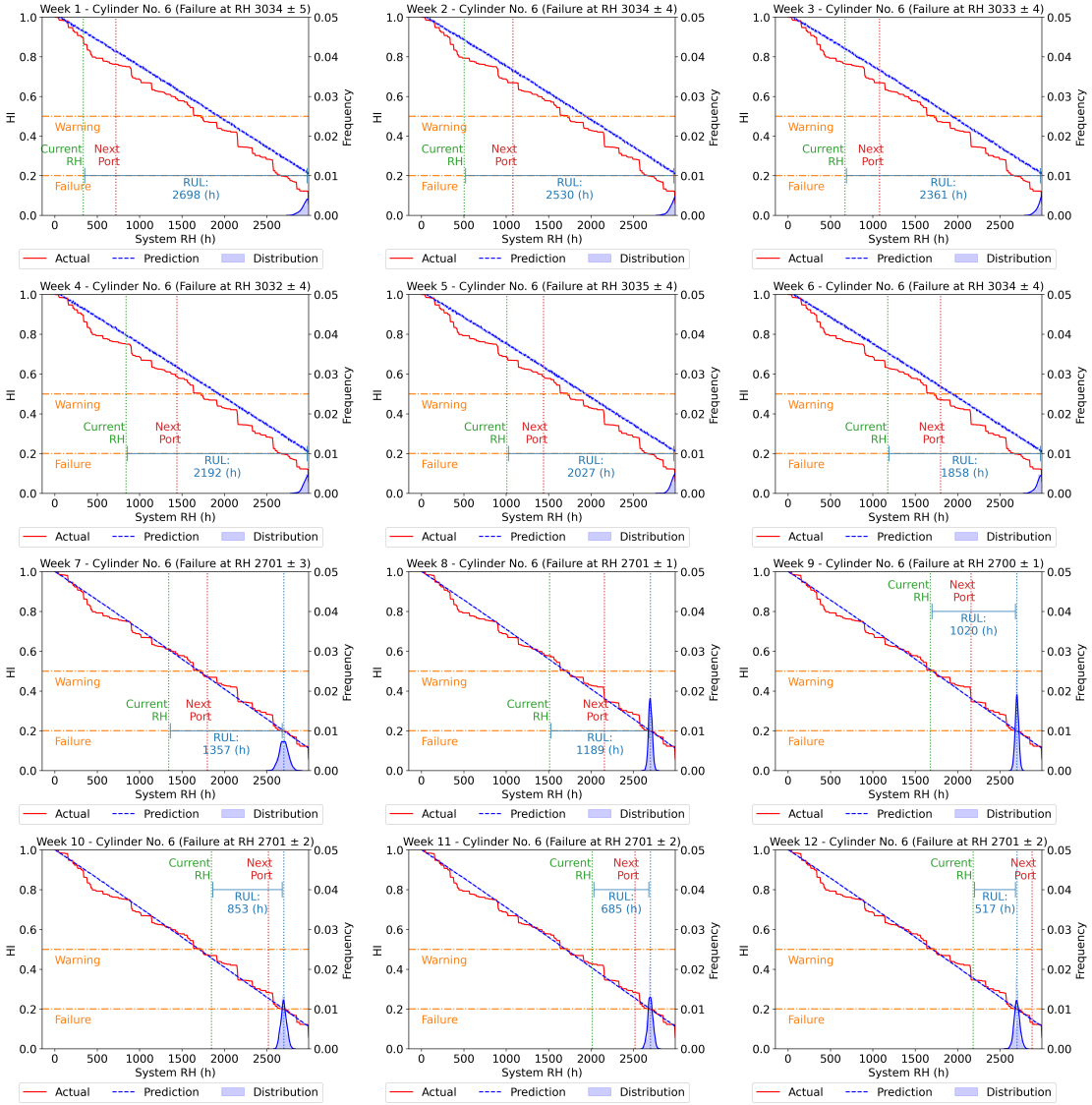
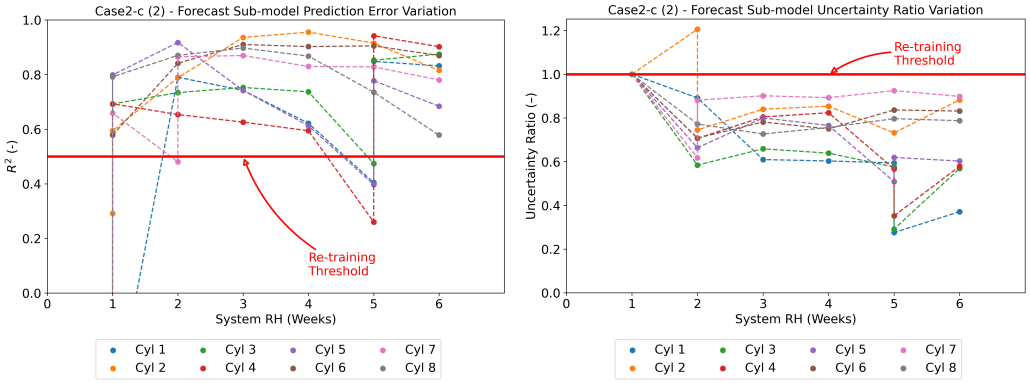


Figure 7.11: Case 2-c (1) – HI variation trend for Cylinder 6

Different initial conditions

Figure 7.12 presents the prediction error and uncertainty ratio variations of the HI forecast sub-models in Case 2-c (2). A negative R^2 value was observed for Cylinder 1 despite re-training in week 1, but the R^2 value improved to 0.79 in week 2. In week 5, the HI forecast models for Cylinders 1, 3, 4 and 5 did not achieve the minimum required R^2 value of 0.5. However, the data-driven model management process improved the R^2 value and retained it above the minimum requirement for all cylinders in the subsequent weeks. The uncertainty ratio of the HI forecast sub-model for Cylinder 2 exceeded the re-training threshold in week 2 but was reduced to 0.7 through re-training.



(a) Prediction error variations **(b)** Robustness variations

Figure 7.12: Case 2-c (2) – HI forecast sub-models performance variations

The HI forecast sub-models results and re-training decision for Cylinder 3 are summarised in Table 7.13, whereas the HI variations for Cylinder 3 are presented in Figure 7.13. The data-driven model management process updated the HI forecast sub-models twice during operations, in weeks 1 and 5. In Week 6 (1,176 hours), a maintenance decision was made for Cylinder 3 based on the forecasted HI, which indicated it would cross the failure threshold at 1,706 hours. The RUL at the time of decision-making was 530 hours, which was smaller than the ETA to the next port. The margin of error with 95% confidence was reduced from 7 hours in Week 1 to 4 hours in Week 6.

Table 7.13: Case 2-c (2) – HI forecast sub-model results and management decisions for Cylinder 3

Results	**Unit**	**W1**[a]	**W2**	**W3**	**W4**	**W5**	**W6**
Re-training	–	Yes	No	No	No	Yes	No
R^2	–	0.69	0.73	0.75	0.74	0.85	0.87
Failure RH	hours	2,204	2,200	2,197	2,197	1,705	1,706
RUL	hours	1,868	1,696	1,525	1,357	697	530
RUL Error	hours	– 512	– 508	– 505	– 505	– 13	– 14
MoE [c]	hours	7	4	5	5	2	4

[a] Week Number
[b] Actual Failure RH: 1,692 (hours)
[c] Margin of Error

7.3 Case Study 3 - Demonstration of PHM Model and Frameworks

The integrated PHM model, which employs frameworks for trustworthy DT and PHM models, was demonstrated through a complex degradation scenario, characterised by

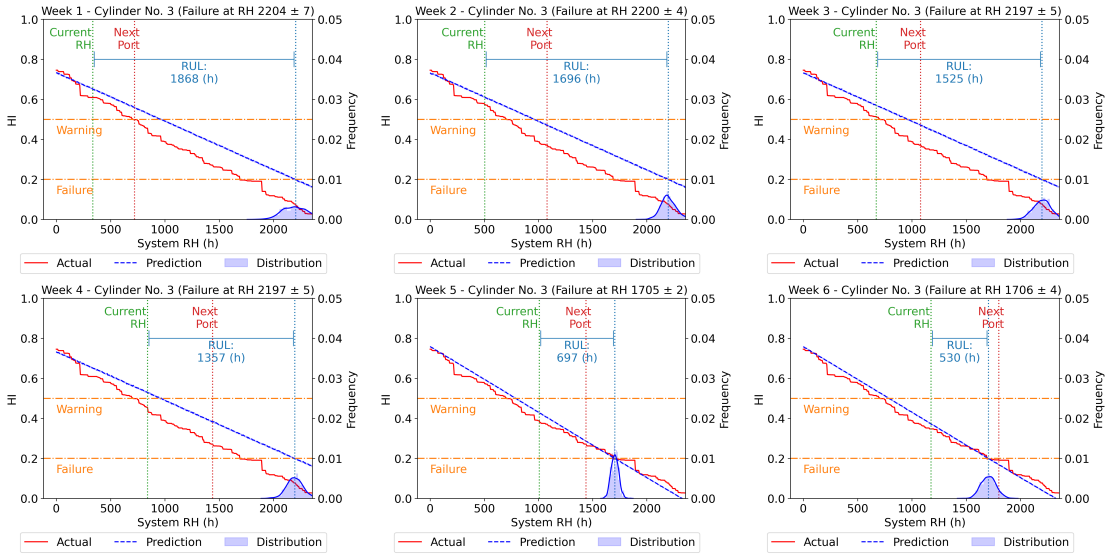


Figure 7.13: Case 2-c (2) – HI variation trends for Cylinder 3

simultaneous degradation of multiple components.

7.3.1 Simulation-based Data Generation

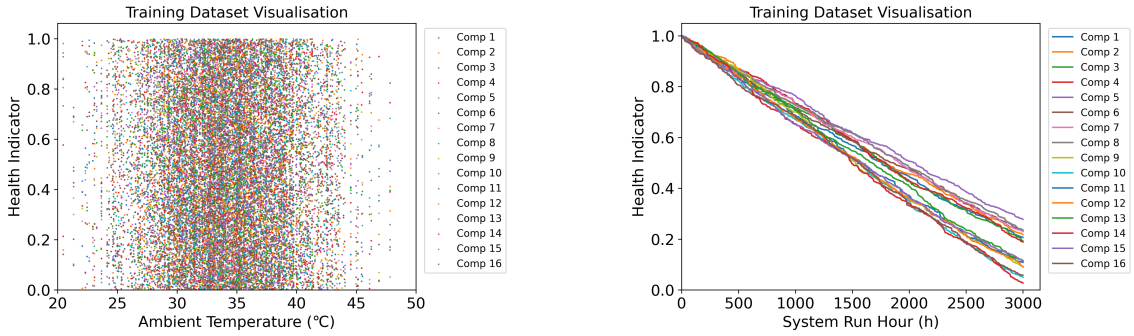
For extending operating envelope of limited envelope dataset, the trustworthy DT generated three datasets for training and testing the PHM model: HI construction training, HI forecast training, and Testing datasets. For the HI construction training dataset, the Latin hypercube method [181] is utilised to sample 6400 operating points considering the degradation range and engine operating envelope. For the HI forecast training dataset and testing dataset, the gamma process model is utilised to generate stocahstic degradation datasets with 500 operating points and 394 operating points, respectively. The employed degradation stochasticity and the initial HI for each dataset are presented in Table 7.14.

Table 7.14: Degradation characteristics for training and testing datasets

Dataset Types	Degradation Stochasticity	Initial HI							
		Comp 1[b]	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8
		Comp 9	Comp 10	Comp 11	Comp 12	Comp 13	Comp 14	Comp 15	Comp 16
Training	$\Gamma(1, 1)$ [a]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Testing	$\Gamma(0.1, 10)$	1.00	0.96	0.93	0.90	0.86	0.90	0.93	0.96
		0.86	0.90	0.93	0.96	1.00	0.96	0.93	0.90

[a] Γ: Gamma Distribution (shape, scale)
[b] Component Number

The training degradation profile includes low stochastic degradation trajectories with similar initial conditions for each component, whereas the testing degradation



(a) Case 3 – HI construction training dataset **(b)** Case 3 – HI forecast training dataset

Figure 7.14: Case 3 – Training datasets visualisation

profiles are combinations of different initial conditions for each component and high stochastic degradation trajectories. The training datasets for HI construction and HI forecast are visualised in Figure 7.14.

The degradation datasets for testing the selected methods are visualised in Figure 7.15. This dataset is useful for evaluating the performance of trained RUL estimation models in complex degradation scenarios, characterised by both irregular degradation rates and maintenance schedules.

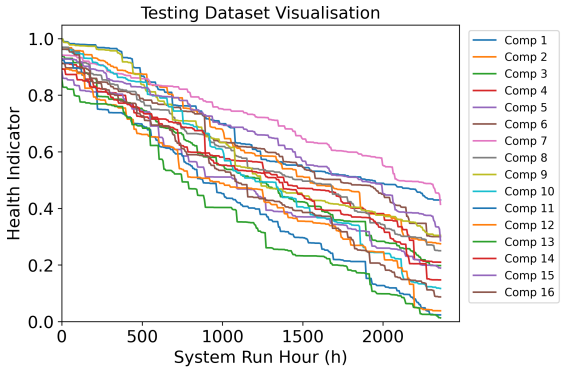


Figure 7.15: Case 3 – Testing datasets visualisation

7.3.2 PHM Model Development and Management

In Case 3, hyperparameters are searched using the Python Keras-tuner library, which comes with Bayesian optimisation and random search algorithms. Number of neurons is searched between 32 and 512, epoch is searched between 30 and 100, batch size is searched between 16 and 64, activation functions are selected among ReLu, Tanh, and Sigmoid, learning rate is searched between 0.0001 and 0.01, The selected hyperparameters for HI construction and forecast models are summarised in Table 7.15.

Table 7.15: Hyperparameter settings

Sub-model	**Fault Diagnosis**	**HI Construction**	**HI Forecast**
Methods	MLP	MLP	MLP-Dropout
Number of Layers	3	2	2
Hidden Size	256	480	384
Epoch	50	50	100
Batch Size	32	32	64
Activation	Sigmoid	ReLU	ReLU
Learning Rate	0.001	0.003	0.001

Table 7.16: PHM model performance table

Sub-models	**Performance**								
	Comp[a]1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8	Overall
	Comp 9	Comp 10	Comp 11	Comp 12	Comp 13	Comp 14	Comp 15	Comp 16	
Fault Diagnosis[b]	0.82	0.94	0.95	0.97	0.92	0.92	0.88	0.99	0.89
	0.94	0.99	0.82	0.86	0.74	0.87	0.74	0.93	
HI Construction[c]	0.99	0.96	1.00	0.99	0.98	0.98	0.96	0.99	0.97
	0.92	0.95	0.98	0.97	0.98	0.98	0.95	0.98	
HI Forecast[c]	1.00	0.85	0.99	0.96	0.97	0.96	0.95	0.98	0.95
	0.98	0.92	1.00	0.80	0.96	1.00	0.89	0.99	

[a] Component Number

[b] Percentage Accuracy

[c] Coefficient of Determination (R^2)

Table 7.16 provides a comprehensive summary of the performance metrics for individual sub-models associated with each component, in addition to presenting the overall performance. The fault diagnosis sub-model demonstrated an overall classification accuracy of 0.89, with the lowest accuracy of 0.74 at Components 13 and 15. The HI construction and forecast sub-models exhibited overall regression accuracies of 0.97 and 0.95, respectively. In the HI construction, all components had an R^2 value greater than 0.92. However, in the HI forecast, the lowest R^2 value was 0.8, occurring in Component 12, which employed as the re-training threshold in Case 3.

Fault Diagnosis

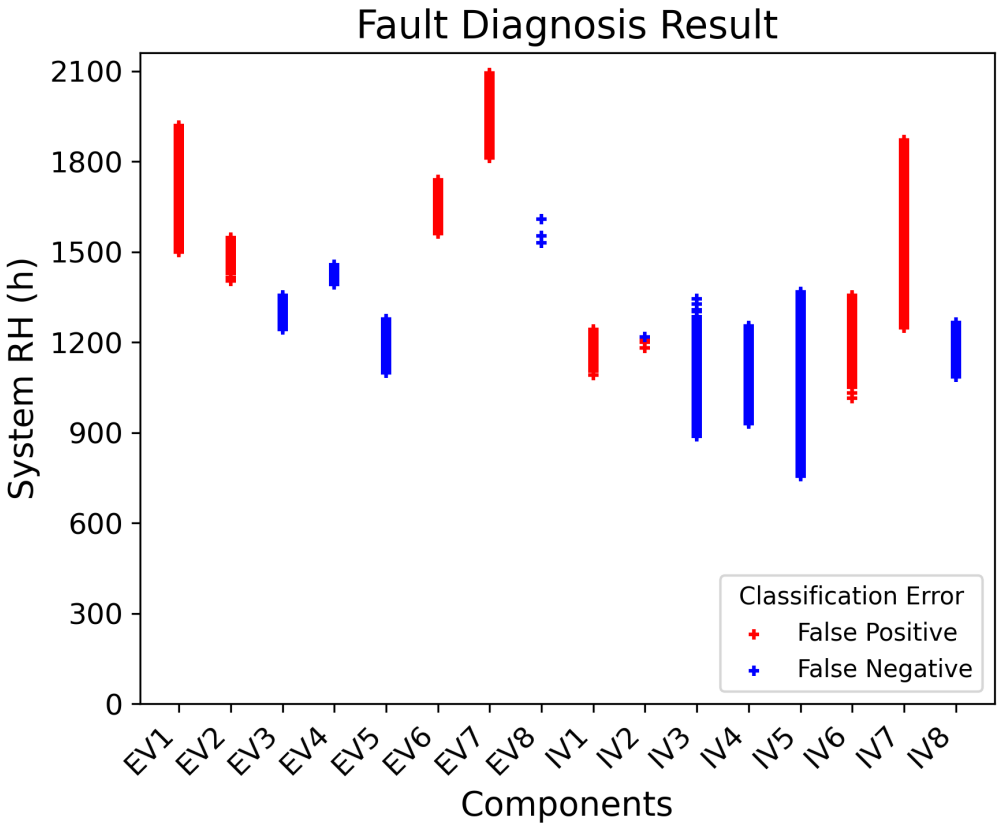


Figure 7.16: Case 3 – Fault diagnosis result visualisation

Figure 7.16 illustrates the classification results of the fault diagnosis sub-model. The red colour denotes the false positive classification error, whereas the blue colour denotes the false negative classification error. The most significant classification error was observed at Component IV7, which diagnosed the anomaly 104 time steps later than actual anomaly.

HI Construction

Figure 7.17 illustrates the results of the HI construction for each component. Despite exhibiting fluctuations in the HI construction, the constructed HI consistently aligned with the actual HI across all components. An abrupt decrease in the HI trend was detected in Component 3; however, a similar decrease was not detected in Component 11.

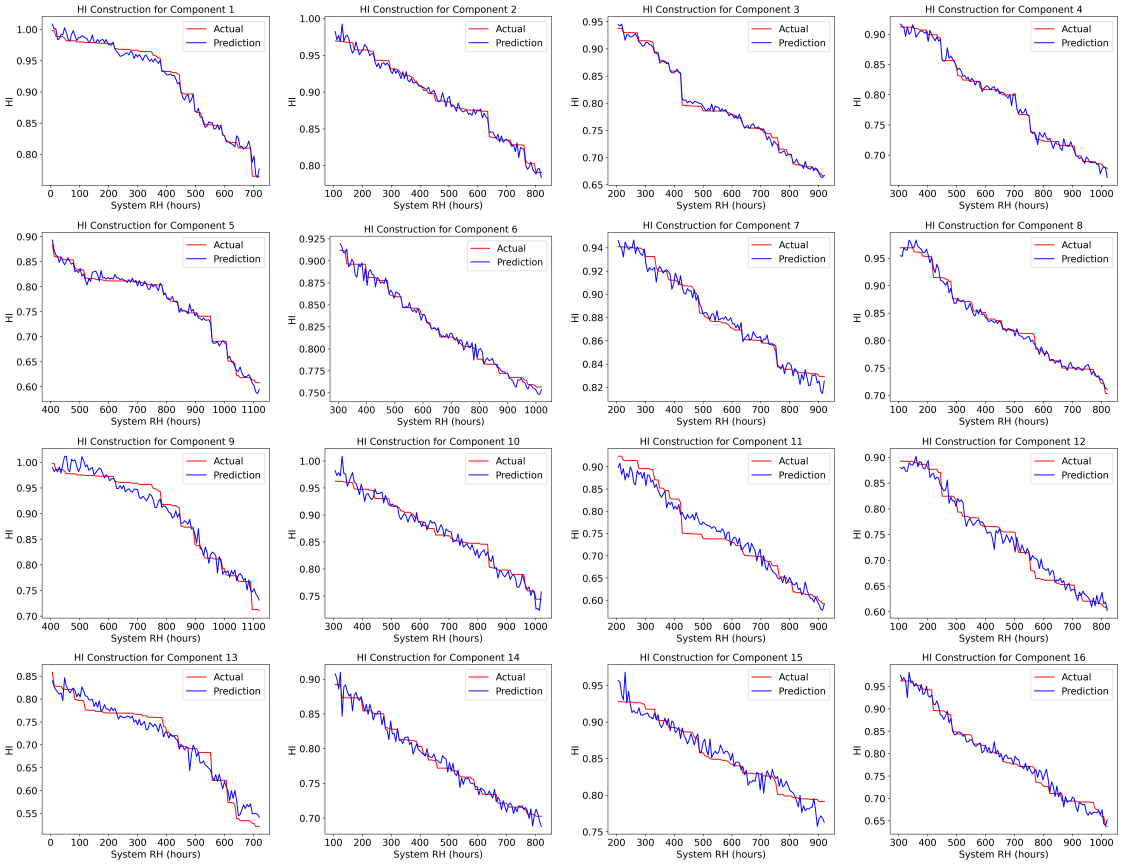
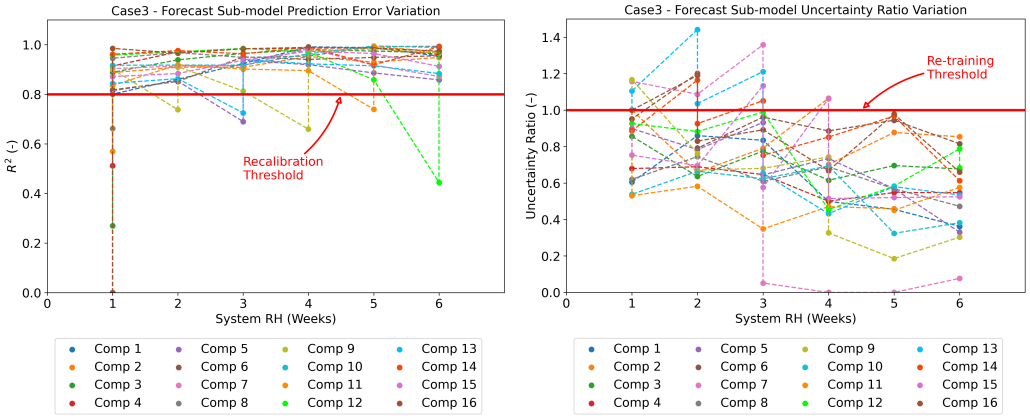


Figure 7.17: Case 3 - HI construction result

HI Forecast

Figure 7.18 presents the prediction error and uncertainty variations of the HI forecast sub-models in Case 3. It is inferred that the performances of the HI forecast sub-models were managed above the pre-defined re-training threshold throughout the 6 weeks of operation. The HI forecast sub-models were re-trained six times by the error criterion and 11 times by the robustness criterion, excluding the first week. In the first week, most of the components, with the exception of Components 2 and 6, required re-training and achieved R^2 values above the R^2 threshold. The lowest R^2 value of 0.44 was observed for Component 12 at Week 6; however, it increased to 0.97 after re-training. In the initial two weeks, the re-training process managed to decrease the uncertainty ratio; however, this reduction was insufficient to meet the established threshold for Components 7 and 13.



(a) Prediction error variations **(b)** Robustness variations

Figure 7.18: Case 3 – HI forecast sub-models performance variations

The HI forecast sub-model results and re-training decision for Component 13 are summarised in Table 7.17, while the HI variation trends for Component 13 are presented in Figure 7.19. The data-driven model management process updated the HI forecast sub-model for Component 13 three times over a six-week operational period. A maintenance decision was made at Week 6 (1,176 hours) in anticipation of the HI of Component 13 crossing the failure threshold at 1,797 hours. At the moment of decision-making, RUL was predicted at 621 hours, which was shorter than the ETA to the next port. The predicted RUL exceeded the actual RUL by 69 hours, representing a significant improvement from the 650-hour discrepancy observed in the projection of the first week. The margin of error, with 95% confidence, was reduced from 12 hours in the first week to 6 hours in Week 6.

Table 7.17: HI forecast sub-model results and management for Component 13

Results	**Unit**	**W1**[a]	**W2**	**W3**	**W4**	**W5**	**W6**
Re-training	–	Yes	Yes	Yes	No	No	No
R^2	–	0.84	0.89	0.94	0.92	0.91	0.88
Warning[b]/Failure[c]RH	hours	1,406	1,240	847	856	1,798	1,797
RUL	hours	1,070	736	175	16	790	621
RUL Error	hours	- 650	- 484	- 91	- 100	- 70	- 69
MoE	hours	12	11	7	5	6	6

[a] Week Number

[b] Actual Warning RH: 756 (hours)

[c] Actual Warning RH: 1,728 (hours)

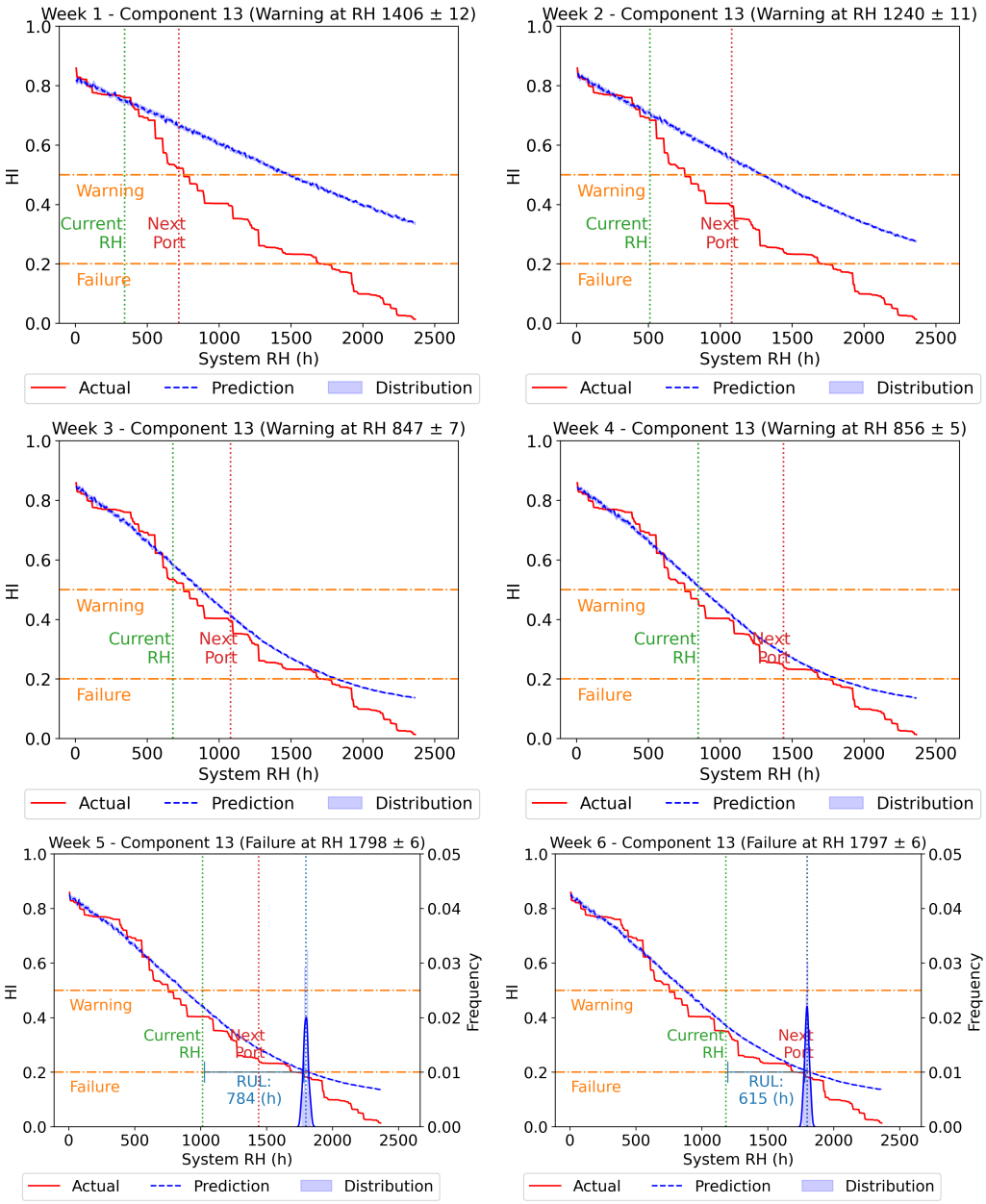


Figure 7.19: Case 3 – HI variation trend for Component 13

7.4 Discussion

The framework for ensuring the trustworthiness of physics-based DT includes the progressivity phase that allows lifetime updates (extension, modification, and addition) for future applications. Furthermore, physics-based DTs in the framework can be substituted with data-driven or hybrid DTs to reduce the modelling efforts and computational costs. In such cases, the trustworthiness criteria need to be updated to include elements such as explainability, robustness against adversarial noise, and physical plausibility. The framework enables a partial trustworthiness assurance

when the DT meets the criteria under limited operating conditions; however, such a DT is capable of generating datasets within the specified operational boundaries.

In this study, the demonstration of the trustworthy DT framework includes a limitation of the validation envelope. The validation of the DT cover only limited operating conditions due to the limited measurement availability. Future studies are expected to use datasets acquired during engine operations to validate the DT under broader operational conditions and to improve the quality of both the DT and generated datasets.

The data-driven model development and management framework provides comprehensive guidance on the method search, selection, integration, training, and management process. Comparison of applicable methods using various degradation profiles reveals that the most effective methods vary according to the particular applications. To develop a trustworthy PHM model, an independent assessment of the applicable methods is recommended, as the effectiveness of a particular combination may not be universal in all scenarios. The comparison results determine the strengths and weaknesses of the considered methods, thus supporting method selection and their integration. In future studies, it is advisable to examine state-of-the-art derivatives of data-driven methods to test the efficiency of the PHM model.

The results in the demonstration of data-driven model management (Case 2-c) suggest that a trained data-driven model tends to preserve the degradation trend obtained from the training dataset. Cylinder 1 in Figure A.6 and Cylinder 7 in Figure A.8 represent relatively late degradation scenarios, and Cylinder 6 in Figure A.6 and Cylinder 4 in Figure A.8 represent relatively fast degradation scenarios. The forecasted HI trends are located below the actual degradation trajectories in the late degradation scenarios, and the forecast HI trends are located above the actual degradation trajectories in the fast degradation scenarios. Although the preservation tendency of the forecast model confers benefits on long-term prediction, it simultaneously introduces a latent risk for prediction failure in real-world operational settings.

Cylinder 7 in Figure A.6, and Cylinder 5 in Figure A.8 have abrupt drops in the degradation trajectories. Such abrupt changes are beyond the predictive capabilities of previously trained forecast models. The data-driven model management

strategy within the framework effectively addresses stochastic changes in degradation trajectories, thereby improving the performance of the data-driven models and their long-term prognostic capability. The iterative performance monitoring and re-training with the latest data adjust the forecasted HI trend, reducing the impact of the stochasticity in degradation trajectories.

The results of the case study underscore the inherent limitations of the data-driven approach. Specifically, even models that achieve perfect error metrics in the training phase may exhibit inadequate predictive performance for future degradation trajectories when these trajectories deviate from historically observed patterns. This finding highlights the necessity of the data operating envelope extension to train various patterns within the model, thereby reducing the risk of discrepancies between historical and future trends.

In the demonstration of the PHM model (Case 3), the uncertainty ratio for particular components did not meet the established threshold even after retraining. The uncertainty of the RUL estimation cannot be mitigated if the training data do not include sufficient information. Future research is recommended to integrate the data envelope extension process into the prognostic procedure to enhance the data information and reduce uncertainty.

In this study, the trustworthiness of the data-driven PHM model was evaluated based on two criteria: error and robustness. However, different applications may require additional criteria, including explainability. It is advisable to research the trustworthiness criteria and incorporating them into the framework, in future studies. Furthermore, comprehensive research on the re-training threshold is expected to effectively maintain the trustworthiness of PHM models within the proposed PHM model management framework.

Chapter 8

Conclusions

8.1 Summary

This study developed a methodology for developing and managing PHM models that overcomes practical challenges in the maritime industry, including limited data availability and stochastic degradation patterns. The methodology contains six phases, including physics-based DT, trustworthiness assurance framework, simulation-based data generation, data-driven PHM model development, PHM model management, and the demonstration of PHM model with management framework. The trustworthiness of the DT was evaluated and ensured using a framework based on the steps of validation, verification, and robustness appraisal, while the trustworthiness of data-driven PHM model was maintained by the iterative monitoring and re-training, based on the accuracy and robustness criteria. In case studies, simulation-generated datasets with diverse degradation profiles and data-driven methods selected through comparative assessments were used to demonstrate the developed frameworks and digital tools, using the 8L50DF marine engine as a reference system.

8.2 Main Findings

The developed frameworks and digital tools are applicable beyond marine engine systems to any machinery system in any industry, as long as performance parameters and the corresponding health information are available. The development of digitalisation platforms across diverse machinery systems is expected to manage the

health condition of the complex machinery system intelligently and effectively.

The DT development contains a degradation model of engine cylinder valves by integrating the degradation law model and the stochastic process model. The degradation modelling process, including the anomaly summary table, DT integration, and trustworthiness assurance, is applicable to any other anomalies. This study considered only a limited number of degradation components; however, real-world machinery operations include simultaneous degradation from multiple components. In future studies, the addition of multiple degradation components and their interactions is anticipated to expand the operating envelope of DT and generate realistic datasets.

This study utilises the simulation-based data generation approach to supplement information to training datasets and to expand the operational capability of data-driven PHM models. As the performance of data-driven models is strongly affected by the quantity and quality of training datasets, this strategy is advantageous not only when the quantity of measurement data is not sufficient but also when the quantity is sufficient, to enhance the capability of data-driven models in handling operating conditions that have not been experienced.

The diagnosis sub-model can analyse current health classes with problematic components and locations; however, the progress of the anomalies cannot be tracked. In contrast, the prognosis sub-model can estimate health indicators of future time slices with RULs; however, it necessitates access to substantial historical time series data and is sensitive to the operating environment. For effective health monitoring and maintenance decision making, the integration of both diagnosis and prognosis sub-models is imperative.

This study uses trend modelling approach, including both HI construction and HI forecast sub-models, to estimate RUL using unlabelled data. Consequently, the forecast errors are attributable to the performances of both HI construction and HI forecast sub-models. The accurate HI construction sub-model is fundamentally requisite for accomplishing accurate HI forecast. The development of the HI construction sub-model necessitates a substantial amount of labelled data for training, thereby making it essential to employ simulation-based data generation using trustworthy DT.

In prognosis methods, especially for the HI forecast sub-task, probabilistic methods

are essential for the uncertainty quantification; however, this process is inherently time consuming. To mitigate computational efforts, it is advisable to employ dropout methods. Note that this method is applicable only to deep learning algorithms, as it relies on randomly dropping connections between layers.

Consideration of real-world operating conditions is important, as the operational environment, accessible data, and data quality in shipboard applications are different compared to virtual and laboratory tests. Shipboard measurements with health labels are generally unobtainable in real-world maritime operations. The results of the case study demonstrated that the simulation-based data generation approach can mitigate the challenges associated with limited data availability.

The proposed framework uses the assumption that the marine engine is part of an autonomous ship with a one-month voyage duration and a one-week data-driven model management interval. However, every ship has a specific operating environment and voyage schedules. It is essential to customise the model management parameters considering the specific operating environment and voyage schedules of each application.

8.3 Novelties & Contributions

The novelty of this research comes from the following four points.

- Trustworthiness of DT is ensured by considering validation, verification, and robustness steps, while also accounting for DT progressivity, rather than relying solely on conventional validation against limited measurement data.
- Integration of degradataion law model and stochastic process model to represent actual degradation patterns of marine engine components.
- Methodology for data-driven PHM model development integrates sub-models of Diagnosis, HI construction, HI forecast with uncertainty quantification, RUL calculation, and maintenance decision making.
- Trustworthiness of data-driven model is evaluated and maintained throughout the system lifetime, considering error and robustness.

This study made contributions toward addressing particular gaps identified in the literature review in Chapter 2, as listed in Table 8.1.

Table 8.1: Addressed literature review gaps

Gap	Description	Addressed?
1	Lack of systematic approach for selecting and integrating data-driven methods	✓
2	Lack of a framework to retain the performance of the PHM model regardless of the varying operating conditions	✓
3	Inefficient approach for developing DT under anomaly conditions	✓
4	Lack of a framework to ensure the trustworthiness of the DT	✓
5	Lack of PHM method tests in complex degradation scenarios, including simultaneous degradation in multiple components and stochastic degradation patterns	△ [a]
6	Lack of method to apply the health information derived from PHM models in the decision making phase	–
7	Lack of methods to identify the need for data generation considering the quality and quantity of data	–
8	Lack of methods to identify the decision sensitivity of fault diagnosis models	–
9	Lack of methods to address sensing errors	–

[a] Partially addressed

1. & 2. A data-driven model development and management framework was proposed and demonstrated. The framework includes data-driven method selection and integration to develop a comprehensive PHM model, along with iterative trustworthiness evaluation and re-training, taking into account error and robustness criteria.

3. & 4. A systematic framework was proposed and demonstrated to develop DTs under anomaly conditions and to ensure the trustworthiness of these DTs. The framework incorporates degradation modelling by integrating the degradation law model and the stochastic process model, as well as ensures the trustworthiness by employing three steps of validation, verification, and robustness.

5. A complex degradation scenario, including simultaneous degradation with different initial health conditions, was tested to demonstrate the effectiveness of the developed frameworks and tools. However, this research gap was only partially addressed by focusing on the degradation of the exhaust and intake valves, which do not fully represent the intricate degradation processes of the marine engine.

8.3.1 Contribution to the Maritime Research Domain and Industry

A framework that evaluates the trustworthiness of DTs and generates simulation-based datasets contributes maritime research domain by mitigating data scarcity and facilitating the research of data-driven approaches. To the author's best knowledge, the trustworthiness of DTs has not been reported in the maritime research domain, consequently the proposed framework provides validity of the outcome of the trustworthy DT. Although several researchers have developed promising data-driven methods, their demonstration is constrained by scarce data. The utilisation of simulation-based data generation from the trustworthy DT framework provides a feasible solution to the limited data availability challenge.

Furthermore, a framework to manage the trustworthiness of data-driven models increases the practical application of the PHM model. In real-world maritime operation, the operation of complex systems, such as marine engines, is generally more intricate than the controlled environmental tests of laboratory experiments and simulations. The maritime industry requires data-driven models that performs in any operating conditions. To the author's best knowledge, this is the initial research considering data availability variations during operations. This study developed a PHM model management framework that incorporates data availability, consequently enhancing the PHM system feasibility in maritime applications.

8.4 Achievement of Aim & Objectives

The objectives 1 and 2 were to identify state-of-the-art methods for the development of PHM models and to analyse the behaviours of the reference marine engine

system under anomaly conditions, with particular attention to autonomous shipping operating profiles. These objectives were achieved through a comprehensive review of the literature on PHM systems, data-driven methods for PHM, and simulation-based data generation under anomaly conditions. Consequently, the identified and analysed information contributed to the other objectives, including the development of DT and data-driven PHM models.

The objectives 3, 4, and 5 focused on the development of trustworthy DT and the generation of simulation datasets under various operating conditions. These objectives employed 1D simulation tools with a physics-based DT grounded in first principles. The DOE with the Latin hypercube method facilitated the DT trustworthiness assessment, as well as the generation of the extended envelope simulation datasets. Consequently, the generated datasets facilitated the development of data-driven models and the model management framework, as well as their demonstrations.

The objectives 6 and 7 focused on the development and management of data-driven PHM models. These objectives took into account the real-world operating conditions of ship machinery systems, including stochastic degradation patterns and multi-components degradation with different initial conditions. MLP and dropout methods were selected and integrated to achieve an accurate RUL estimation while incorporating uncertainty quantification.

The last objective 8 was the demonstration of the previous objectives, aligned with the aim of this research. This objective was achieved by case studies that developed PHM models using only engine shop test records and maintained the effectiveness of the PHM models under dynamic operating conditions.

8.5 Limitations

This study focused on the development of digital tools and their trustworthiness assurance frameworks. The effectiveness of the developed digital tools and frameworks was demonstrated through case studies; however, the following limitations were identified.

- The optimisation of data driven model is only lightly explored, providing opportunities for performance improvement through advanced optimisation

techniques.

- Rule-based methods used in maintenance decision-making do not guarantee optimal maintenance planning.
- Simulation-based data generation with Physics-based DT requires substantial computational effort.
- The developed methodology does not explore both the suitability of acquired data and the required data quantity through simulations.
- Full-scale validation of the developed digital tools using shipboard measurements was not conducted.

8.6 Future Studies

Based on the limitations of this study, the following research topics are expected to advance and implement PHM systems in real-world ship machinery operations.

- The health information derived from the PHM model is expected to facilitate health-aware control. Such intelligent control systems, which jointly aim to maximise operational efficiency and minimise the risk of breakdown, will be essential to replace human operators in autonomous shipping. The development of a health-aware control system is expected to employ optimisation techniques, coupled with extensive analysis of complex system behaviours to mitigate unintended consequences.
- The development of an online PHM methodology with hybrid DT is expected in future studies, aimed at enabling prompt health monitoring and maintenance decision-making. This online PHM necessitates the use of efficient methods with high accuracy and low computational cost, alongside the availability of real-time measurements from the marine engine.
- Future research is expected to develop a holistic PHM methodology that considers data suitability assessment and DT with maintenance modelling. The data suitability assessment incorporates the data quality evaluation and

simulation-based data generation based on the acquired data. The DT is anticipated to simulate not only degradation behaviours but also influence of maintenance to represent long-term ship operations. Additionally, the methodology is anticipated to address the challenge of sensor degradation, which is a frequent event in real-world operations.

Bibliography

[1] United Nations Conference on Trade and Development. Review of Maritime Transport 2023. Technical report, United Nations, New York, 9 2023. URL https://unctad.org/system/files/official-document/rmt2023_en.pdf.

[2] Clarissa A. González Chávez, Selma Brynolf, Mélanie Despeisse, Björn Johansson, Anna Öhrwall Rönnbäck, Jonathan Rösler, and Johan Stahre. Advancing sustainability through digital servitization: An exploratory study in the maritime shipping industry. *Journal of Cleaner Production*, 436:140401, 1 2024. ISSN 0959-6526. doi: 10.1016/J.JCLEPRO.2023.140401.

[3] Pedro Luis Sanchez-Gonzalez, David Díaz-Gutiérrez, Teresa J. Leo, and Luis R. Núñez-Rivas. Toward Digitalization of Maritime Transport? *Sensors 2019, Vol. 19, Page 926*, 19(4):926, 2 2019. ISSN 1424-8220. doi: 10.3390/S19040926. URL https://www.mdpi.com/1424-8220/19/4/926/htmhttps://www.mdpi.com/1424-8220/19/4/926.

[4] Leonard Heilig, Eduardo Lalla-Ruiz, and Stefan Voß. Digital transformation in maritime ports: analysis and a game theoretic framework. *Netnomics*, 18: 227–254, 2017. doi: 10.1007/s11066-017-9122-x. URL https://doi.org/10.1007/s11066-017-9122-x.

[5] Erik Veitch and Ole Andreas Alsos. A systematic review of human-AI interaction in autonomous ship systems. *Safety science*, 152, 2022. doi: 10.1016/j.ssci.2022.105778. URL http://creativecommons.org/licenses/by/4.0/.

[6] P Zhang, L A De Silva, and S Rajagopal. Challenges in the integration of

Data Management Systems (DMS) in ship operations. *Maritime Technology and Research*, 2(4):187–207, 2020.

[7] Marija Jović, Edvard Tijan, David Brčić, and Andreja Pucihar. Digitalization in Maritime Transport and Seaports: Bibliometric, Content and Thematic Analysis. *Journal of Marine Science and Engineering 2022, Vol. 10, Page 486*, 10(4):486, 4 2022. ISSN 2077-1312. doi: 10.3390/JMSE10040486. URL https://www.mdpi.com/2077-1312/10/4/486/htmhttps://www.mdpi.com/2077-1312/10/4/486.

[8] Tobias Rye Torben, Jon Arne Glomsrud, Tom Arne Pedersen, Ingrid B Utne, and Asgeir J Sørensen. Automatic simulation-based testing of autonomous ships using Gaussian processes and temporal logic. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, page 1748006X211069277, 2022. ISSN 1748-006X.

[9] André Listou Ellefsen, Vilmar Æsøy, Sergey Ushakov, and Houxiang Zhang. A comprehensive survey of prognostics and health management based on deep learning for autonomous ships. *IEEE Transactions on Reliability*, 68(2):720–740, 2019. ISSN 0018-9529.

[10] Ikram Remadna, Labib Sadek Terrissa, Soheyb Ayad, and Nourddine Zerhouni. RUL estimation enhancement using hybrid deep learning methods. *International Journal of Prognostics and Health Management*, 12(1), 2021.

[11] Risto Tiusanen, Eetu Heikkilä, and Timo Malm. System Safety Engineering Approach for Autonomous Mobile Machinery. In *World Congress on Engineering Asset Management*, pages 239–251. Springer, 2019.

[12] Ingrid Bouwer Utne, Asgeir J Sørensen, and Ingrid Schjølberg. Risk management of autonomous marine systems and operations. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 57663, page V03BT02A020. American Society of Mechanical Engineers, 2017. ISBN 0791857662.

[13] Marco Molica Colella, Anastasiya Azarko, Ørnulf Jan Rødseth, Håvard Nordahl, Gerasimos Theotokatos, Thoralf Ruud, and Vidar Helgas. Speeding-up the

transition towards a next generation of autonomous ships in the EU. Technical report, AUTOSHIP Project, 2023. URL https://www.autoship-project.eu/wp-content/uploads/2023/03/AUTOSHIP-Ebook.pdf.

[14] James C Dowd and Michael E Fitzgerald. Rethinking Propulsion Machinery Instrumentation for Autonomous Operations. *Naval Engineers Journal*, 133 (1):83–89, 2021. ISSN 0028-1425.

[15] Paul Topping. Autonomous Ships: Where Is It Going from Here? In *Autonomous Vessels in Maritime Affairs: Law and Governance Implications*, pages 397–410. Springer, 2023. doi: https://doi.org/10.1007/978-3-031-24740-8{_}20.

[16] Stig Eriksen, Ingrid Bouwer Utne, and Marie Lützen. An RCM approach for assessing reliability challenges and maintenance needs of unmanned cargo ships. *Reliability Engineering & System Safety*, 210:107550, 2021. ISSN 0951-8320.

[17] Peng Wu and R W G Bucknall. Marine propulsion using battery power. In *Shipping in Changing Climates Conference 2016*. Shipping in Changing Climates Conference 2016, 2016.

[18] C. Kooij, A. A. Kana, and R. G. Hekkenberg. A task-based analysis of the economic viability of low-manned and unmanned cargo ship concepts. *Ocean Engineering*, 242:110111, 12 2021. ISSN 0029-8018. doi: 10.1016/J.OCEANENG.2021.110111.

[19] Kum Fai Yuen, Hui Shan Loh, Qingji Zhou, and Yiik Diew Wong. Determinants of job satisfaction and performance of seafarers. *Transportation research part A: policy and practice*, 110:1–12, 2018. ISSN 0965-8564.

[20] Heather Leggate. The future shortage of seafarers: will it become a reality? *Maritime Policy & Management*, 31(1):3–13, 2004. ISSN 0308-8839.

[21] Jiri de Vos, Robert G. Hekkenberg, and Osiris A. Valdez Banda. The Impact of Autonomous Ships on Safety at Sea – A Statistical Analysis. *Reliability Engineering & System Safety*, 210:107558, 6 2021. ISSN 0951-8320. doi: 10.1016/J.RESS.2021.107558.

[22] Ørnulf Jan Rødseth and Hans-Christoph Burmeister. Developments toward the unmanned ship. In *Proceedings of International Symposium Information on Ships–ISIS*, volume 201, pages 30–31, 2012.

[23] Esa Jokioinen, J Poikonen, R Jalonen, and J Saarni. Remote and autonomous ships-the next steps. *AAWA Position Paper, Rolls Royce plc, London*, 2016.

[24] H A Tvete. The ReVolt-a new, innovative ship concept. *focus-The future is hybrid*, 2015.

[25] F B Scurt, T Vesselenyi, R C Tarca, H Beles, and G Dragomir. Autonomous vehicles: classification, technology and evolution. *IOP Conference Series: Materials Science and Engineering*, 1169(1), 2021. ISSN 1757-8981. doi: 10.1088/1757-899x/1169/1/012032.

[26] EMSA. Annual overview of marine casualties and incidents. Technical report, European Maritime Safety Agency, Lisbon, Portugal, 6 2023. URL https://www.emsa.europa.eu/publications/item/5052-annual-overview-of-marine-casualties-and-incidents.html.

[27] I. Yanchin and O. Petrov. Towards autonomous shipping: Benefits and challenges in the field of information technology and telecommunication. *TransNav*, 14(3), 2020. ISSN 20836481. doi: 10.12716/1001.14.03.12.

[28] Rolf Skjong. Risk Acceptance Criteria: current proposals and IMO position. *Surface transport technologies for sustaninable development, Spain*, pages 4–6, 2002.

[29] Marko Höyhtyä, Jyrki Huusko, Markku Kiviranta, Kenneth Solberg, and Juha Rokka. Connectivity for autonomous ships: Architecture, use cases, and research challenges. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 345–350. IEEE, 2017. ISBN 1509040323.

[30] Guangrong Zou and Saara H{\"a}nninen. Integrated Energy Solutions to Smart and Green Shipping: 2021 Edition. Technical report, VTT Technical Research Centre of Finland, 2021.

[31] Joao L.D. Dantas and Gerasimos Theotokatos. A framework for the economic-environmental feasibility assessment of short-sea shipping autonomous vessels. *Ocean Engineering*, 279:114420, 7 2023. ISSN 0029-8018. doi: 10.1016/J.OCEANENG.2023.114420.

[32] Çağlar Karatuğ, Yasin Arslanoğlu, and C. Guedes Soares. Review of maintenance strategies for ship machinery systems. *Journal of Marine Engineering and Technology*, 22(5), 2023. ISSN 20568487. doi: 10.1080/20464177.2023.2180831.

[33] Muntazir Abbas and Mahmood Shafiee. An overview of maintenance management strategies for corroded steel structures in extreme marine environments. *Marine Structures*, 71:102718, 5 2020. ISSN 0951-8339. doi: 10.1016/J.MARSTRUC.2020.102718.

[34] Mohammad Yazdi. Maintenance Strategies and Optimization Techniques. In *Advances in Computational Mathematics for Industrial System Reliability and Maintainability*, chapter 3, pages 43–58. Springer, Cham, 2024. ISBN 978-3-031-53514-7. doi: 10.1007/978-3-031-53514-7{_}3. URL https://link.springer.com/chapter/10.1007/978-3-031-53514-7_3.

[35] John Moubray. *Reliability Centered Maintenance II*. Industrial Press Inc., 1997.

[36] Yanrong Li, Shizhe Peng, Yanting Li, and Wei Jiang. A review of condition-based maintenance: Its prognostic and operational aspects. *Frontiers of Engineering Management 2020 7:3*, 7(3):323–334, 7 2020. ISSN 2096-0255. doi: 10.1007/S42524-020-0121-5. URL https://link.springer.com/article/10.1007/s42524-020-0121-5.

[37] Elena Quatrini, Francesco Costantino, Giulio Di Gravio, and Riccardo Patriarca. Condition-Based Maintenance—An Extensive Literature Review. *Machines 2020, Vol. 8, Page 31*, 8(2):31, 6 2020. ISSN 2075-1702. doi: 10.3390/MACHINES8020031. URL https://www.mdpi.com/2075-1702/8/2/31/htmhttps://www.mdpi.com/2075-1702/8/2/31.

[38] Amitkumar Patil, Gunjan Soni, Anuj Prakash, and Kritika Karwasra. Maintenance strategy selection: a comprehensive review of current paradigms and solu-

tion approaches. *International Journal of Quality and Reliability Management*, 39(3):675–703, 2 2022. ISSN 0265671X. doi: 10.1108/IJQRM-04-2021-0105/FULL/PDF.

[39] Tanja Nemeth, Fazel Ansari, Wilfried Sihn, Bernhard Haslhofer, and Alexander Schindler. PriMa-X: A reference model for realizing prescriptive maintenance and assessing its maturity enhanced by machine learning. *Procedia CIRP*, 72: 1039–1044, 1 2018. ISSN 2212-8271. doi: 10.1016/J.PROCIR.2018.03.280.

[40] Peng Zhang, Zeyu Gao, Lele Cao, Fangyang Dong, Yongjiu Zou, Kai Wang, Yuewen Zhang, and Peiting Sun. Marine Systems and Equipment Prognostics and Health Management: A Systematic Review from Health Condition Monitoring to Maintenance Strategy. *Machines*, 10(2):72, 2022.

[41] Khanh T P Nguyen and Kamal Medjaher. A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering & System Safety*, 188:251–262, 2019.

[42] Pengfei Wen, Shuai Zhao, Shaowei Chen, and Yong Li. A generalized remaining useful life prediction method for complex systems based on composite health indicator. *Reliability Engineering & System Safety*, 205:107241, 2021.

[43] Ahmad BahooToroody, Mohammad Mahdi Abaei, Osiris Valdez Banda, Pentti Kujala, Filippo De Carlo, and Rouzbeh Abbassi. Prognostic health management of repairable ship systems through different autonomy degree; From current condition to fully autonomous ship. *Reliability Engineering and System Safety*, 221, 2022. ISSN 09518320. doi: 10.1016/j.ress.2022.108355.

[44] 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. ISBN 1424419352.

[45] Oihane C Basurko and Zigor Uriondo. Condition-based maintenance for medium speed diesel engines used in vessels in operation. *Applied Thermal Engineering*, 80:404–412, 2015. ISSN 1359-4311.

[46] Erik Vanem and Andreas Brandsæter. Unsupervised anomaly detection based on clustering methods and sensor data on a marine diesel engine. *Journal of Marine Engineering & Technology*, 20(4):217–234, 2021. ISSN 2046-4177.

[47] Andrea Coraddu, Luca Oneto, Davide Ilardi, Sokratis Stoumpos, and Gerasimos Theotokatos. Marine dual fuel engines monitoring in the wild through weakly supervised data analytics. *Engineering Applications of Artificial Intelligence*, 100:104179, 2021. ISSN 0952-1976.

[48] Zhixiong Li, Xinping Yan, Chengqing Yuan, Jiangbin Zhao, and Zhongxiao Peng. Fault detection and diagnosis of a gearbox in marine propulsion systems using bispectrum analysis and artificial neural networks. *Journal of Marine Science and Application*, 10(1):17–24, 2011. ISSN 1993-5048.

[49] Ameneh Forouzandeh Shahraki, Om Parkash Yadav, and Haitao Liao. A review on degradation modelling and its engineering applications. *International Journal of Performability Engineering*, 13(3):299, 2017. ISSN 0973-1318.

[50] Brendan P. Sullivan, Shantanoo Desai, Jordi Sole, Monica Rossi, Lucia Ramundo, and Sergio Terzi. Maritime 4.0 – Opportunities in Digitalization and Advanced Manufacturing for Vessel Development. *Procedia Manufacturing*, 42: 246–253, 1 2020. ISSN 2351-9789. doi: 10.1016/J.PROMFG.2020.02.078.

[51] Damoon Razmjooei, Moslem Alimohammadlou, Habib Allah Ranaei Kordshouli, and Kazem Askarifar. Industry 4.0 research in the maritime industry: a bibliometric analysis. *WMU Journal of Maritime Affairs*, 22(3):385–416, 9 2023. ISSN 16541642. doi: 10.1007/S13437-022-00298-8/FIGURES/6. URL https://link.springer.com/article/10.1007/s13437-022-00298-8.

[52] Kevin Heffner and Ernulf Jan Rødseth. Enabling Technologies for Maritime Autonomous Surface Ships. *Journal of Physics: Conference Series*, 1357(1): 012021, 10 2019. ISSN 1742-6596. doi: 10.1088/1742-6596/1357/1/012021. URL https://iopscience.iop.org/article/10.1088/1742-6596/1357/1/012021https://iopscience.iop.org/article/10.1088/1742-6596/1357/1/012021/meta.

[53] Yuki Ichimura, Dimitrios Dalaklis, Momoko Kitada, and Anastasia Christodoulou. Shipping in the era of digitalization: Mapping the future strategic plans of major maritime commercial actors. *Digital Business*, 2(1): 100022, 3 2022. ISSN 2666-9544. doi: 10.1016/J.DIGBUS.2022.100022.

[54] Edvard Tijan, Marija Jović, Saša Aksentijević, and Andreja Pucihar. Digital transformation in the maritime transport sector. *Technological Forecasting and Social Change*, 170, 2021. ISSN 00401625. doi: 10.1016/j.techfore.2021.120879.

[55] Andrzej Felski and Karolina Zwolak. The Ocean-Going Autonomous Ship—Challenges and Threats. *Journal of Marine Science and Engineering 2020, Vol. 8, Page 41*, 8(1):41, 1 2020. ISSN 2077-1312. doi: 10.3390/JMSE8010041. URL https://www.mdpi.com/2077-1312/8/1/41/htmhttps://www.mdpi.com/2077-1312/8/1/41.

[56] Tae-eun Kim and Jens-Uwe Schröder-Hinrichs. Research Developments and Debates Regarding Maritime Autonomous Surface Ship: Status, Challenges and Perspectives. In *New Maritime Business: Uncertainty, Sustainability, Technology and Big Data*, pages 175–197. Springer, Cham, 11 2021. ISBN 978-3-030-78957-2. doi: 10.1007/978-3-030-78957-2{_}10. URL https://link.springer.com/chapter/10.1007/978-3-030-78957-2_10.

[57] Goran Vojković and Melita Milenković. Autonomous ships and legal authorities of the ship master. *Case Studies on Transport Policy*, 8(2):333–340, 6 2020. ISSN 2213-624X. doi: 10.1016/J.CSTP.2019.12.001.

[58] Hasan Mahbub Tusher, Ziaul Haque Munim, Theo E. Notteboom, Tae Eun Kim, and Salman Nazir. Cyber security risk assessment in autonomous shipping. *Maritime Economics and Logistics*, 24(2):208–227, 6 2022. ISSN 1479294X. doi: 10.1057/S41278-022-00214-0/FIGURES/3. URL https://link.springer.com/article/10.1057/s41278-022-00214-0.

[59] Maria Lambrou, Daisuke Watanabe, and Junya Iida. Shipping digitalization management: conceptualization, typology and antecedents. *Journal of Shipping and Trade 2019 4:1*, 4(1):1–17, 11 2019. ISSN 2364-4575. doi: 10.1186/S41072-019-0052-7. URL https:

//link.springer.com/articles/10.1186/s41072-019-0052-7https://link.springer.com/article/10.1186/s41072-019-0052-7.

[60] Xiaojian Xu, Xinping Yan, Kun Yang, Jiangbin Zhao, Chenxing Sheng, and Chengqing Yuan. Review of condition monitoring and fault diagnosis for marine power systems. *Transportation Safety and Environment*, 2021.

[61] Yogeswaran Sinnasamy, Mohd Razali Mat Yassin, Noor Aishah Sa'at, Hasril Nain, Faiz Azmi Sutarji, Azmahani Sulaiman, Ibrahim Tahir, Rosdi Yaakob, Ahmad Subardi Mohd Wazir, Kamil Azwan Salehuddin, Mohd Ridzuan Mohd Rashid, Abbas Zubir, Hanizah Kasmoni, Elizabeth Louisnaden, and Khairul Anuar Ahmad. Recognition of most common diesel engine condition monitoring methods. *Defence S and T Technical Bulletin*, 10(3), 2017. ISSN 19856571.

[62] Yanghui Tan, Hui Tian, Ruizheng Jiang, Yejin Lin, and Jundong Zhang. A comparative investigation of data-driven approaches based on one-class classifiers for condition monitoring of marine machinery system. *Ocean Engineering*, 201:107174, 2020. ISSN 0029-8018.

[63] Stylianos Oikonomou, Iraklis Lazakis, and George Papadakis. An innovative machine learning system for real time condition monitoring of ship machinery. In *Practical Design of Ships and Other Floating Structures*, pages 753–768. Springer, 2019.

[64] Chao Fu, Xiaoxia Liang, Qian Li, Kuan Lu, Fengshou Gu, Andrew D Ball, and Zhaoli Zheng. Comparative Study on Health Monitoring of a Marine Engine Using Multivariate Physics-Based Models and Unsupervised Data-Driven Models. *Machines*, 11(5):557, 2023.

[65] Christian Velasco-Gallego and Iraklis Lazakis. Development of a time series imaging approach for fault classification of marine systems. *Ocean Engineering*, pages 1–32, 2022. ISSN 0029-8018.

[66] Yulong Zhan, Zhubin Shi, and Mlngming Liu. The application of support vector machines (SVM) to fault diagnosis of marine main engine cylinder cover. In *IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society*, pages 3018–3022. IEEE, 2007. ISBN 1424407834.

[67] Zhiming Cheng, Houlin Liu, Runan Hua, Liang Dong, Qijiang Ma, and Jiancheng Zhu. Research on Multi-Fault Identification of Marine Vertical Centrifugal Pump Based on Multi-Domain Characteristic Parameters. *Journal of Marine Science and Engineering*, 11(3):551, 2023.

[68] Soroush Senemmar and Jie Zhang. Deep learning-based fault detection, classification, and locating in shipboard power systems. In *2021 IEEE Electric Ship Technologies Symposium (ESTS)*, pages 1–6, 2021.

[69] Peihua Han, André Listou Ellefsen, Guoyuan Li, Vilmar Æsøy, and Houxiang Zhang. Fault prognostics using LSTM networks: Application to marine diesel engine. *IEEE Sensors Journal*, 21(22), 2021. ISSN 15581748. doi: 10.1109/JSEN.2021.3119151.

[70] Shilong Yang, Baoping Tang, Weiying Wang, Qichao Yang, and Cheng Hu. Physics-informed multi-state temporal frequency network for RUL prediction of rolling bearings. *Reliability Engineering & System Safety*, 242:109716, 2 2024. ISSN 0951-8320. doi: 10.1016/J.RESS.2023.109716.

[71] Ningbo Zhao, Shuying Li, Yunpeng Cao, and Hui Meng. Remote intelligent expert system for operation state of marine gas turbine engine. In *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pages 3210–3215, 2014.

[72] Tiedo Tinga and Richard Loendersloot. Physical model-based prognostics and health monitoring to enable predictive maintenance. *Predictive Maintenance in Dynamic Systems: Advanced Methods, Decision Support Tools and Real-World Applications*, pages 313–353, 2019.

[73] Zhonghai Ma, Haitao Liao, Jianhang Gao, Songlin Nie, and Yugang Geng. Physics-Informed Machine Learning for Degradation Modeling of an Electro-Hydrostatic Actuator System. *Reliability Engineering and System Safety*, 229, 2023. ISSN 09518320. doi: 10.1016/j.ress.2022.108898.

[74] Amgad Muneer, Shakirah Mohd Taib, Suliman Mohamed Fati, and Hitham Alhussian. Deep-Learning Based Prognosis Approach for Remaining Useful Life Prediction of Turbofan Engine. *Symmetry*, 13(10):1861, 2021.

[75] Omar Bougacha, Christophe Varnier, and Noureddine Zerhouni. A review of post-prognostics decision-making in prognostics and health management. *International Journal of Prognostics and Health Management*, 11(15):31, 2020.

[76] Pascal Vrignat, Frédéric Kratz, and Manuel Avila. Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. *Reliability Engineering & System Safety*, 218:108140, 2022.

[77] Kavindu Ranasinghe, Roberto Sabatini, Alessandro Gardi, Suraj Bijjahalli, Rohan Kapoor, Thomas Fahey, and Kathiravan Thangavel. Advances in Integrated System Health Management for mission-essential and safety-critical aerospace applications. *Progress in Aerospace Sciences*, 128:100758, 2022.

[78] Ali Saleh, Manuel Chiachío, Juan Fernández Salas, and Athanasios Kolios. Self-adaptive optimized maintenance of offshore wind turbines by intelligent Petri nets. *Reliability Engineering and System Safety*, 231, 2023. ISSN 09518320. doi: 10.1016/j.ress.2022.109013.

[79] M Hung Do and Dirk Söffker. State-of-the-art in integrated prognostics and health management control for utility-scale wind turbines. *Renewable and Sustainable Energy Reviews*, 145:111102, 2021.

[80] Ramin Moradi and Katrina M Groth. Modernizing risk assessment: A systematic integration of PRA and PHM techniques. *Reliability Engineering & System Safety*, 204:107194, 2020.

[81] André Listou Ellefsen, Emil Bjørlykhaug, Vilmar Æsøy, Sergey Ushakov, and Houxiang Zhang. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliability Engineering & System Safety*, 183:240–251, 3 2019. ISSN 0951-8320. doi: 10.1016/J.RESS.2018.11.027.

[82] Christian Velasco-Gallego and Iraklis Lazakis. Mar-RUL: A remaining useful life prediction approach for fault prognostics of marine machinery. *Applied Ocean Research*, 140:103735, 11 2023. ISSN 0141-1187. doi: 10.1016/J.APOR.2023.103735.

[83] Kiran Maharana, Surajit Mondal, and Bhushankumar Nemade. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1):91–99, 2022.

[84] Jose Ignacio Aizpurua, Knut Erik Knutsen, Markus Heimdal, and Erik Vanem. Integrated machine learning and probabilistic degradation approach for vessel electric motor prognostics. *Ocean Engineering*, 275:114153, 2023.

[85] Seokgoo Kim, Nam Ho Kim, and Joo Ho Choi. Prediction of remaining useful life by data augmentation technique based on dynamic time warping. *Mechanical Systems and Signal Processing*, 136:106486, 2 2020. ISSN 0888-3270. doi: 10.1016/J.YMSSP.2019.106486.

[86] 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*, 178:255–268, 10 2018. ISSN 0951-8320. doi: 10.1016/J.RESS.2018.06.021.

[87] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 4. Springer, 2017. doi: https://doi.org/10.1007/978-3-319-52452-8.

[88] 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. ISSN 00189529. doi: 10.1109/TR.2014.2299152.

[89] Zhixiong Li, Xinping Yan, Chengqing Yuan, and Zhongxiao Peng. Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed. *Journal of Mechanical Science and Technology*, 26(8):2413–2423, 8 2012. ISSN 1738494X. doi: 10.1007/S12206-012-0621-2/METRICS. URL https://link.springer.com/article/10.1007/s12206-012-0621-2.

[90] Yiannis Raptodimos and Iraklis Lazakis. Using artificial neural network-self-organising map for data clustering of marine engine condition monitoring applications. *Ships and Offshore Structures*, 13(6):649–656, 2018. ISSN 1744-5302.

[91] Pan Zhang, Wenzhi Gao, Yong Li, and Yanjun Wang. Misfire detection of diesel engine based on convolutional neural networks. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 235 (8):2148–2165, 7 2021. ISSN 20412991. doi: 10.1177/0954407020987077/ASSET/IMAGES/LARGE/10.1177{_}0954407020987077-FIG13.JPEG. URL https://journals.sagepub.com/doi/full/10.1177/0954407020987077.

[92] Peng Zhang, Chaozhe Li, Huanyun Xu, Yongjiu Zou, Kai Wang, Yuewen Zhang, and Peiting Sun. Bi-AAE: A binary adversarial autoencoder deep neural network model for anomaly detection in system-levels marine diesel engines. *Ocean Engineering*, 302:117700, 6 2024. ISSN 0029-8018. doi: 10.1016/J.OCEANENG.2024.117700.

[93] Peter A Lindahl, Daisy H Green, Gregory Bredariol, Andre Aboulian, John S Donnal, and Steven B Leeb. Shipboard fault detection through nonintrusive load monitoring: A case study. *IEEE Sensors Journal*, 18(21):8986–8995, 2018. ISSN 1530-437X.

[94] Stephen Adams, Peter A. Beling, Kevin Farinholt, Nathan Brown, Sherwood Polter, and Qing Dong. Condition based monitoring for a hydraulic actuator. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*, volume 2016-October, 2016. doi: 10.36001/phmconf.2016.v8i1.2581.

[95] Ruihan Wang, Hui Chen, Cong Guan, Wenfeng Gong, and Zehui Zhang. Research on the fault monitoring method of marine diesel engines based on the manifold learning and isolation forest. *Applied Ocean Research*, 112:102681, 7 2021. ISSN 0141-1187. doi: 10.1016/J.APOR.2021.102681.

[96] Yanghui Tan, Jundong Zhang, Hui Tian, Dingyu Jiang, Lei Guo, Gaoming Wang, and Yejin Lin. Multi-label classification for simultaneous fault diagnosis of marine machinery: a comparative study. *Ocean Engineering*, 239:109723, 2021. ISSN 0029-8018.

[97] Wennian Yu, I I Yong Kim, and Chris Mechefske. An improved similarity-based

prognostic algorithm for RUL estimation using an RNN autoencoder scheme. *Reliability Engineering & System Safety*, 199:106926, 2020.

[98] Ahin Banerjee, Sanjay K Gupta, and Chandrasekhar Putcha. Degradation Data–Driven Analysis for Estimation of the Remaining Useful Life of a Motor. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 7(2):4021012, 2021. ISSN 2376-7642.

[99] M. A. Djeziri, S. Benmoussa, and M. EH Benbouzid. Data-driven approach augmented in simulation for robust fault prognosis. *Engineering Applications of Artificial Intelligence*, 86:154–164, 11 2019. ISSN 0952-1976. doi: 10.1016/J.ENGAPPAI.2019.09.002.

[100] Racha Khelif, Brigitte Chebel-Morello, Simon Malinowski, Emna Laajili, Farhat Fnaiech, and Noureddine Zerhouni. Direct Remaining Useful Life Estimation Based on Support Vector Regression. *IEEE Transactions on Industrial Electronics*, 64(3):2276–2285, 3 2017. ISSN 02780046. doi: 10.1109/TIE.2016.2623260.

[101] Wennian Yu, II Yong Kim, and Chris Mechefske. Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mechanical Systems and Signal Processing*, 129, 2019. ISSN 10961216. doi: 10.1016/j.ymssp.2019.05.005.

[102] Haoxuan Zhou, Xin Huang, Guangrui Wen, Zihao Lei, Shuzhi Dong, Ping Zhang, and Xuefeng Chen. Construction of health indicators for condition monitoring of rotating machinery: A review of the research. *Expert Systems with Applications*, 203:117297, 10 2022. ISSN 0957-4174. doi: 10.1016/J.ESWA.2022.117297.

[103] Mohand Arab Djeziri, Samir Benmoussa, and Enrico Zio. Review on health indices extraction and trend modeling for remaining useful life estimation. *Artificial Intelligence Techniques for a Scalable Energy Transition: Advanced Methods, Digital Technologies, Decision Support Tools, and Applications*, pages 183–223, 2020.

[104] Khanh T.P. Nguyen and Kamal Medjaher. An automated health indicator construction methodology for prognostics based on multi-criteria optimization.

ISA Transactions, 113:81–96, 7 2021. ISSN 0019-0578. doi: 10.1016/J.ISATRA.2020.03.017.

[105] Georgios Koutroulis, Belgin Mutlu, and Roman Kern. Constructing robust health indicators from complex engineered systems via anticausal learning. *Engineering Applications of Artificial Intelligence*, 113:104926, 8 2022. ISSN 0952-1976. doi: 10.1016/J.ENGAPPAI.2022.104926.

[106] Lin Huang, Xin Pan, Yajie Liu, and Li Gong. An Unsupervised Machine Learning Approach for Monitoring Data Fusion and Health Indicator Construction. *Sensors 2023, Vol. 23, Page 7239*, 23(16):7239, 8 2023. ISSN 1424-8220. doi: 10.3390/S23167239. URL https://www.mdpi.com/1424-8220/23/16/7239/htmhttps://www.mdpi.com/1424-8220/23/16/7239.

[107] 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. ISSN 0888-3270.

[108] Wei Wu, Jingtao Hu, and Jilong Zhang. Prognostics of machine health condition using an improved ARIMA-based prediction method. *ICIEA 2007: 2007 Second IEEE Conference on Industrial Electronics and Applications*, pages 1062–1067, 2007. doi: 10.1109/ICIEA.2007.4318571.

[109] Márcio Das Chagas Moura, Enrico Zio, Isis Didier Lins, and Enrique Droguett. Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, 96(11):1527–1534, 11 2011. ISSN 0951-8320. doi: 10.1016/J.RESS.2011.06.006.

[110] Yongzhi Zhang, Rui Xiong, Hongwen He, and Michael G. Pecht. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Transactions on Vehicular Technology*, 67(7): 5695–5705, 7 2018. ISSN 00189545. doi: 10.1109/TVT.2018.2805189.

[111] Lei Ren, Yaqiang Sun, Hao Wang, and Lin Zhang. Prediction of bearing remaining useful life with deep convolution neural network. *IEEE Access*, 6: 13041–13049, 2 2018. ISSN 21693536. doi: 10.1109/ACCESS.2018.2804930.

[112] Xiang Li, Qian Ding, and Jian Qiao Sun. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 172:1–11, 4 2018. ISSN 0951-8320. doi: 10.1016/J.RESS.2017.11.021.

[113] Bo Sun, Yu Li, Zili Wang, Yi Ren, Qiang Feng, and Dezhen Yang. An improved inverse Gaussian process with random effects and measurement errors for RUL prediction of hydraulic piston pump. *Measurement*, 173:108604, 3 2021. ISSN 0263-2241. doi: 10.1016/J.MEASUREMENT.2020.108604.

[114] Venkat P. Nemani, Hao Lu, Adam Thelen, Chao Hu, and Andrew T. Zimmerman. Ensembles of probabilistic LSTM predictors and correctors for bearing prognostics using industrial standards. *Neurocomputing*, 491:575–596, 6 2022. ISSN 0925-2312. doi: 10.1016/J.NEUCOM.2021.12.035.

[115] Andy Rivas, Gregory Kyriakos Delipei, and Jason Hou. Predictions of component Remaining Useful Lifetime Using Bayesian Neural Network. *Progress in Nuclear Energy*, 146:104143, 4 2022. ISSN 0149-1970. doi: 10.1016/J.PNUCENE.2022.104143.

[116] Meng Wei, Hairong Gu, Min Ye, Qiao Wang, Xinxin Xu, and Chenguang Wu. Remaining useful life prediction of lithium-ion batteries based on Monte Carlo Dropout and gated recurrent unit. *Energy Reports*, 7:2862–2871, 11 2021. ISSN 2352-4847. doi: 10.1016/J.EGYR.2021.05.019.

[117] Miyeon Jeon, Yoojeong Noh, Kyunghwan Jeon, Sangbong Lee, and Inwon Lee. Data gap analysis of ship and maritime data using meta learning. *Applied Soft Computing*, 101, 2021. ISSN 15684946. doi: 10.1016/j.asoc.2020.107048.

[118] Boris Kantsepolsky, Itzhak Aviv, Roye Weitzfeld, and Eliyahu Bordo. Exploring Quantum Sensing Potential for Systems Applications. *IEEE Access*, 11:31569–31582, 2023. ISSN 21693536. doi: 10.1109/ACCESS.2023.3262506.

[119] Antonios Kamariotis, Konstantinos Tatsis, Eleni Chatzi, Kai Goebel, and Daniel Straub. A metric for assessing and optimizing data-driven prognostic algorithms for predictive maintenance. *Reliability Engineering & System Safety*, 242:109723, 2 2024. ISSN 0951-8320. doi: 10.1016/J.RESS.2023.109723.

[120] J Jeon and G Theotokatos. Datasets envelope impact on marine engines prognostics and health management models accuracy. In *33rd European Safety and Reliability Conference*, pages 2901–2906, 2023. ISBN 9789811880711. doi: 10.3850/978-981-18-8071-1{_}P592-cd.

[121] Michail Cheliotis, Iraklis Lazakis, and Gerasimos Theotokatos. Machine learning and data-driven fault detection for ship systems operations. *Ocean Engineering*, 216:107968, 2020. ISSN 0029-8018.

[122] Peihua Han, Andre Listou Ellefsen, Guoyuan Li, Finn Tore Holmeset, and Houxiang Zhang. Fault Detection with LSTM-Based Variational Autoencoder for Maritime Components. *IEEE Sensors Journal*, 21(19):21903–21912, 10 2021. ISSN 15581748. doi: 10.1109/JSEN.2021.3105226.

[123] Ruihan Wang, Hui Chen, and Cong Guan. A Bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine. *ISA transactions*, 118:159–173, 2021.

[124] Marco Altosole, Flavio Balsamo, Maria Acanfora, Luigia Mocerino, Ugo Campora, and Francesco Perra. A Digital Twin Approach to the Diagnostic Analysis of a Marine Diesel Engine. In *Progress in Marine Science and Technology*, volume 6, pages 198–206, 2022. doi: 10.3233/PMST220025.

[125] Carlos Ferreira and Gil Gonçalves. Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods. *Journal of Manufacturing Systems*, 63:550–562, 4 2022. ISSN 0278-6125. doi: 10.1016/J.JMSY.2022.05.010.

[126] Yu Guo, Jundong Zhang, Bin Sun, and Yongkang Wang. A universal fault diagnosis framework for marine machinery based on domain adaptation. *Ocean Engineering*, 302:117729, 6 2024. ISSN 0029-8018. doi: 10.1016/J.OCEANENG.2024.117729.

[127] Thomas W. Lucas, W. David Kelton, Paul J. Sánchez, Susan M. Sanchez, and Ben L. Anderson. Changing the paradigm: Simulation, now a method of first resort. *Naval Research Logistics (NRL)*, 62(4):

293–303, 6 2015. ISSN 1520-6750. doi: 10.1002/NAV.21628. URL https://onlinelibrary.wiley.com/doi/full/10.1002/nav.21628https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.21628https://onlinelibrary.wiley.com/doi/10.1002/nav.21628.

[128] Gary Horne and Klaus Peter Schwierz. Summary of Data Farming. *Axioms 2016, Vol. 5, Page 8*, 5(1):8, 3 2016. ISSN 2075-1680. doi: 10.3390/AXIOMS5010008. URL https://www.mdpi.com/2075-1680/5/1/8/htmhttps://www.mdpi.com/2075-1680/5/1/8.

[129] Tobias Lechler, Martin Sjarov, and Jörg Franke. Data Farming in Production Systems - A Review on Potentials, Challenges and Exemplary Applications. *Procedia CIRP*, 96:230–235, 1 2021. ISSN 2212-8271. doi: 10.1016/J.PROCIR.2021.01.156.

[130] Susan M. Sanchez. Data Farming. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 30(4), 11 2020. ISSN 15581195. doi: 10.1145/3425398. URL https://dl.acm.org/doi/10.1145/3425398.

[131] Kamyar Nikzadfar and Amir H. Shamekhi. Investigating the relative contribution of operational parameters on performance and emissions of a common-rail diesel engine using neural network. *Fuel*, 125:116–128, 6 2014. ISSN 0016-2361. doi: 10.1016/J.FUEL.2014.02.021.

[132] Yaqing Bo, Han Wu, Weifan Che, Zeyu Zhang, Xiangrong Li, and Leonid Myagkov. Methodology and application of digital twin-driven diesel engine fault diagnosis and virtual fault model acquisition. *Engineering Applications of Artificial Intelligence*, 131:107853, 5 2024. ISSN 0952-1976. doi: 10.1016/J.ENGAPPAI.2024.107853.

[133] Eric VanDerHorn, Zhenghua Wang, and Sankaran Mahadevan. Towards a digital twin approach for vessel-specific fatigue damage monitoring and prognosis. *Reliability Engineering & System Safety*, 219:108222, 2022. ISSN 0951-8320.

[134] Jiju Antony. *Design of experiments for engineers and scientists*. Elsevier, 2014. ISBN 0080994199.

[135] Marco Altosole, Silvia Donnarumma, Valentina Spagnolo, and Stefano Vignolo. Performance Simulation of Marine Cycloidal Propellers: A Both Theoretical and Heuristic Approach. *Journal of Marine Science and Engineering 2022, Vol. 10, Page 505*, 10(4):505, 4 2022. ISSN 2077-1312. doi: 10.3390/JMSE10040505. URL https://www.mdpi.com/2077-1312/10/4/505/htmhttps://www.mdpi.com/2077-1312/10/4/505.

[136] Jeff C Jensen, Danica H Chang, and Edward A Lee. A model-based design methodology for cyber-physical systems. In *2011 7th international wireless communications and mobile computing conference*, pages 1666–1671. IEEE, 2011. ISBN 1424495385.

[137] Francesco Baldi, Gerasimos Theotokatos, and Karin Andersson. Development of a combined mean value–zero dimensional model and application for a large marine four-stroke Diesel engine simulation. *Applied Energy*, 154:402–415, 2015. ISSN 0306-2619.

[138] Imran Mahmood, Tameen Kausar, Hessam S Sarjoughian, Asad Waqar Malik, and Naveed Riaz. An integrated modeling, simulation and analysis framework for engineering complex systems. *IEEE Access*, 7:67497–67514, 2019. ISSN 2169-3536.

[139] Hassan Moussa Nahim, Rafic Younes, Hassan Shraim, and Mustapha Ouladsine. Oriented review to potential simulator for faults modeling in diesel engine. *Journal of Marine Science and Technology*, 21(3):533–551, 2016. ISSN 1437-8213.

[140] Kaitano Dzinavatonga, Ke Christ Obileke, Golden Makaka, Patrick Mukumba, and Benny Munyaradzi Nyambo. Determination of Thermal Time Constant of a Pt100 Temperature Sensor in Still Oil Using the Time Derivative Method. *Chemical Engineering & Technology*, 46(8):1673–1678, 8 2023. ISSN 1521-4125. doi: 10.1002/CEAT.202200119. URL https://onlinelibrary.wiley.com/doi/full/10.1002/ceat.202200119https://onlinelibrary.wiley.com/doi/abs/10.1002/ceat.202200119https://onlinelibrary.wiley.com/doi/10.1002/ceat.202200119.

[141] Sergejus Lebedevas, Justas Žaglinskis, and Martynas Drazdauskas. Development and Validation of Heat Release Characteristics Identification Method of Diesel Engine under Operating Conditions. *Journal of Marine Science and Engineering 2023, Vol. 11, Page 182*, 11(1):182, 1 2023. ISSN 2077-1312. doi: 10.3390/JMSE11010182. URL https://www.mdpi.com/2077-1312/11/1/182/htmhttps://www.mdpi.com/2077-1312/11/1/182.

[142] Pramod Ghimire, Mehdi Zadeh, and Eilif Pedersen. Co-Simulation of a Marine Hybrid Power System for Real-Time Virtual Testing. In *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*, pages 1–6. IEEE, 2021. ISBN 1728175836.

[143] Nicolai Pedersen, Jan Madsen, and Morten Vejlgaard-Laursen. Co-simulation of distributed engine control system and network model using FMI & SCNSL. *IFAC-PapersOnLine*, 48(16):261–266, 2015. ISSN 2405-8963.

[144] Stian Skjong and Eilif Pedersen. Co-simulation of a marine offshore vessel in dp-operations including hardware-in-the-loop (HIL). In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 57731, page V07AT06A038. American Society of Mechanical Engineers, 2017. ISBN 0791857735.

[145] Jens Bastian, Christop Clauß, Susann Wolf, and Peter Schneider. Master for co-simulation using FMI. *Proceedings of the 8th International Modelica Conference; March 20th-22nd; Technical Univeristy; Dresden; Germany*, 2011 (63):115–120, 2011.

[146] W Chen, M Klomp, and S Ran. Real-time co-simulation method study for vehicle steering and chassis system. *IFAC-PapersOnLine*, 51(9):273–278, 2018. ISSN 2405-8963.

[147] Cong Guan, Gerasimos Theotokatos, and Hui Chen. Analysis of two stroke marine diesel engine operation including turbocharger cut-out by using a zero-dimensional model. *Energies*, 8(6):5738–5764, 2015.

[148] Fayez Shakil Ahmed, Salah Laghrouche, Adeel Mehmood, and Mohammed El Bagdouri. Estimation of exhaust gas aerodynamic force on the variable geometry

turbocharger actuator: 1D flow model approach. *Energy conversion and management*, 84:436–447, 2014. ISSN 0196-8904.

[149] Laura von Rueden, Sebastian Mayer, Rafet Sifa, Christian Bauckhage, and Jochen Garcke. Combining machine learning and simulation to a hybrid modelling approach: Current and future directions. In *Advances in Intelligent Data Analysis XVIII: 18th International Symposium on Intelligent Data Analysis, IDA 2020, Konstanz, Germany, April 27–29, 2020, Proceedings 18*, pages 548–560. Springer, 2020. ISBN 3030445836.

[150] Andrea Coraddu, Miltiadis Kalikatzarakis, Gerasimos Theotokatos, Rinze Geertsma, and Luca Oneto. Physical and data-driven models hybridisation for modelling the dynamic state of a four-stroke marine diesel engine, 2022.

[151] Konstantinos-Marios Tsitsilonis, Gerasimos Theotokatos, Chaitanya Patil, and Andrea Coraddu. Health assessment framework of marine engines enabled by digital twins. *International Journal of Engine Research*, 2022. ISSN 1468-0874.

[152] William Q Meeker and Luis A Escobar. *Statistical methods for reliability data*. John Wiley \& Sons, 1998.

[153] KANG Rui, GONG Wenjun, and CHEN Yunxia. Model-driven degradation modeling approaches: Investigation and review. *Chinese Journal of Aeronautics*, 33(4):1137–1153, 2020. ISSN 1000-9361.

[154] Long P. Chin and Viswanath R. Katta. Numerical modeling of deposition in fuel-injection nozzles. *33rd Aerospace Sciences Meeting and Exhibit*, 1995. doi: 10.2514/6.1995-497. URL https://arc.aiaa.org/doi/10.2514/6.1995-497.

[155] Wenjie Qin and Lunjing Duan. Simulation of Cam Wear in Valve Trains of Diesel Engines. *SAE International Journal of Fuels and Lubricants*, 7(3), 2014. ISSN 19463960. doi: 10.4271/2014-01-2879.

[156] R Lewis and R S Dwyer-Joyce. Design tools for prediction of valve recession and solving valve/seat failure problems. *SAE Transactions*, pages 1868–1877, 2001. ISSN 0096-736X.

[157] Simon C. Tung and Yong Huang. Modeling of Abrasive Wear in a Piston Ring and Engine Cylinder Bore System©. *Tribology Transactions*, 47(1): 17–22, 2004. ISSN 1547397X. doi: 10.1080/05698190490279074. URL https://www.tandfonline.com/doi/abs/10.1080/05698190490279074.

[158] O. F. Eker, F. Camci, and I. K. Jennions. Physics-Based Degradation Modelling for Filter Clogging. *PHM Society European Conference*, 2(1), 7 2014. ISSN 2325-016X. doi: 10.36001/PHME.2014.V2I1.1478. URL https://www.papers.phmsociety.org/index.php/phme/article/view/1478.

[159] Rohan Patel, Jack Hadley, Austin Gabhart, Deepika Singla, Xiao Wei, Jacob Grant, Nicole Robertson, Neil Weston, Michael Steffens, and Dimitri Mavris. Maritime Autonomous System Design Methods and Technology Forecasting. *SNAME 14th International Marine Design Conference, IMDC 2022*, 6 2022. doi: 10.5957/IMDC-2022-352. URL https://dx.doi.org/10.5957/IMDC-2022-352.

[160] Christina Diakaki, Natalia Panagiotidou, Anastasios Pouliezos, Georgios D. Kontes, George S. Stavrakakis, Kostas Belibassakis, Theodoros P. Gerostathis, George Livanos, Dimitrios Nikolaos Pagonis, and Gerasimos Theotokatos. A decision support system for the development of voyage and maintenance plans for ships. *International Journal of Decision Support Systems*, 1(1):42–71, 2015. ISSN 2050-6988. doi: 10.1504/IJDSS.2015.067274.

[161] Francesco Di Cerbo, Pascal Bisson, Alan Hartman, Sebastien Keller, Per Håkon Meland, Micha Moffie, Nazila Gol Mohammadi, Sachar Paulus, and Stuart Short. Towards trustworthiness assurance in the cloud. In *Cyber Security and Privacy Forum*, pages 3–15. Springer, 2013.

[162] Brian Connett and Bryan O'Halloran. Systems engineering design: Architecting trustworthiness in cyber physical systems using an extended aggregated modality. *Procedia computer science*, 140:4–12, 2018. ISSN 1877-0509.

[163] Fred B Schneider, Steven M Bellovin, and Alan S Inouye. Building trustworthy systems: Lessons from the PTN and Internet. *IEEE Internet Computing*, 3(6): 64–72, 1999. ISSN 1089-7801.

[164] Brydon T. Wang and Mark Burdon. Automating trustworthiness in digital twins. *Advances in 21st Century Human Settlements*, pages 345–365, 2021. ISSN 21982554. doi: 10.1007/978-981-15-8670-5{_}14/FIGURES/1. URL https://link.springer.com/chapter/10.1007/978-981-15-8670-5_14.

[165] Radu F Babiceanu and Remzi Seker. Trustworthiness requirements for manufacturing cyber-physical systems. *Procedia Manufacturing*, 11:973–981, 2017. ISSN 2351-9789.

[166] Jose Luis de la Vara, Alejandra Ruiz, and Gaël Blondelle. Assurance and certification of cyber–physical systems: The AMASS open source ecosystem. *Journal of Systems and Software*, 171, 2021. ISSN 01641212. doi: 10.1016/j.jss.2020.110812.

[167] Nima Gorjian, Lin Ma, Murthy Mittinty, Prasad Yarlagadda, and Yong Sun. A review on degradation models in reliability analysis. *Engineering asset lifecycle management*, pages 369–384, 2010.

[168] Gamma Technologies. GT-SUITE Manual, 2022.

[169] Sokratis Stoumpos, Gerasimos Theotokatos, Evangelos Boulougouris, Dracos Vassalos, Iraklis Lazakis, and George Livanos. Marine dual fuel engine modelling and parametric investigation of engine settings effect on performance-emissions trade-offs. *Ocean Engineering*, 157:376–386, 2018. ISSN 0029-8018.

[170] Sokratis Stoumpos, Gerasimos Theotokatos, Christoforos Mavrelos, and Evangelos Boulougouris. Towards marine dual fuel engines digital twins–integrated modelling of thermodynamic processes and control system functions. *Journal of Marine Science and Engineering*, 8(3):200, 2020.

[171] Gerhard Woschni. A universally applicable equation for the instantaneous heat transfer coefficient in the internal combustion engine. Technical report, SAE Technical paper, 1967.

[172] Günter P Merker, Christian Schwarz, Gunnar Stiesch, and Frank Otto. *Simulating Combustion: Simulation of combustion and pollutant formation for engine-development*. Springer Science & Business Media, 2005. ISBN 3540306269.

[173] Kai Song, Peng Xu, Guo Wei, Yinsheng Chen, and Qi Wang. Health Management Decision of Sensor System Based on Health Reliability Degree and Grey Group Decision-Making. *Sensors 2018, Vol. 18, Page 2316*, 18(7):2316, 7 2018. ISSN 1424-8220. doi: 10.3390/S18072316. URL https://www.mdpi.com/1424-8220/18/7/2316/htmhttps://www.mdpi.com/1424-8220/18/7/2316.

[174] Kshitij Jerath, Sean Brennan, and Constantino Lagoa. Bridging the gap between sensor noise modeling and sensor characterization. *Measurement*, 116: 350–366, 2018. ISSN 0263-2241.

[175] Wärtsilä. Wärtsilä 50DF Product Guide, 2012. URL https://cdn.wartsila.com/docs/default-source/product-files/engines/df-engine/product-guide-o-e-w50df.pdf?sfvrsn=9.

[176] Siri Solem, Kjetil Fagerholt, Stein Ove Erikstad, and Øyvind Patricksson. Optimization of diesel electric machinery system configuration in conceptual ship design. *Journal of Marine Science and Technology*, 20:406–416, 2015.

[177] Robert Borgovini, Stephen Pemberton, and Michael Rossi. Failure mode, effects, and criticality analysis (FMECA). *Reliability Analysis Center*, 1993.

[178] Hu-Chen Liu. *FMEA Using Uncertainty Theories and MCDM Methods*. Springer, Singapore, 2016. ISBN 978-981-10-1466-6. doi: 10.1007/978-981-10-1466-6{_}2. URL https://link.springer.com/chapter/10.1007/978-981-10-1466-6_2.

[179] T Slatter, R Lewis, and R S Dwyer-Joyce. Valve recession modelling. Technical report, SAE Technical Paper, 2006.

[180] Jan M van Noortwijk. A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1):2–21, 2009.

[181] D E Huntington and C S Lyrintzis. Improvements to and limitations of Latin hypercube sampling. *Probabilistic engineering mechanics*, 13(4):245–253, 1998. ISSN 0266-8920.

[182] Sigrún Andradóttir. Simulation optimization. In *Handbook of simulation: Principles, methodology, advances, applications, and practice*, pages 307–333. JOHN WILEY & SONS, INC, 1998.

[183] Kalyanmoy Deb and Himanshu Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. *IEEE transactions on evolutionary computation*, 18(4):577–601, 2013. ISSN 1089-778X.

[184] ISO. ISO 15550 - Internal Combustion Engines: Determination and method for the measurement of engine power-General requirements. Technical report, International Organization for Standardization, 2016.

[185] Jack P C Kleijnen. Sensitivity analysis of simulation models. *SSRN Electronic Journal*, 2009.

[186] Jan Hauke and Tomasz Kossowski. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, 30(2):87, 2011. ISSN 2082-2103.

[187] Chengwei Xiao, Jiaqi Ye, Rui Máximo Esteves, and Chunming Rong. Using Spearman's correlation coefficients for exploratory data analysis on big dataset. *Concurrency and Computation: Practice and Experience*, 28(14):3866–3878, 2016. ISSN 1532-0626.

[188] Arthur Vrijdag, Paul Schulten, Douwe Stapersma, and Tom Van Terwisga. Efficient uncertainty analysis of a complex multidisciplinary simulation model. *Journal of Marine Engineering & Technology*, 6(2):79–88, 9 2007. ISSN 14761548. doi: 10.1080/20464177.2007.11020204. URL https://www.tandfonline.com/doi/abs/10.1080/20464177.2007.11020204.

[189] Eric Smith. Uncertainty analysis. *Encyclopedia of environmetrics*, 4:2283–2297, 2002.

[190] David C Cox and Paul Baybutt. Methods for uncertainty analysis: a comparative survey. *Risk Analysis*, 1(4):251–258, 1981. ISSN 0272-4332.

[191] Alice Zheng and Amanda Casari. *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc.", 2018. ISBN 1491953195.

[192] David Roxbee Cox and Nancy Reid. *The theory of the design of experiments.* Chapman and Hall/CRC, 2000. ISBN 042912628X.

[193] Weiming Meng, Keshun Li, Wei Feng, and Xingjia Jiang. Study on the evaluation module of ship operation management under big data view. In *Proceedings - 2020 International Conference on E-Commerce and Internet Technology, ECIT 2020*, pages 82–84, 2020. doi: 10.1109/ECIT50008.2020.00027.

[194] Neal B Gallagher. Savitzky-Golay smoothing and differentiation filter. *Eigenvector Research Incorporated*, 2020.

[195] SGOPAL Patro and Kishore Kumar Sahu. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015.

[196] Qin Liang, Knut Erik Knutsen, Erik Vanem, Vilmar Æsøy, and Houxiang Zhang. A review of maritime equipment prognostics health management from a classification society perspective. *Ocean Engineering*, 301:117619, 6 2024. ISSN 0029-8018. doi: 10.1016/J.OCEANENG.2024.117619.

[197] Haedong Jeong, Bumsoo Park, Seungtae Park, Hyungcheol Min, and Seungchul Lee. Fault detection and identification method using observer-based residuals. *Reliability Engineering & System Safety*, 184:27–40, 2019. ISSN 0951-8320.

[198] Chih Wei Hsu and Chih Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 2002. ISSN 10459227. doi: 10.1109/72.991427.

[199] Sebastian Raschka and Vahid Mirjalili. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2.* Packt Publishing Ltd., Birmingham, third edition edition, 2019. ISBN 978-1-78995-575-0.

[200] Oliver Kramer. *Dimensionality Reduction with Unsupervised Nearest Neighbors*, volume 51. Springer, 2013. ISBN 9783642386510. doi: 10.1007/978-3-642-38652-7/COVER.

[201] Chengtao Cai, Xiangyu Weng, and Chuanbin Zhang. A novel approach for marine diesel engine fault diagnosis. *Cluster computing*, 20(2):1691–1702, 2017. ISSN 1573-7543.

[202] Iraklis Lazakis, Christos Gkerekos, and Gerasimos Theotokatos. Investigating an SVM-driven, one-class approach to estimating ship systems condition. *Ships and Offshore Structures*, 14(5):432–441, 2019. ISSN 1744-5302.

[203] Bernard Mulgrew. Applying radial basis functions. *IEEE Signal Processing Magazine*, 13(2), 1996. ISSN 10535888. doi: 10.1109/79.487041.

[204] Ruonan Liu, Boyuan Yang, Enrico Zio, and Xuefeng Chen. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108:33–47, 8 2018. ISSN 0888-3270. doi: 10.1016/J.YMSSP.2018.02.016.

[205] Jan Monieta and Lech Kasyk. Application of Machine Learning to Classify the Technical Condition of Marine Engine Injectors Based on Experimental Vibration Displacement Parameters. *Energies 2023, Vol. 16, Page 6898*, 16(19):6898, 9 2023. ISSN 1996-1073. doi: 10.3390/EN16196898. URL https://www.mdpi.com/1996-1073/16/19/6898/htmhttps://www.mdpi.com/1996-1073/16/19/6898.

[206] Joel Grus. *Data science from scratch: first principles with python*. O'Reilly Media, 2019. ISBN 9781492041139. URL http://oreilly.com.

[207] Leandro L Minku, George Cabral, Marcella Martins, and Markus Wagner. *Introduction to Computational Intelligence*. An IEEE Computational Intelligence Society Open Book., 2023. doi: 10.5281/zenodo.7537827. URL https://github.com/ieee-cis/IEEE-CIS-Open-Access-Book-Volume-1.

[208] Fazl Gökgöz and Fahrettin Filiz. Electricity price forecasting: A comparative

analysis with shallow-ann and dnn. *International Journal of Energy and Power Engineering*, 12(6):421–425, 2018.

[209] Mohammadreza Heydarian, Thomas E Doyle, and Reza Samavi. MLCM: multi-label confusion matrix. *IEEE Access*, 10:19083–19095, 2022. ISSN 2169-3536.

[210] A. C. Davison. *Statistical Models*, volume 11. Cambridge University Press, 2003. ISBN 9780521773393. doi: 10.1017/CBO9780511815850.

[211] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 10 2001. ISSN 08856125. doi: 10.1023/A:1010933404324/METRICS. URL https://link.springer.com/article/10.1023/A:1010933404324.

[212] Hoang Pham. *Springer Handbook of Engineering Statistics*. Springer London, 2006. doi: 10.1007/978-1-84628-288-1.

[213] A Liaw and M Wiener. Classification and Regression by randomForest. *R News*, 2(3), 2002.

[214] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 8 2004. ISSN 09603174. doi: 10.1023/B:STCO.0000035301.49549.88/METRICS. URL https://link.springer.com/article/10.1023/B:STCO.0000035301.49549.88.

[215] Zihao Wang, Haitong Xu, Li Xia, Zaojian Zou, and C. Guedes Soares. Kernel-based support vector regression for nonparametric modeling of ship maneuvering motion. *Ocean Engineering*, 216:107994, 11 2020. ISSN 0029-8018. doi: 10.1016/J.OCEANENG.2020.107994.

[216] Zhengmin Kong, Yande Cui, Zhou Xia, and He Lv. Convolution and Long Short-Term Memory Hybrid Deep Neural Networks for Remaining Useful Life Prognostics. *Applied Sciences 2019, Vol. 9, Page 4156*, 9(19):4156, 10 2019. ISSN 2076-3417. doi: 10.3390/APP9194156. URL https://www.mdpi.com/2076-3417/9/19/4156/htmhttps://www.mdpi.com/2076-3417/9/19/4156.

[217] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[218] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. A comparison of ARIMA and LSTM in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pages 1394–1401, 2018.

[219] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data 2021 8:1*, 8(1): 1–74, 3 2021. ISSN 2196-1115. doi: 10.1186/S40537-021-00444-8. URL https://link.springer.com/articles/10.1186/s40537-021-00444-8https://link.springer.com/article/10.1186/s40537-021-00444-8.

[220] Zhixiong Li, Dazhong Wu, Chao Hu, and Janis Terpenny. An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction. *Reliability Engineering & System Safety*, 184:110–122, 4 2019. ISSN 0951-8320. doi: 10.1016/J.RESS.2017.12.016.

[221] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell Deepmind. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Advances in Neural Information Processing Systems*, 30, 2017.

[222] Hao Wang, Yi Ming Zhang, Jian Xiao Mao, and Hua Ping Wan. A probabilistic approach for short-term prediction of wind gust speed using ensemble learning. *Journal of Wind Engineering and Industrial Aerodynamics*, 202:104198, 7 2020. ISSN 0167-6105. doi: 10.1016/J.JWEIA.2020.104198.

[223] Zhi-Hua Zhou and Zhi-Hua Zhou. *Ensemble learning*. Springer, 2021.

[224] Laurent Valentin Jospin, Hamid Laga, Farid Boussaid, Wray Buntine, and Mohammed Bennamoun. Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, 17(2):29–48, 2022.

[225] Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14, 2013. ISSN 15324435.

[226] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *33rd International Conference on Machine Learning, ICML 2016*, volume 3, pages 1050–1059, 2016.

[227] Lingxue Zhu and Nikolay Laptev. Deep and Confident Prediction for Time Series at Uber. *IEEE International Conference on Data Mining Workshops, ICDMW*, 2017-November:103–110, 12 2017. ISSN 23759259. doi: 10.1109/ICDMW.2017.19.

[228] Jacques Wainer and Pablo Fonseca. How to tune the RBF SVM hyperparameters? An empirical evaluation of 18 search algorithms. *Artificial Intelligence Review*, 54(6), 2021. ISSN 15737462. doi: 10.1007/s10462-021-10011-5.

[229] Hussain Alibrahim and Simone A Ludwig. Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization. In *2021 IEEE Congress on Evolutionary Computation (CEC)*, pages 1551–1559, 2021.

[230] Morteza Moradi, Agnes Broer, Juan Chiachío, Rinze Benedictus, Theodoros H. Loutas, and Dimitrios Zarouchas. Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data. *Engineering Applications of Artificial Intelligence*, 117:105502, 1 2023. ISSN 0952-1976. doi: 10.1016/J.ENGAPPAI.2022.105502.

[231] Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 2021. ISSN 23765992. doi: 10.7717/PEERJ-CS.623.

[232] Iain Pardoe. *Applied Regression Modeling, Third Edition*. John Wiley \& Sons, 2020. ISBN 9781119615941. doi: 10.1002/9781119615941. URL https://onlinelibrary.wiley.com/doi/book/10.1002/9781119615941.

[233] Peterson K Ozili. The acceptable R-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native english speakers*, pages 134–143. IGI Global, 2023.

[234] R.J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia, 2018.

[235] Gerasimos Theotokatos, Sokratis Stoumpos, Victor Bolbot, and Evangelos Boulougouris. Simulation-based investigation of a marine dual-fuel engine. *Journal of Marine Engineering & Technology*, 19(sup1):5–16, 2020. ISSN 2046-4177.

[236] Roger Lewis. *Wear of diesel engine inlet valves and seats*. PhD thesis, University of Sheffield, Sheffield, 2000.

[237] Jerzy Kowalski. An experimental study of emission and combustion characteristics of marine diesel engine in case of cylinder valves leakage. *Polish Maritime Research*, 22(3):90–98, 2015. ISSN 1233-2585.

[238] Jose Antonio Pagán Rubio, Francisco Vera-García, Jose Hernandez Grau, Jose Muñoz Cámara, and Daniel Albaladejo Hernandez. Marine diesel engine failure simulator based on thermodynamic model. *Applied Thermal Engineering*, 144:982–995, 2018. ISSN 1359-4311.

[239] N D Whitehouse, A Stotter, and M S Janota. Estimating the effects of altitude, ambient temperature and turbocharger match on engine performance. *Proceedings of the Institution of Mechanical Engineers*, 178(1):483–500, 1963. ISSN 0020-3483.

[240] J R Serrano, C Guardiola, V Dolz, A Tiseira, and C Cervelló. Experimental study of the turbine inlet gas temperature influence on turbocharger performance. Technical report, SAE Technical Paper, 2007.

[241] MAN. Influence of ambient temperature conditions. *Copenhagen: MAN Diesel and Turbo*, 2014.

[242] Sun Tong, Chen Yanqiao, and Zhou Yuan. Fault prediction of marine diesel engine based on time series and support vector machine. In *Proceedings - 2020 International Conference on Intelligent Design, ICID 2020*, 2020. doi: 10.1109/ICID52250.2020.00023.

Appendix A

Method Comparison Result Plot

This appendix visualised methods comparison results that were described in Section 7.2.2. The selected methods for each sub-model were compared employing four distinctive testing datasets. Figures A.1, A.2, A.3, and A.4 illustrated the results for the HI construction model, while Figures A.5, A.6, A.7, and A.8 illustrated the results for the HI Forecast model. Figures A.9, A.10, A.11, and A.12 illustrated the HI forecast sub-task results of the integrated prognosis, along with the actual HI values.

A.1 HI Construction Method Comparison Result Plots

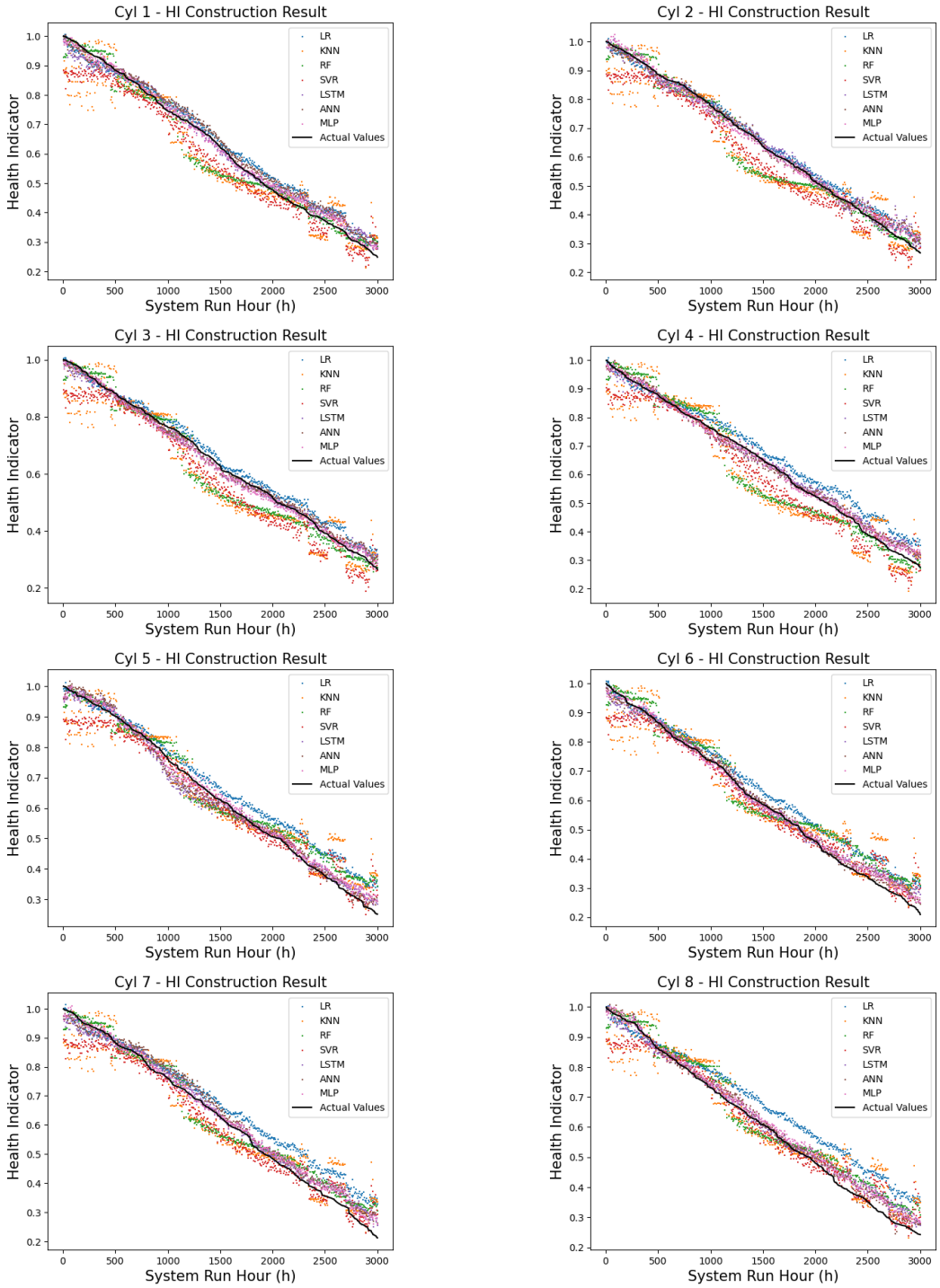


Figure A.1: Dataset 1 – HI construction model results for cylinders 1–8

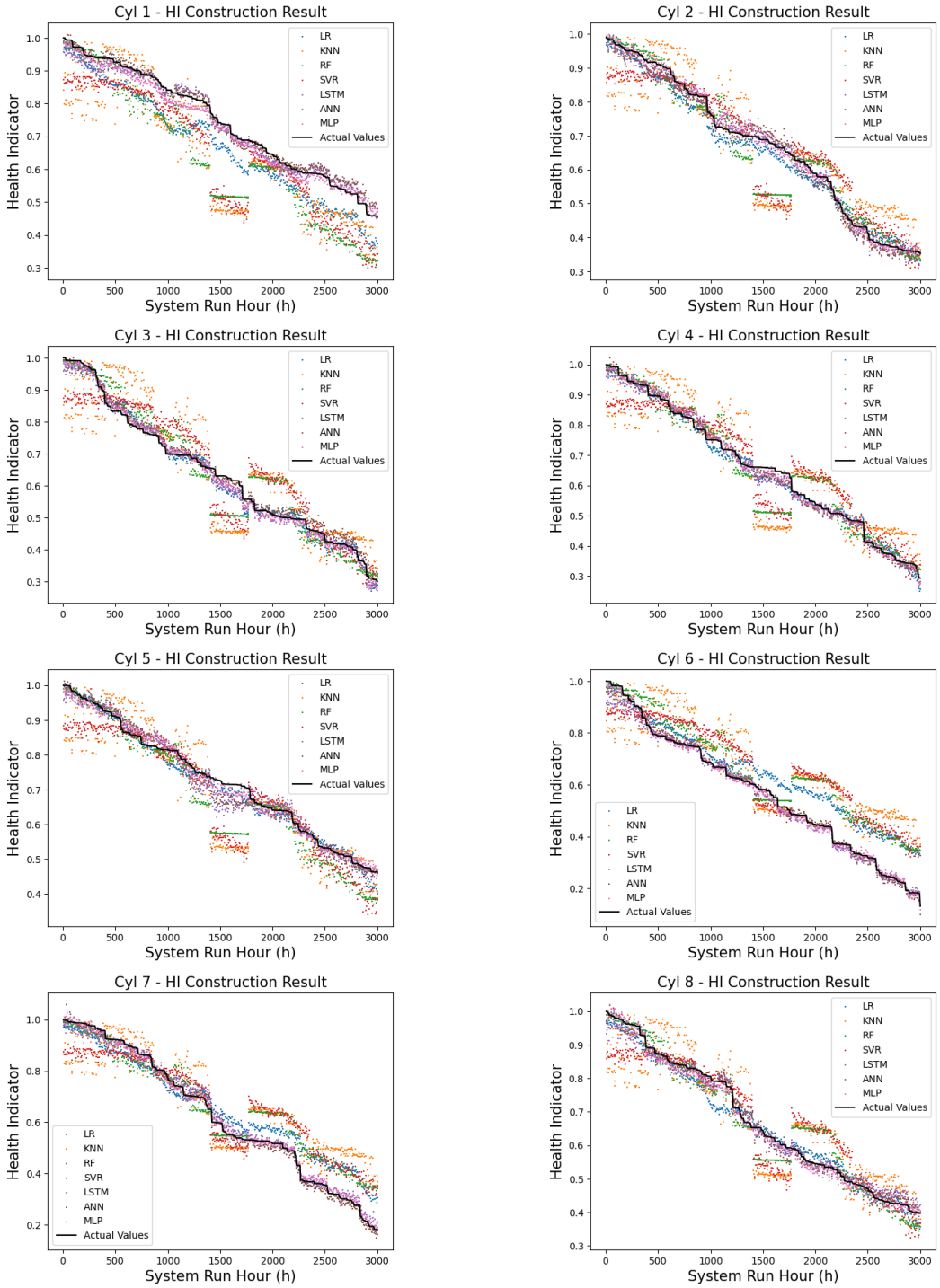


Figure A.2: Dataset 2 – HI construction model results for cylinders 1–8

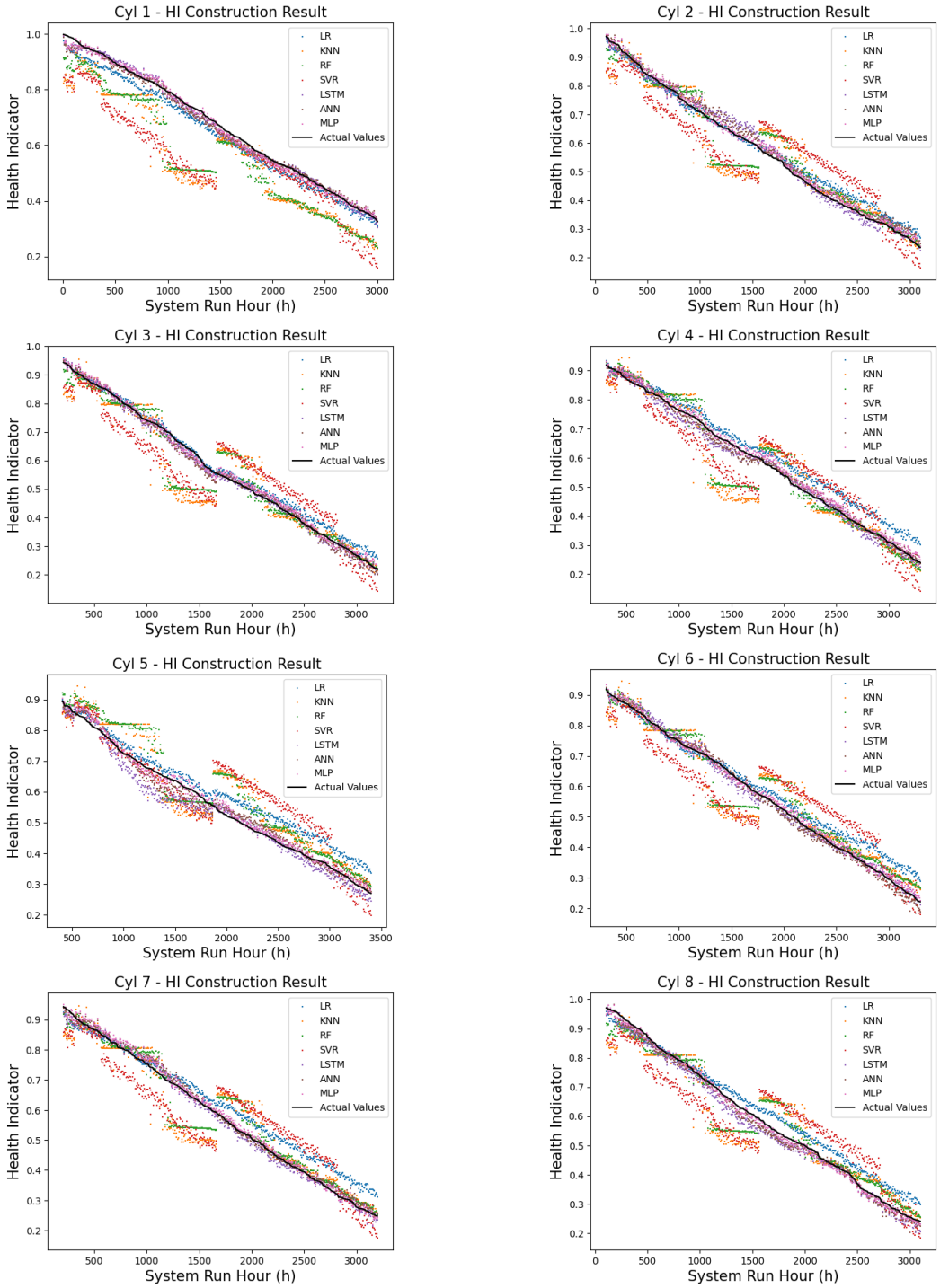


Figure A.3: Dataset 3 – HI construction model results for cylinders 1–8

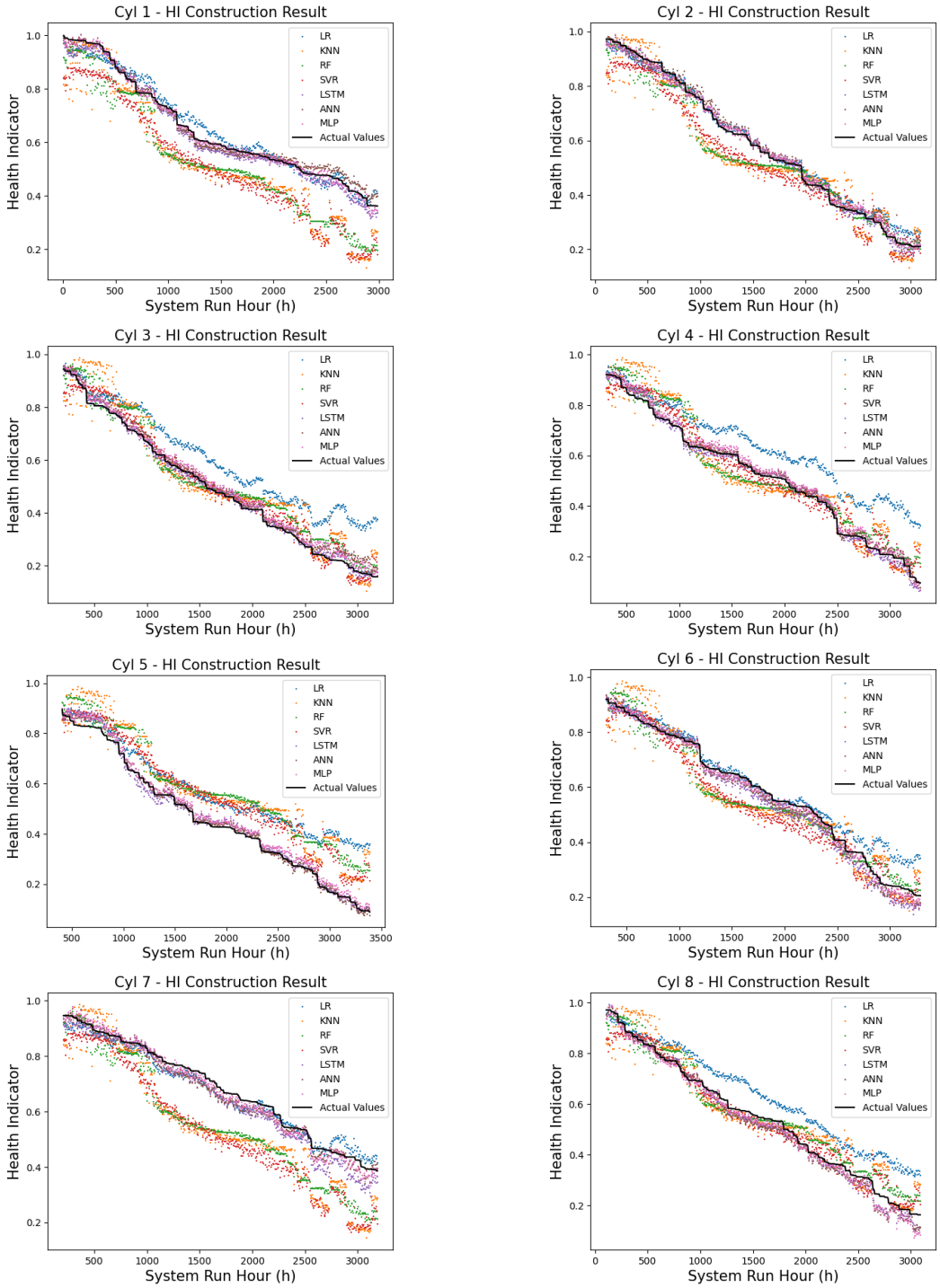


Figure A.4: Dataset 4 – HI construction model results for cylinders 1–8

A.2 HI Forecast Method Comparison Result Plots

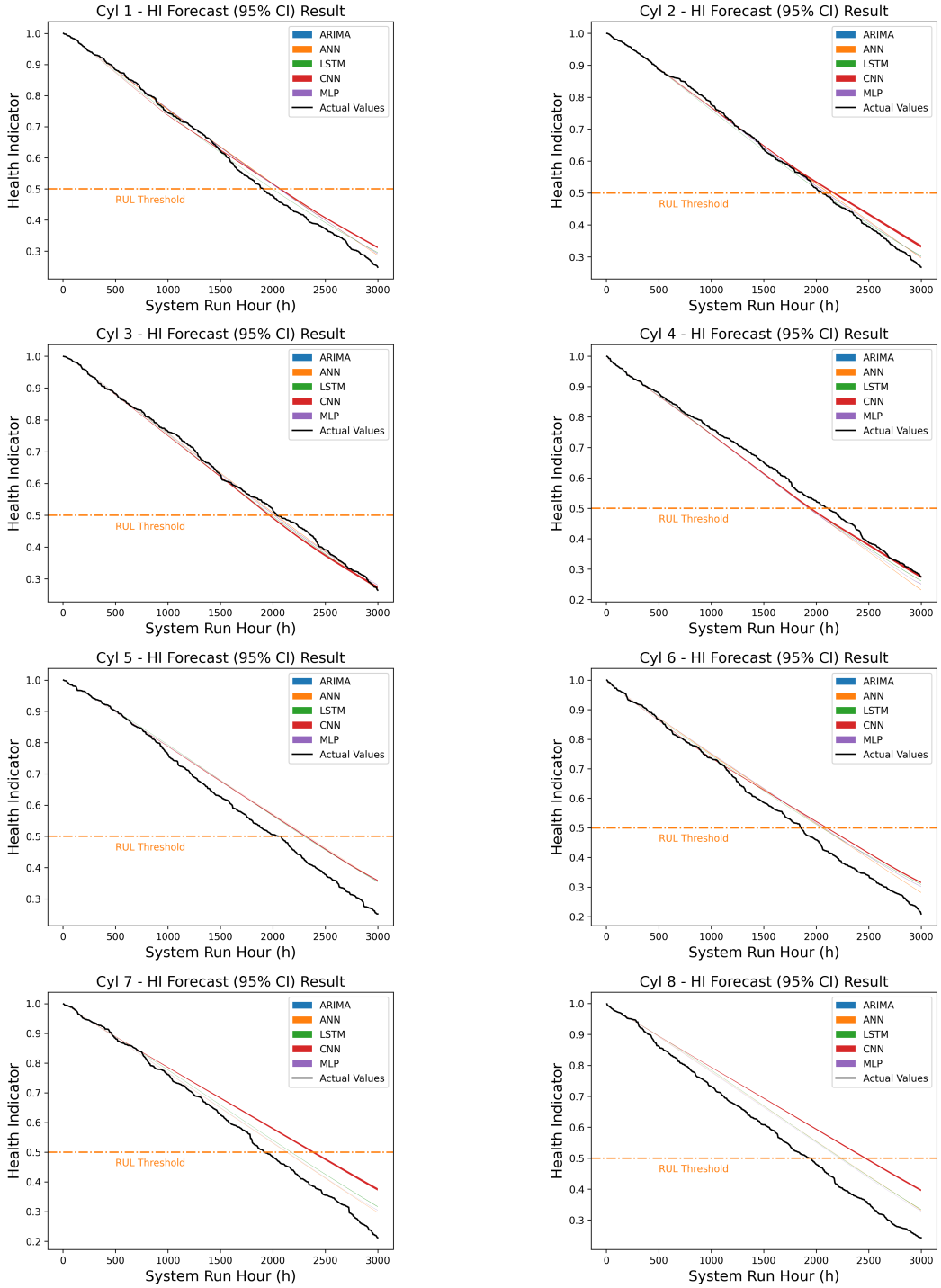


Figure A.5: Dataset 1 – HI forecast model results for cylinders 1–8

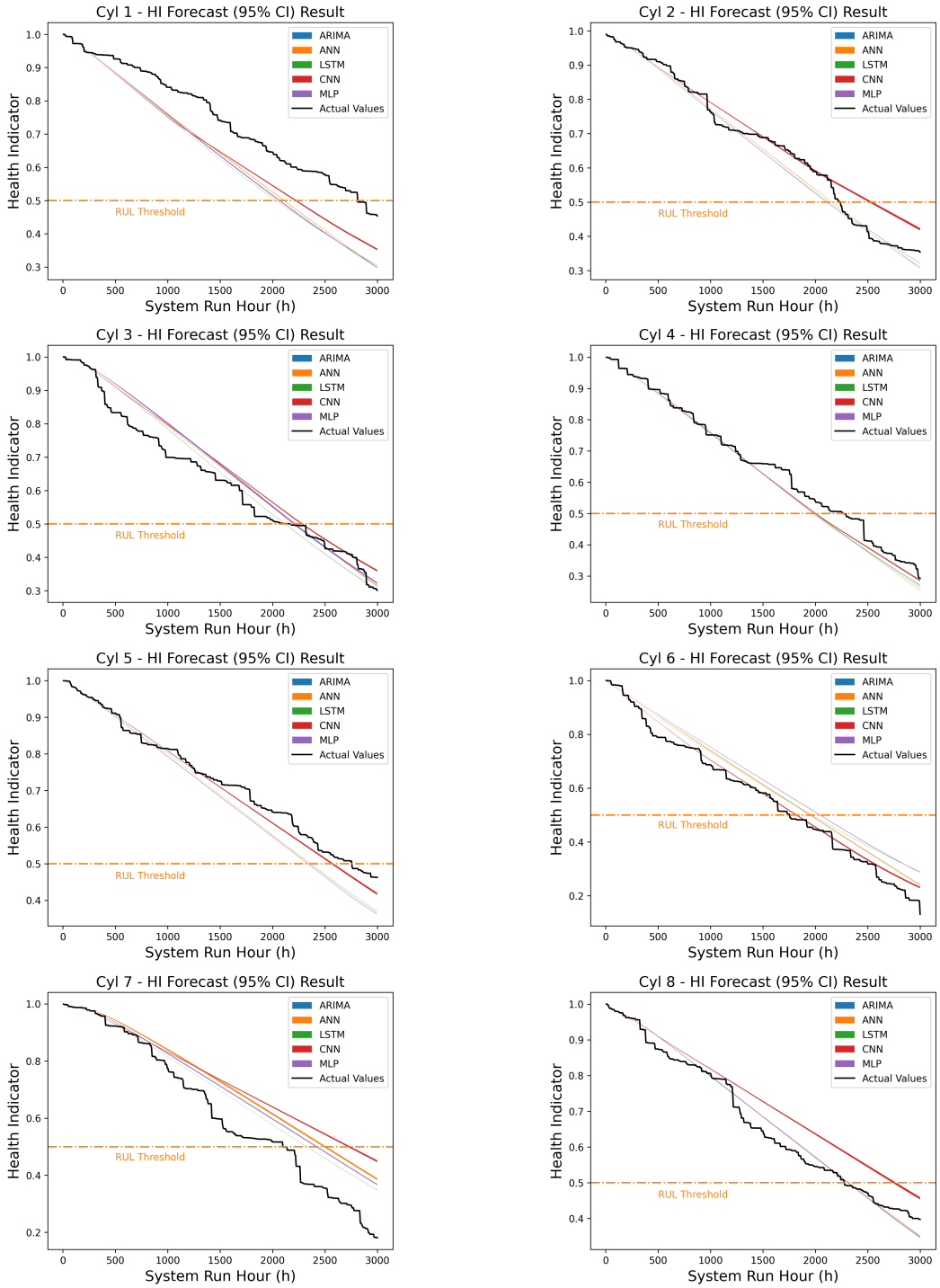


Figure A.6: Dataset 2 – HI forecast model results for cylinders 1–8

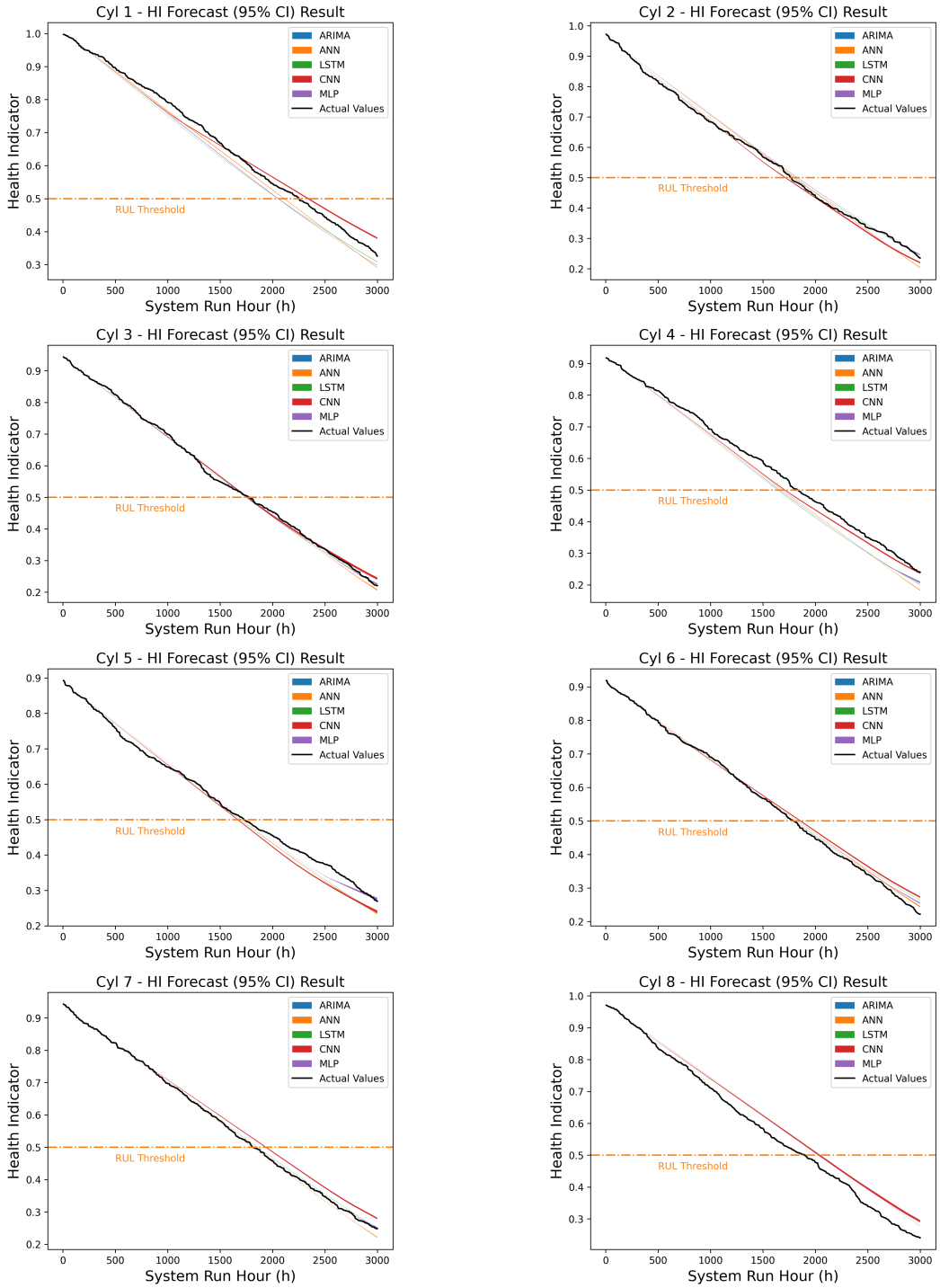


Figure A.7: Dataset 3 – HI forecast model results for cylinders 1–8

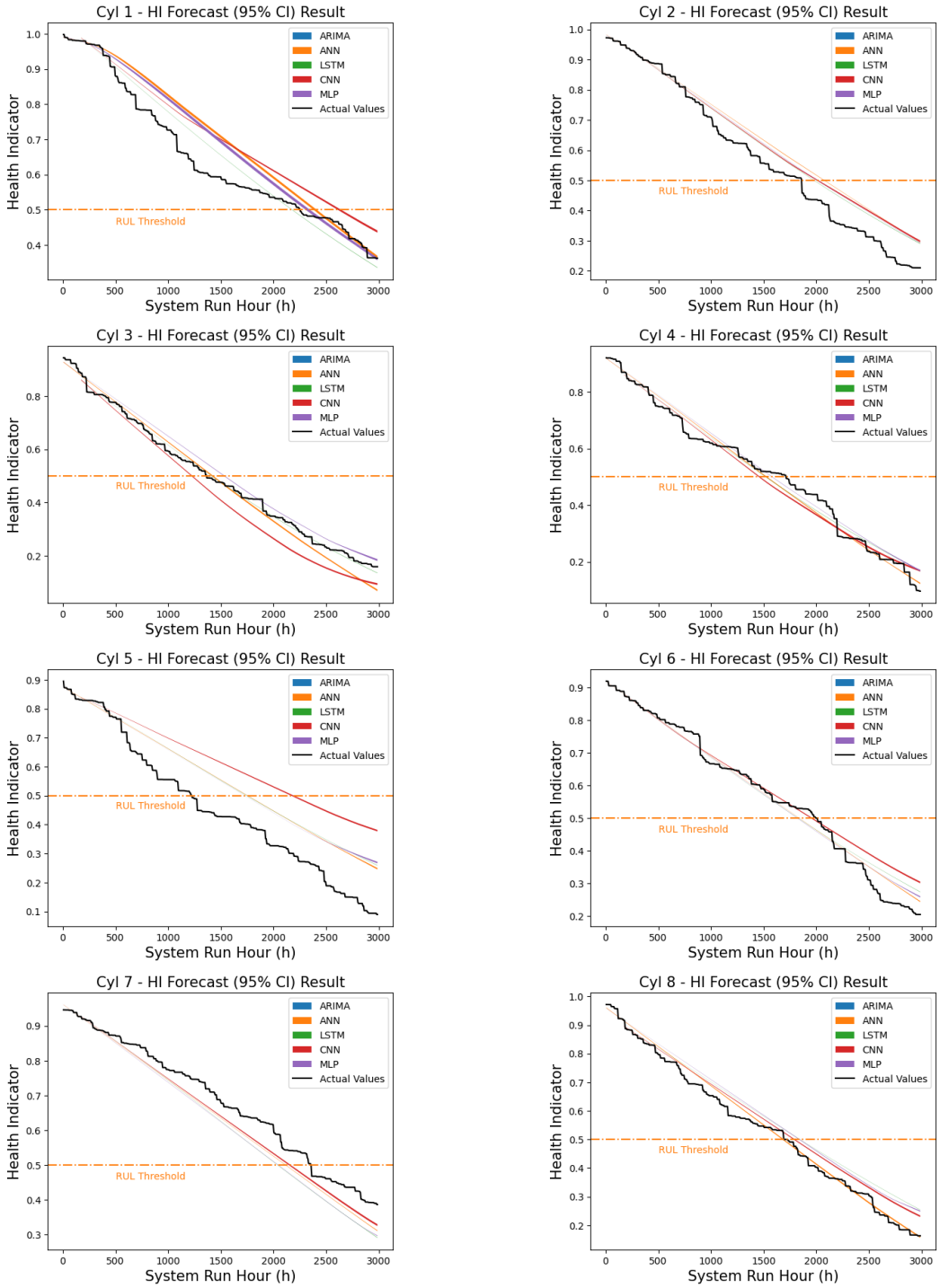


Figure A.8: Dataset 4 – HI forecast model results for cylinders 1–8

A.3 Integrated Prognosis Model HI Forecast Result Plots

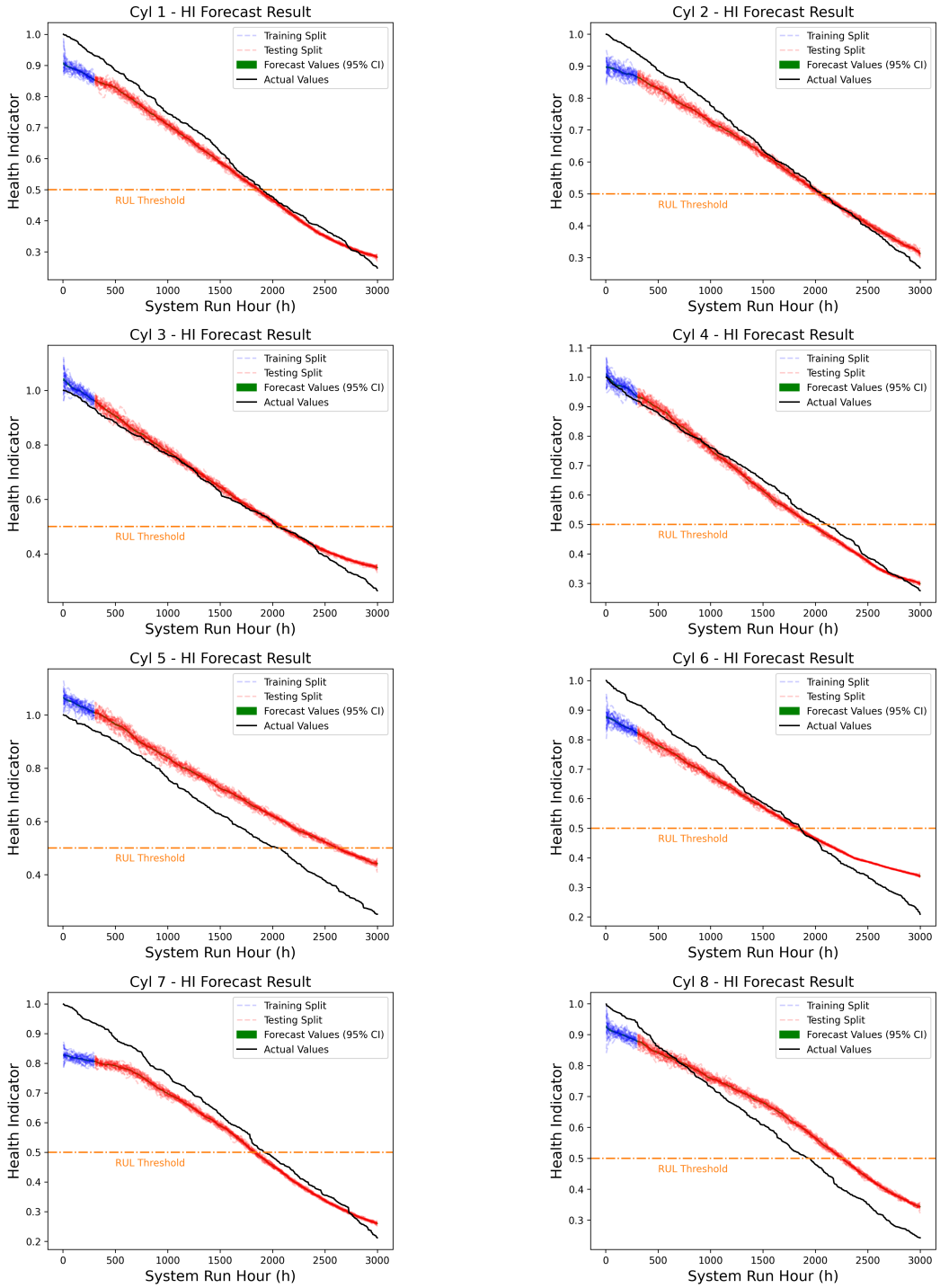


Figure A.9: Dataset 1 – Combined model results for cylinders 1–8

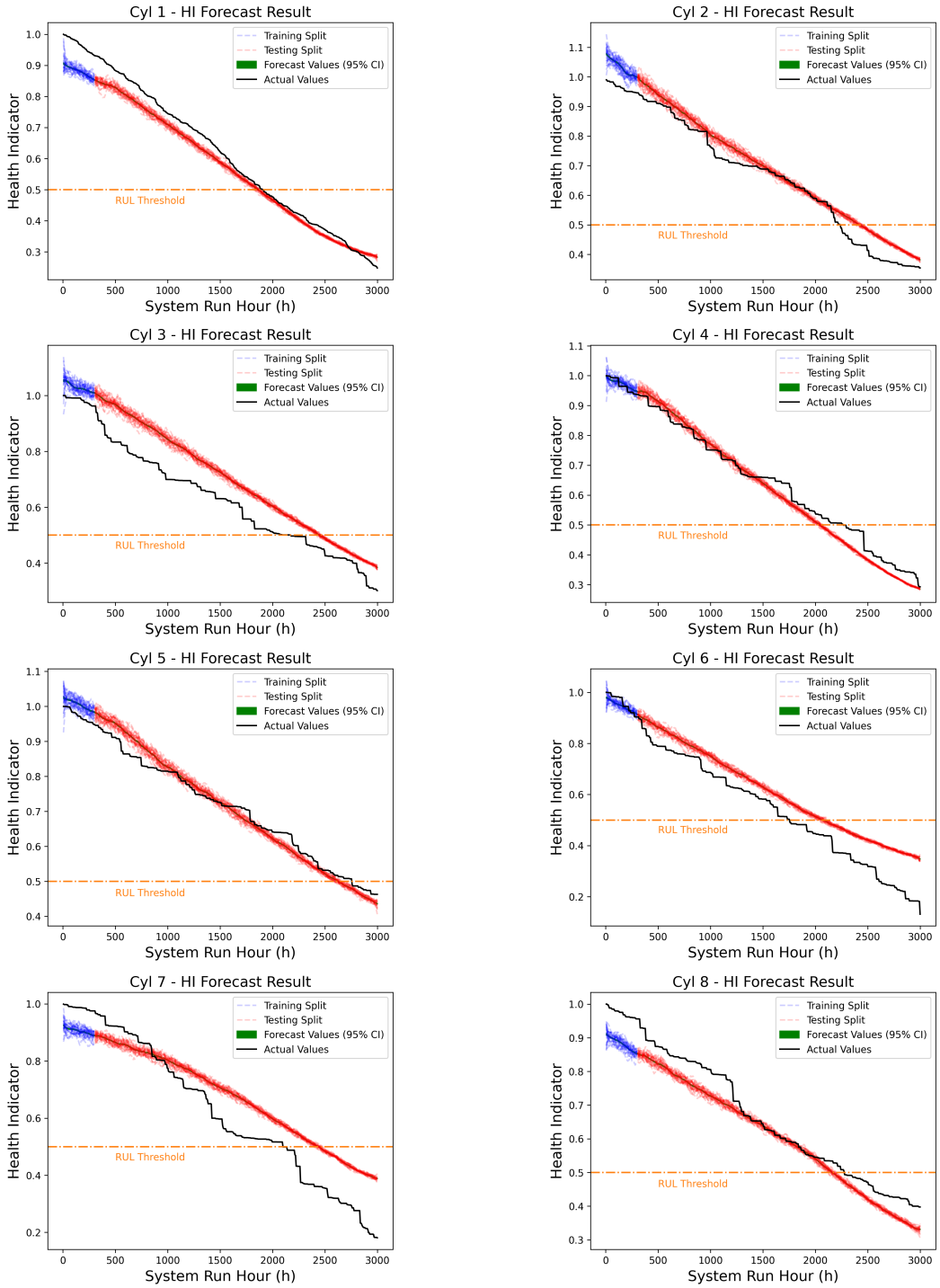


Figure A.10: Dataset 2 – Combined model results for cylinders 1–8

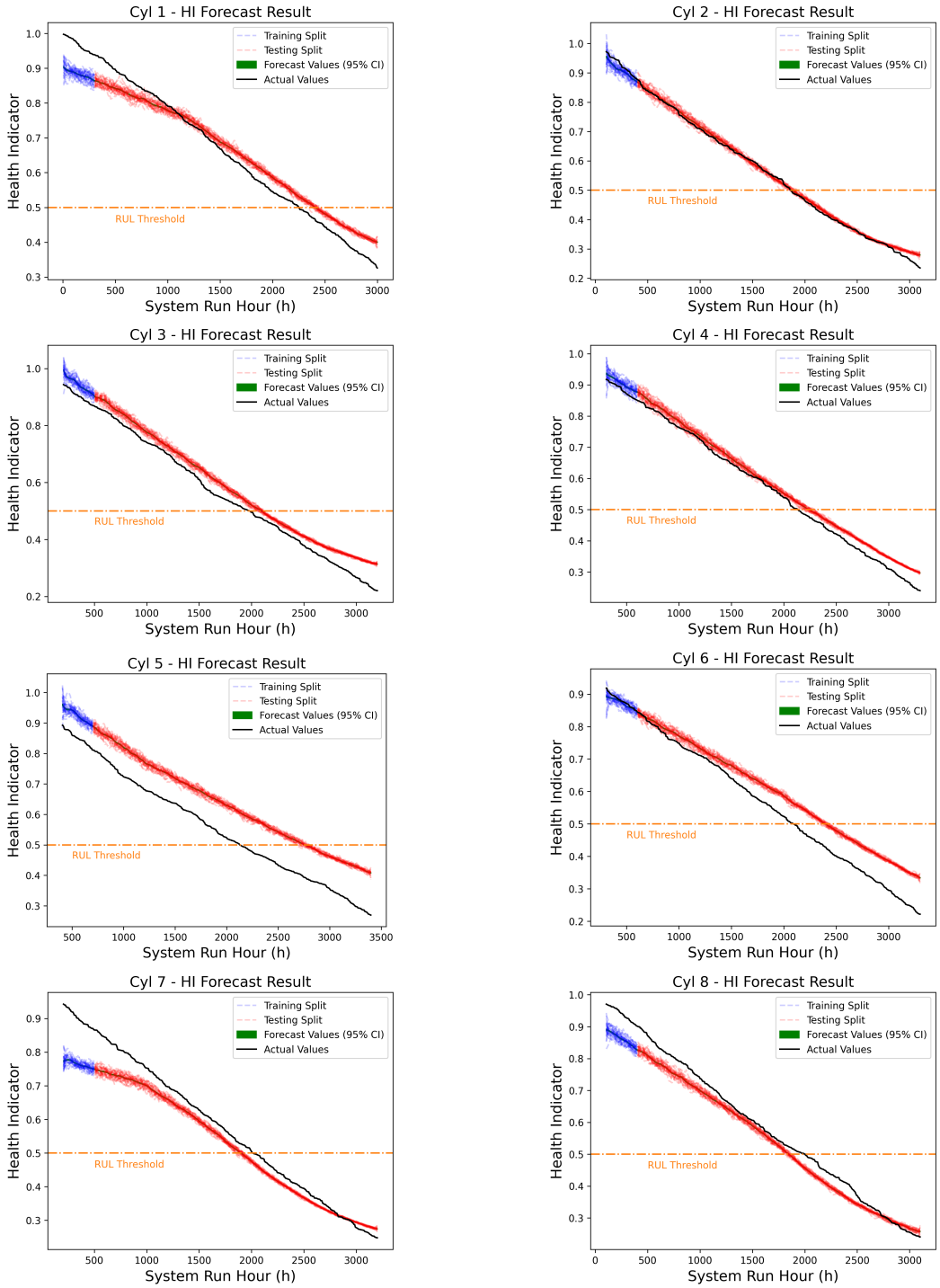


Figure A.11: Dataset 3 – Combined model results for cylinders 1–8

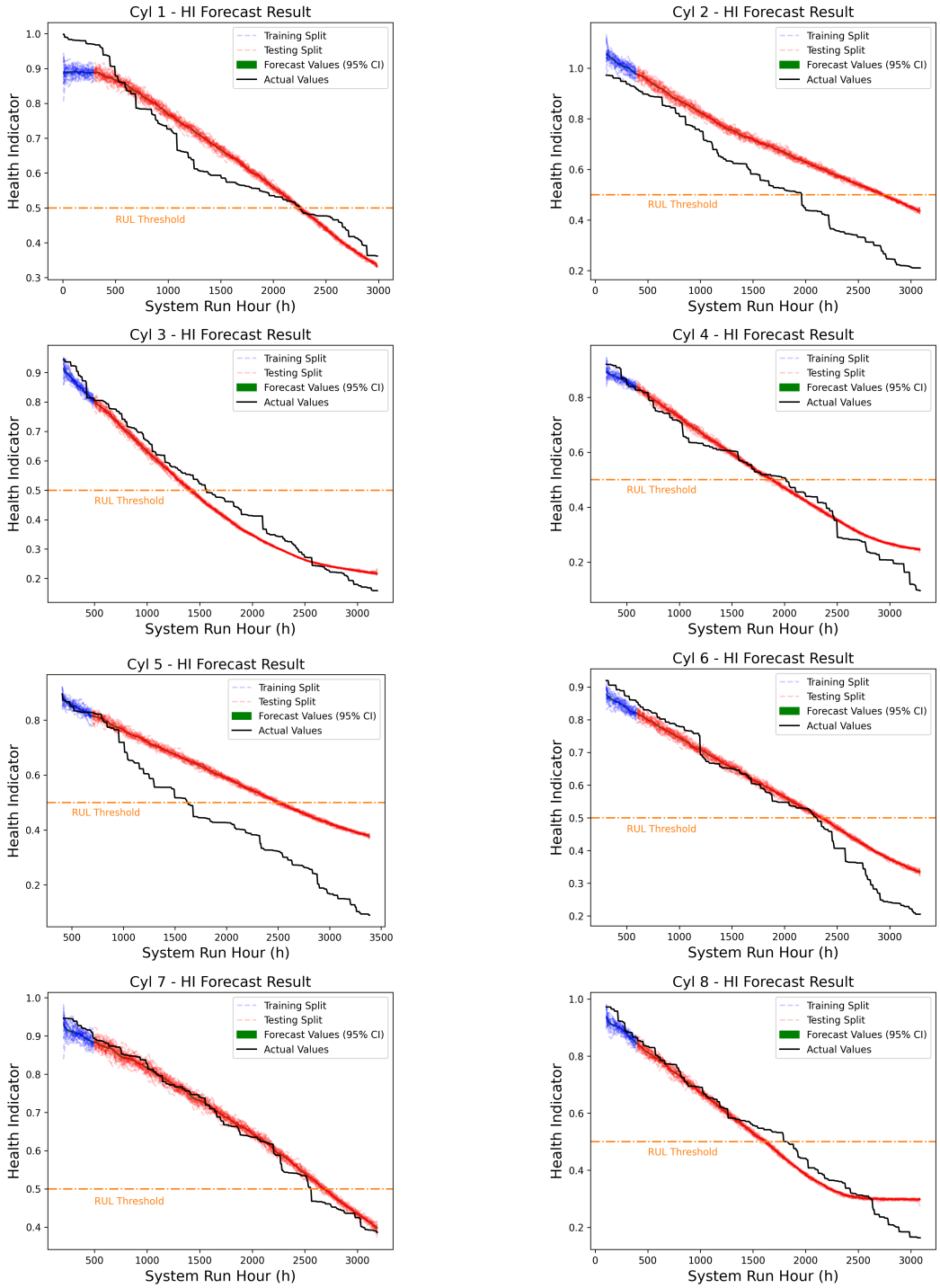


Figure A.12: Dataset 4 – Combined model results for cylinders 1–8