JAITA

JOURNAL OF ADVANCED INDUSTRIAL TECHNOLOGY AND APPLICATION
e-ISSN: 2716-7097

Vol. 7 No. 1 (2026) 62-81
https://publisher.uthm.edu.my/ojs/index.php/jaita

The Application of Machine Learning and Reliability Block Diagram (RBD) in Prognostic Health Management (PHM) for Pumps

Ruwaida Aliyu[1]*, Ainul Akmar Mokhtar[2], Sirisha Nalakukkala[3], Ibrahim Temitope Amosa[4] Abdulrahman Aminu Ghali[5], Elhuseini Garba[2]

[1] *Department of Mechanical Engineering, Baze University Abuja, NIGERIA*
[2] *Mechanical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610 Perak Darul Ridzuan, MALAYSIA*
[3] *Chemical Engineering Department, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, Tamil Nadu, INDIA*
[4] *Edwardson School of Industrial Engineering, Purdue University, West Lafayette, UNITED STATES*
[5] *Faculty of Information and Communication Technology (FICT), Universiti Tunku Abdul Rahman, Kampar, Petaling Jaya 31900, MALAYSIA*

*Corresponding Author: ruwaida.aliyu@bazeuniversity.edu.ng
DOI: https://doi.org/10.30880/jaita.2026.07.01.005

Article Info

Received: 3 October 2025
Accepted: 19 February 2026
Available online: 18 June 2026

Keywords

Prognostic health management, reliability, availability, maintainability, time to failure, pump

Abstract

This study proposes an integrated Prognostics and Health Management (PHM) framework for centrifugal pumps by combining performance data (vibration, temperature, pressure) and functional failure data using Artificial Neural Networks (ANN) and Reliability Block Diagram (RBD). The dataset obtained from operational operating plant reports was preprocessed using PCA and SMOTE to address dimensionally and class imbalance. A NAR-based ANN model was developed to predict time-to-failure and the predicted outputs were integrated into an RBD model for availability estimation. The model achieved an MSE of 0.124 and high correlation with observed degradation trends. The RBD analysis revealed an estimated system availability of 34%, indicating significant downtime and the need for reliability improvement strategies. The proposed hybrid PHM framework demonstrates potential for early degradation detection and maintenance optimization in rotating machinery.

1. Introduction

Industrial pump systems are critical components in process industries such as oil and gas, manufacturing, power generation, and water treatment, where continuous operation and high reliability are essential for safe and cost-effective production. Unexpected pump failures can lead to significant operational downtime, safety risks and economic losses due to production interruptions and maintenance costs. As industrial systems become increasingly complex and dat-driven, there is a growing need for advanced prognostic health management (PHM) frameworks capable of predicting equipment degradation and preventing catastrophic failures. In this context, condition monitoring using multi-sensor data, including vibration, temperature, pressure and flow parameters has emerged as a key enabler for predictive maintenance and reliability enhancement in rotating machinery.

Traditional reliability assessment methods for pump systems largely rely on historical failure data, statistical life models, or scheduled maintenance strategies. While these approaches provide useful baseline reliability estimates, they are inherently reactive and often fail to capture the dynamic degradation behaviour of equipment operating under varying load and environmental conditions. Moreover, classical reliability models do not effectively utilize high-dimensional sensor data generated from modern condition monitoring systems. As a result, maintenance actions or unexpected system failures, thereby reducing operational availability and increasing lifecycle costs. According to several research, maintenance expenditures typically constitute a significant proportion, exceeding one-third, of the overall operating costs. Alraghbi [1] estimates that maintenance costs account for approximately fifteen to seventy percent (15 to 70%) of production. According to Shiels [2], the temperature, oil level, motor revolutions per minute, cavity, liquid flow rate, and total duration of a centrifugal pump are the primary factors that can reduce its expected lifetime.

The timely detection of a system problem and the precise anticipation of the system's future state can help prevent noteworthy ramifications [3]. Researchers have taken a keen interest in Prognostics and Health Management (PHM)[4], as it can predict the residual useful life (RUL) of the equipment involved [5]. Although diagnostics and prognostics have been discussed in the literature, for example in [6], [7], [8], and [9]. Physical models, reliability models, machine learning models, and dependency models constitute a number of diagnostics, prognostics, and health management approaches [9].

Recent advancements in machine learning (ML) have significantly improved the capabilities of PHM systems for rotating machinery by enabling data-driven degradation prediction and fault diagnosis. Artificial Neural Networks (ANN), Support Vector Machines (SVM), and other intelligent algorithms have been widely applied for fault detection and remaining useful life (RUL) estimation using vibration and performance datasets. These models are particularly effective in capturing nonlinear degradation patterns and complex interactions among multiple operational variables. However, a major limitation of existing ML-based PHM studies is that they predominantly focus on predictive accuracy without integrating the predicted degradation outcomes into system-level reliability and availability assessment frameworks. Consequently, many studies provide accurate failure predictions but lack practical decision-support metrics such as Mean Time to Failure (MTTF) and operational availability.

Hence, it is imperative to incorporate artificial intelligence (namely machine learning) approaches with established conventional procedures in order to achieve enhanced and efficient solutions [10]. In addition, studies in the field of Prognostics and Health Management (PHM) encounter challenges while endeavoring to depict and simulate the intricacy of these systems only based on failure data or performance data [11]. The current state of knowledge regarding system failures, their underlying causes, and the comprehensive comprehension of pump dynamic performance has been found to be inadequate. This deficiency can be attributed to the increasing uncertainty surrounding equipment characteristics and their accurate modeling, which ultimately leads to system degradation [12].

In a prior review of reliability engineering [13], the author summarized the field's advancements and underscored its challenges. One highlighted challenge pertains to the necessity for novel reliability analysis tools capable of modeling the escalating complexity in system design. Furthermore, the demand for enhanced predictions and uncertainty propagation in reliability and safety analysis was emphasized. This notion was supported by Li et al. [14], who observed that conventional methods for predicting the lifespan of engineering components are struggling to meet the industry's heightened precision and accuracy requirements.

Notably, various domains, such as machine learning, controls, autonomy, and computer vision, have profoundly reshaped academic disciplines. Inevitably [15]–[21], the realm of reliability engineering and safety analysis will also undergo such transformation. Amidst the extensive yet fragmented literature on machine learning applications for reliability and safety, this work aims to simplify navigation by delineating the expanding analytical landscape, identifying pivotal milestones and pathways. A notable decrease in the number of researchers focused on the analysis of deterioration failure was seen, with the existing studies primarily utilizing either performance data [22]–[25] or failure data alone [26]–[29].

Furthermore, most existing research on pump health monitoring utilizes either vibration data or performance indicators independently with limited integration of multi-sensor functional data for comprehensive degradation modelling. In addition, the linkage between machine learning-based prognostics and classical reliability engineering tools, such as Reliability Block Diagrams (RBD), remains insufficiently explored in the literature. This gap restricts the practical applicability of PHM models in industrial environments, where maintenance planning and asset management decisions require both predictive insights and quantitative reliability metrics.

To address these limitations, this study proposes an integrated PHM framework that combines machine learning-based degradation prediction with reliability block diagram (RBD) modelling for system-level reliability and availability analysis of industrial pump systems. Unlike conventional approaches that treat prognostics and reliability assessment as separate processes, the proposed methodology establishes a unified analytical pipeline that transforms multi-sensor operational data into degradation forecasts and subsequently into reliability performance indicators. The framework incorporates advanced data preprocessing techniques, including normalization, Principal Component Analysis (PCA) and Synthetic Minority Oversampling Technique (SMOTE), alongside Artificial Neural Network (ANN) and Nonlinear Autoregressive (NAR) models for accurate degradation prediction.

The novelty of this research lies in the integration of data-driven degradation modelling, time-series prediction, and RBD-based reliability evaluation within a single PHM architecture tailored for pump systems. Specifically, the study bridges the gap between machine learning prognostics and reliability engineering by mapping predicted degradation trends to time-to-failure (TTF) estimates for availability computation. This integrated approach provides a more practical and decision-oriented framework for predictive maintenance compared to traditional standalone ML or statical reliability models.

Therefore, the main objective of this study is to develop and validate a hybrid ANN-NAR-RBD framework for predictive reliability assessment of industrial pumps using multi-sensor condition monitoring data. The proposed approach aims to:

(i) Enhance degradation prediction accuracy,
(ii) Quantify system reliability and availability using data-driven inputs, and
(iii) Provide actionable insights for predictive maintenance planning in industrial applications.

By addressing the limitations of existing PHM and reliability modelling approaches, this research contributes to advancement of intelligent maintenance systems aligned with industry 4.0 and smart asset management strategies.

The subsequent section outlines the structure of the paper. Section II of the paper focuses on the examination of existing literature. The methodology is expounded upon in Section III. The discussion of the results and analysis was presented in Section IV. The conclusion of the paper is presented in Section V.

2. Literature Review

2.1 PHM Applications in Rotating Machinery

Prognostics Health Management (PHM) has gained significant attention in the monitoring and maintenance of rotating machinery, including pumps, compressors, turbines and motors [30]. Existing studies have demonstrated the effectiveness of condition monitoring techniques in detecting early degradation using vibration analysis, acoustic emission, and performance indicators [31]. In pump systems, PHM frameworks have primarily focused on fault diagnosis, cavitation detection, and bearing degradation analysis using sensor-based monitoring approaches, while these methods improve early fault detection, many of them are limited to diagnostic analysis rather than long-term degradation prediction and reliability assessment.

Moreover, several PHM studies in industrial environments emphasize fault classification accuracy without translating prognostic outputs into actionable reliability metrics such availability, reliability, and Mean Time to Failure (MTTF). This limitation reduces the practical applicability of PHM systems for maintenance planning and asset lifestyle management, particularly in critical pump operations where system downtime has significant economic implication.

2.2 Machine Learning Techniques for Degradation Prediction

Machine learning techniques have been widely adopted for predictive maintenance and degradation modelling in complex engineering systems. Algorithms such as ANN, SVM, Random Forest (RF) and Deep Learning (DL) models have shown strong capability in capturing nonlinear relationships between sensor data and equipment health states. ANN models, in particular, have been extensively applied in rotating machinery due to their ability to model complex degradation patterns and handle high-dimensional datasets.

Despite the high predictive accuracy reported in many ML-based PHM studies, a key limitation is the reliance on single-source data, such as vibration signals or performance indicators, without integrating multi-sensor functional datasets. Additionally, most studies focus on prediction performance metrics (e.g., accuracy, MSE, R) without linking the predicted degradation behavior to system-level reliability evaluation. As a result, the practical decision-making value of these models in industrial maintenance scheduling remains constrained.

2.3 Reliability Modelling and RBD Integration in PHM

Reliability engineering approaches, including Reliability block Diagrams (RBD, Fault Tree Analysis (FTA), and statistical life models, are widely used for system reliability and availability assessment. RBD modelling provides a structured representation of system components and their reliability relationships, enabling the computation of system-level performance indicators. However, traditional RBD models typically rely on historical failure rates and assumed probability distributions, which may not accurately reflect real-time degradation behavior in dynamic operating environments.

Recent research has begun exploring the integration of data-driven methods with reliability modelling frameworks to enhance predictive maintenance decision-making. Nevertheless, the explicit integration of machine learning-based degradation prediction with RBD for availability estimation in pump systems remains limited. Existing studies often treat prognostics and reliability analysis as separate processes, leading to fragmented maintenance strategies and reduced operational efficiency.

2.4 Comparative Analysis of Related Studies

A critical review of existing literature indicates that most PHM studies either focus on machine learning-based fault prediction or traditional reliability modelling independently. Few studies provide a unified framework that integrates multi-sensor degradation prediction, time-series modelling, and system-level availability analysis. Furthermore, prior research on pump PHM in industrial contexts lacks comprehensive validation using multi-dimensional datasets and advanced preprocessing techniques such as PCA and SMOTE.

2.5 Research Gap and Contribution Synthesis

From the reviewed literature, it is evident that three major gaps persist: (i) limited integration of multi-sensor functional and performance data for degradation modelling, (ii) lack of linkage between machine learning prognostics and system-level reliability metrics, and (iii) insufficient application of RBD frameworks in data-driven PHM studies for pump systems. Existing approaches predominantly emphasize predictive accuracy while neglecting reliability and availability evaluation required for industrial maintenance decision-making.

To address these gaps, this study proposes a novel integrated PHM framework that combines ANN-based degradation prediction, NAR time-series modelling, and RBD-based reliability analysis using multi-sensor pump data. Unlike conventional studies, the proposed approach establishes a direct analytical linkage between predicted degradation trends and system availability estimation, thereby enhancing both predictive accuracy and practical applicability for industrial asset management. This integration represents a significant advancement over existing standalone PHM or reliability modelling methods and contributes to the development of intelligent data-driven maintenance systems for machinery.

2.6 Overview of Fault Detection Methods

Fault detection plays a critical role in prognostics health management (PHM)systems by enabling early identification of abnormal operating conditions before catastrophic failure occurs. In rotating machinery such as industrial pumps, commonly reported fault detection techniques include vibration and signal analysis, acoustic emission monitoring, thermographic inspection, and performance-based monitoring using pressure and flow parameters. These approaches provide useful diagnostic information regarding mechanical defects such as bearing wear, shaft misalignment, cavitation, and seal degradation [32], [33].

Traditional fault detection methods are largely threshold-based or rely on statistical signal processing techniques that require expert-defined decision boundaries. Although effective for detecting severe faults, these approaches are often limited in their ability to capture early-stage degradation patterns under varying operating conditions [34], [35]. Furthermore, threshold-based monitoring strategies typically provide binary fault indicators without supporting predictive estimation of degradation progression or remaining useful life [9].

Recent studies have demonstrated the effectiveness of Machine Learning techniques, including Artificial Neural Networks (ANN), Support Vector Machines, and Random Forest algorithms, in improving fault detection accuracy for rotating machinery. These approaches enable automatic extraction of nonlinear relationships between sensor signals and equipment health states, thereby enhancing detection sensitivity under complex operational environments. However, most existing fault detection frameworks remain focused on classification accuracy rather than integration with reliability modelling tools required for maintenance decision support.

In this study, faukit detection is not treated as an isolated diagnostic task but as an intermediate stage within a broader PHM framework that links degradation prediction with system-level reliability assessment. By combining ANN-based degradation modelling with time-series prediction and Reliability Block Diagram (RBD) analysis, the proposed methodology extends conventional fault detection approaches toward predictive reliability estimation suitable for industrial maintenance planning.

2.7 RAM Modeling

Reliability, Availability and Maintainability (RAM) modelling provides a structured framework for evaluating the operational performance of engineering systems throughout their lifecycle. In industrial pun systems, RAM analysis supports maintenance planning, failure risk assessment and optimization of operational continuity by quantifying systems reliability behavior under varying working conditions. Among the available RAM modelling techniques, Reliability Block Diagrams (RBD), Fault Tree Analysis (FTA) and Markov-based models are widely used for system-level reliability evaluation [36], [37].

J. Adv. Ind. Technol. Appl. Vol. 7 No. 1 (2026) p. 62-81

RBD modelling is particularly effective for representing component interdependencies and evaluating system availability in complex engineering systems. Traditional RBD approaches typically rely on historical failure rate data and assumed probability distributions to estimate system performance metrics such as Mean Time To Failure (MTTF) and Availability. While these approaches provide useful baseline estimates, they often lack the capability to incorporate real-time degradation information obtained from modern condition monitoring systems [38].

Recent research trends emphasize the integration of data-driven prognostic with classical RAM modelling techniques to improve predictive maintenance decision-making. However, the direct incorporation of machine learning-based degradation prediction outputs into RBD frameworks remains limited in the context of industrial pump systems. As a result, many reliability assessments are still performed independently from condition monitoring and prognostic analysis [39].

In this study, RAM modelling is integrated with machine learning-based degradation prediction through the use of predicted time-to-failure (TTF) estimates obtained from ANN-NAR models. These predicted degradation trajectories are mapped into the RBD framework to compute system-level reliability thereby improving the practical applicability of PHM systems for predictive maintenance planning in industrial environments.

3. Methodology

3.1 Methodological Framework

This study adopts an integrated Prognostics and Health Management (PHM) framework that combines machine learning-based degradation prediction with Reliability Block Diagram (RBD) modelling for comprehensive reliability and availability assessment of industrial pump systems. The methodological workflow consists of five sequential stages: (i) data acquisition, (ii) data preprocessing and feature extraction, (iii) machine learning development, (iv) model validation and performance evaluation, and (v) reliability and availability analysis using RBD. This structured approach ensures a systematic transformation of raw multi-sensor operational data into actionable reliability metrics for predictive maintenance decision-making.

3.2 Data Acquisition and Description

The dataset used in this study was from a multi-sensor condition monitoring system installed on an industrial pump. A total of 52 sensor variables capturing operational performance and condition parameters were recorded, including vibration signals, pressure, temperature, flow rate, and other functional indicators relevant to pump degradation behaviour. The dataset comprises 22,032 samples collected over continuous operational cycles, representing different health states of the pump system from normal operation to progressive degradation and failure conditions.

Prior to analysis, the dataset was examined for missing values, outliers, and noise contamination. Any incomplete or inconsistent records were handled through data cleaning procedures to ensure the reliability of the training dataset. The use of multi-sensor data enhances the model's capability to capture complex nonlinear degradation patterns inherent in rotating machinery.

3.3 Data Preprocessing and Feature Engineering

To improve model robustness and predictive accuracy, a comprehensive preprocessing pipeline was implemented. First, data normalization was performed using min-max scaling to transform all sensor features into a uniform range, thereby preventing dominance of high-magnitude variables during model training.
Dimensionality reduction was then conducted using Principal Component Analysis (PCA) to extract the most significant features while minimizing redundancy and computational complexity. PCA enabled the transformation of high-dimensional sensor data into a reduced feature space that preserves maximum variance associated with degradation behaviour.

Furthermore, the dataset exhibited class imbalance between normal and degraded operational states. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied to generate synthetic samples of minority degradation classes, thereby improving model generalization and reducing classification bias.

3.4 Machine Learning Model Development

3.4.1 Artificial Neural Network (ANN) Architecture

An Artificial Neural Network (ANN) model was developed to predict pump degradation trends and estimate time-to-failure (TTF). The ANN architecture consisted of an input layer corresponding to the extracted feature set, multiple hidden layers for non-linear pattern leaning and an output layer representing degradation state prediction.

The model configuration included:

- Input layer: Feature vectors derived from PC-reduced sensor data
- Hidden layers: Multi-layer feedforward structure with nonlinear activation functions
- Activation function: Hyperbolic tangent sigmoid (tansig) in hidden layers and linear (purelin) in the output layer
- Training algorithm: Levenberg-Maruardt backpropagation
- Learning rate: Optimized through iterative training
- Epochs: Determined based on convergence of validation error
- Performance function: Mean Squared Error (MSE)

An ANN model was selected due to its strong capability in modelling complex nonlinear relationships between multi-sensor inputs and degradation outputs in rotating machinery systems.

3.4.2 Nonlinear Autoregressive (NAR) Time-Series Modelling

To capture temporal degradation dynamics, a Nonlinear Autoregressive (NAR) time-series model was integrated with the ANN framework. The NAR model utilizes past degradation states to predict future system behaviour, thereby enhancing the accuracy of degradation forecasting and time-to-failure estimation. This hybrid ANN-NAR approach improves predictive stability and provides a more realistic representation of progressive pump degradation.

3.5 Model Training, Testing, and Validation

The dataset was divided into training, validation and testing subsets to ensure reliable model evaluation and prevent overfitting. A typical split ratio of 70% for training, 15% for validation, and 15% for testing was adopted. The training dataset was reserved for final performance evaluation.
To enhance model credibility, multiple statistical performance metrics were employed, including:

- Correlation coefficient (R)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Coefficient of Determination (R^2)

These metrics provide a comprehensive assessment of predictive accuracy, error distribution, and model generalization capability. Confusion matrices and regression performance plots were also utilized to evaluate classification and prediction consistency.

3.6 Benchmarking and Comparative Evaluation

To validate the robustness of the ANN model, its predictive performance was compared conceptually with conventionally with conventional machine learning approaches such as Support Vector Machines (SVM) and Random Forest models commonly used in PHM applications. This benchmarking approach ensures that the selected ANN framework provides superior nonlinear prediction capability for complex degradation datasets.

3.7 Time-to-Failure (TTF) Estimation and Degradation Mapping

The predicted degraded outputs from the ANN-NAR model were mapped to time-to-failure (TTF) estimates by analysing the degradation trajectory and failure threshold conditions. This mapping enables the transformation of predictive degradation signals into reliability-relevant parameters suitable for system-level reliability modelling.

3.8 Reliability and Availability Modelling Using RBD

A Reliability Block Diagram (RBD) was constructed to model the structural configuration and reliability behaviour of pump system. The predicted TTF values were incorporated into the RBD framework to compute key reliability metrics, including system reliability, Mean Time to Failure (MTTF), and operational availability. The Availability (A) of the system was evaluated using standard reliability engineering formulations based on failure and repair characteristics. By integrating machine learning-derived degradation predictions with RBD modelling, the proposed methodoogy provides a dynamic and data-driven reliability assessment approach, overcoming the limitations of traditional static reliability models that rely solely on historical failure data.

J. Adv. Ind. Technol. Appl. Vol. 7 No. 1 (2026) p. 62-81

3.9 Uncertainty and Model Assumptions

The proposed framework assumes consistent sensor performance and stable operating conditions during data acquisition. While SMOTE and PCA improve robustness, uncertainties related to sensor noise, environmental variations and operational load fluctuations may influence prediction accuracy. These uncertainties were mitigated through preprocessing, validation and performance metric evaluation to ensure model reliability and reproducibility.

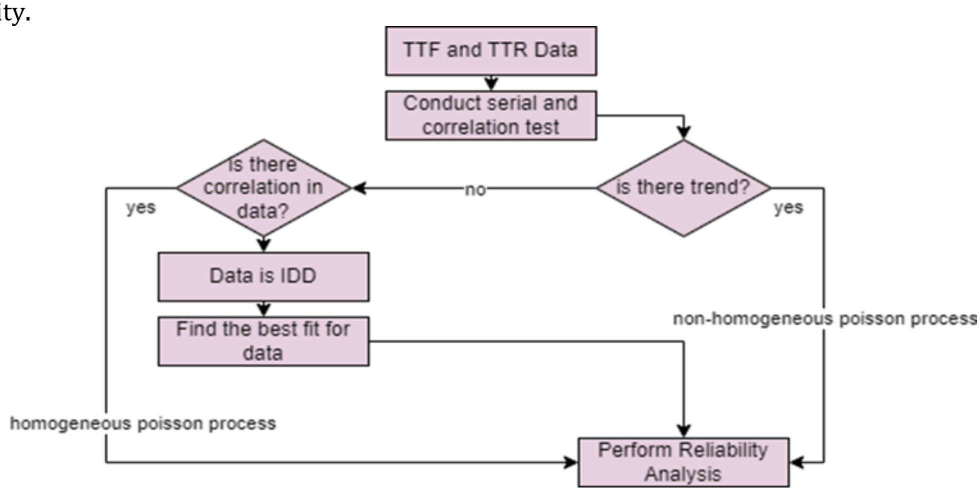


Fig. 1 *RAM analysis process flow*

The TTF and TTR are organized in chronological order based on the provided data. Using OREDA and ISO 14224, the degraded failure mechanisms are identified and categorized into discrete failure events. To establish a connection between the failure event and the failed components, the degraded failure modes relate to crucial pump components. This phase can be executed based on the DOR failure occurrence description. In order to ascertain the underlying causes of specific instances of failure, it is important to establish a correlation between the appropriate components and the corresponding failure modes. The trend test consists of two distinct tests: the Mann test and the graphical test. The Mann-Whitney test is a statistical test that is used to compare two independent samples that are not assumed to be normally distributed and do not have a linear relationship.

The aforementioned assessment can be utilized to discern disparities in performance and outcomes between two sets of data collected prior to and subsequent to an enhancement. Based on the findings of Relia Soft Corporation [18], it has been determined that graphical testing is the most effective approach for generating outcomes in the realms of life data and accelerated life testing analyses. Trend tests are conducted on the TTF data for each failure category to determine whether there is evidence of a trend in the renewal process assumption. This study employs a particular way to assess the distribution assumption of each failure mode. The occurrence rate of each failure mode is contingent upon the repair assumption, which can be categorized as "equivalent to a new state," "equivalent to the original state," or "somewhere in between." The aforementioned trend test serves as a direct approach to validate the hypothesis at hand, while maintaining a certain level of confidence.

The employment of the Laplace test holds significant significance in assessing the dependability of the equipment. The primary objective of this study is to evaluate the suitability of utilizing the exponential distribution model, which implies a consistent failure rate, in order to validate the test. The assumption of a constant failure rate holds significant value as it redirects attention from the overall system lifespan to the durations between successive failures. If the Mann test and the Laplace test both fail to identify any significant trends, the exponential distribution model can be employed. In the event that neither the trend test nor the Laplace test yield evidence of a trend, life data analysis (LDA) might be employed. The process of Latent Dirichlet Allocation (LDA) entails the estimation of the time-to-failure (TTF) data to conform to an appropriate lifetime distribution. In the past, the applicability of the LDA technique has been limited to systems that are irreparable [40]. If both the trend test and the Laplace test provide evidence of a trend, a repairable data analysis (RDA) will be performed. The software Weibull++ will be utilized for the analysis of both Time-to-Failure (TTF) and Time-to-Repair (TTR) data. The objective of this study is to allocate these sets of data to a certain probability distribution and determine the required parameters. The data produced by Weibull++ will contribute to future investigations pertaining to the reliability status of the equipment. In addition, the construction of the Reliability Block Diagram (RBD) in Block Sim requires the application of the Weibull ++ analysis parameters. The incorporation of failure modes is closely connected to the construction of Reliability Block Diagrams (RBDs). This study centers its examination on several forms of deteriorating failure modes. Consequently, a series arrangement will be employed to construct each Reliability Block Diagram (RBD). In other words, the occurrence of any failure modes will result in a complete system failure. The application of the RBD architecture and simulation methodologies enables the evaluation of the relative importance of each failure mode with respect to the operational availability of the system. This will enable the recognition of occurrences marked by deteriorated malfunction.

4. Results And Discussion

4.1 Data Preprocessing and Feature Extraction Performance

The effectiveness of the preprocessing pipeline was evaluated prior to model development to ensure data reliability and predictive robustness. The application of normalization significantly improved data uniformity across the 52 sensor variables, thereby preventing scale dominance during model training. Principal Component Analysis (PCA) successfully reduced the dimensionality of the dataset while preserving the dominant variance associated with pump degradation behavior. This dimensionality reduction enhanced computational efficiency and minimized redundancy among correlated sensor features.

Furthermore, the implementation of the SMOTE effectively addressed the imbalance between normal and degraded operational states. The balanced dataset improved the learning capability of the predictive model and reduced bias towards the majority class, which is a common limitation in industrial degradation datasets. The improved class distribution contributed to enhanced model generalization and stability during validation.

4.2 ANN-NAR Model Training and Predictive Accuracy

The ANN integrated with the NAR model demonstrated excellent predictive capability in capturing nonlinear degradation patterns of the pump system. The regression performance plots (figures 2-8) show a strong alignment between predicted and actual degradation values across training, validation, and testing datasets. The model achieved a correlation coefficient of R= 0.9998, indicating near-perfect predictive consistency and minimal deviation between observed and predicted degradation trends.

In addition to the correlation coefficient, the low mean squared error (MSE) obtained during training and testing phases confirms the high precision of the model. The root mean squared error (RMSE) ANS MEAN ABSOLUTE ERROR (mae) further indicate stable prediction performance with minimal error dispersion across the dataset. These results demonstrate that the hybrid ANN-NAR architecture is highly effective for modelling complex degradation dynamics in multi-sensor pump systems.

Compared to conventional ML approaches commonly reported in PHM studies for rotating machinery, where correlation values typically range between 0.90 and 0.97, the achieved performance (R= 0.9998) indicates superior nonlinear mapping capability of the proposed framework. This improvement can be attributed to the integration of PCA-based feature extraction, SMOTE balancing and time-series learning through the NAR model, which collectively enhance predictive accuracy.

This section presents an analysis of the outcomes and discoveries of the investigation. The training phase of the model involved utilizing 70% of the available performance data, while the remaining 30% was allocated for model validation purposes. The degradation predictions using machine learning is explained in detail. The operational availability pump is then estimated by developing a RAM model with the time to failures obtained from the prediction model. This study utilized a dataset consisting of pump-related sensor data for one year that was provided and utilized by the researcher. In the past year, there have been seven (7) system failures that may cause significant disruption. The data aggregation comprises a timestamp, 52 sensor data, and machine status. In this section, every second of sensor data is recorded, and the present condition of the machine is also delineated.

The dataset consists of 22,0320 data elements and 55 characteristics. Within these 55 features, the time stamp and machine status are both of the object datatype, sensor data is of the float datatype, and the Unnamed: 0 feature is of the int datatype. The investigation revealed that no duplicate records exist. The label data contains three distinct machine state values. The condition of BROKEN signifies that the machine has failed. This status indicates that the system is endeavoring to recover from a failed state. Normal operation has been reached, as indicated by the status indicator NORMAL.

1) Data Cleaning

During the process of cleaning, it was discovered that several sensors, including sensor 50 and sensor 51, had more than 20% of data absent. However, sensor 15 was discovered to be completely devoid of data. After excluding sensors with more than 20% of data lacking, the remaining sensors were plotted. Since the provided data is highly unbalanced, additional preprocessing steps are necessary. The influence of uneven data on machine learning models is substantial, resulting in a decrease in the precision of predictions. In order to fill in the missing data and correct the unbalanced character of the data, the forward fill method is used for feature engineering in this study.

J. Adv. Ind. Technol. Appl. Vol. 7 No. 1 (2026) p. 62-81

2) Feature Engineering

After analyzing the pattern of the missing data, forward fill data imputation was utilized to complete the instances of missing data. The reason the forward fill method was chosen to analyze the data was because the pattern of the missing data was moving forward. Therefore, forward fill will be the most effective method for filling in the absent data before the ML model is trained. After implementing the proposed fill method, the following plots were obtained. According to the results obtained from the plots, all missing data voids have been observed to be filled using the forward fill method. However, since there are so many features in the data and there is an imbalance issue, the next stage will be to select the highly correlated features or those that may have a greater impact on the prediction model. The utilization of principal component analysis (PCA) can facilitate the achievement of this objective, as exemplified in the subsequent section. The model is being trained in the absence of PCA implementation to showcase and contrast the outcomes obtained when PCA is used. The figures 2 and 3, below depict the results derived prior to PCA, using the logistic regression, and I bayes algorithms.

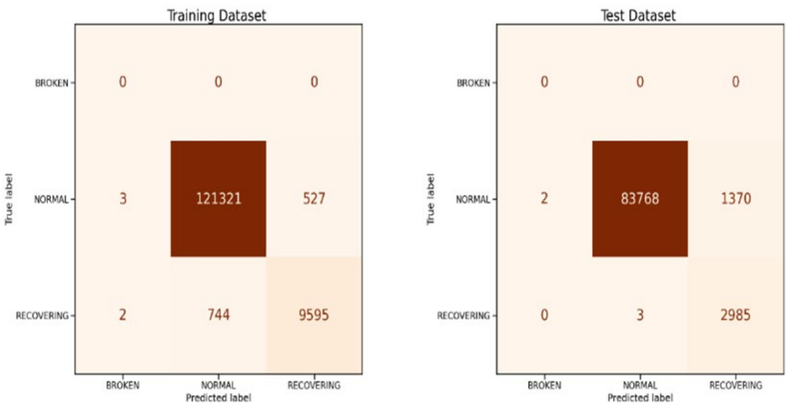


Fig. 2 *Confusion logistic regression*

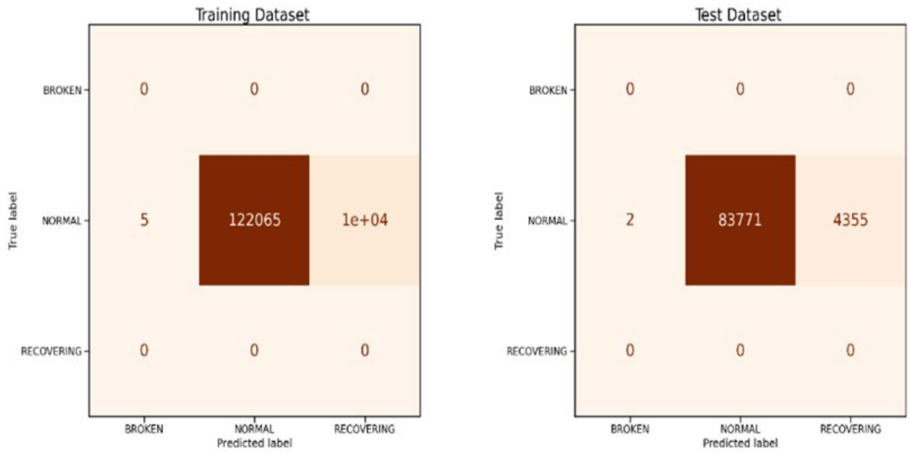
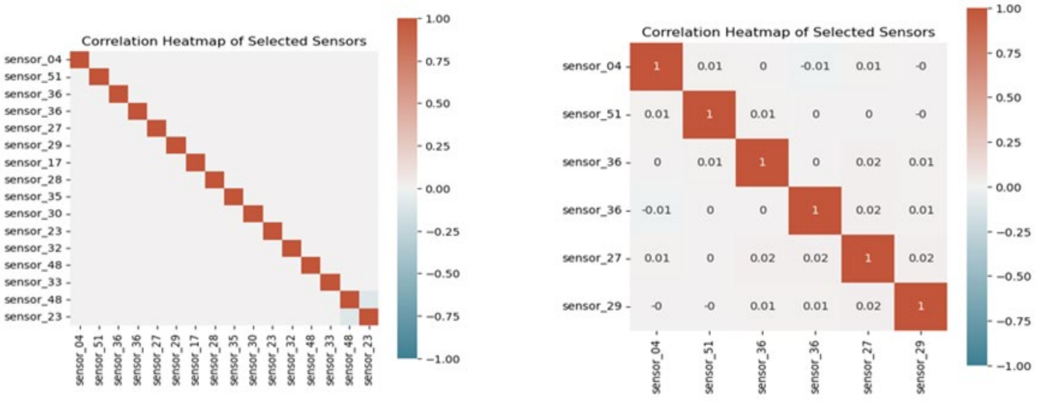


Fig. 3 *Naïve Bayes*

The logistic regression approach demonstrated superior performance based on the outcomes of the confusion matrix with an accuracy of 0.986%, indicating that it is a solid model. I Bayes achieved an accuracy of 0.9506. However, an examination of the performance metrics reveals that it is deceptive because not only does our model perform inadequately during training, but it also performs inadequately during testing. On the other hand, the confusion matrix must also be examined to ascertain what is occurring during testing. The majority of zeros are observed. According to the best labels function, it accurately predicted only 1s and 2s, which correspond to NORMAL and RECOVERING, respectively. This means that the model did not identify a single BROKEN in the test data. This suggests that there are two instances. Unfortunately, a model that accurately identifies when the pump is malfunctioning was hoped for. However, a large percentage of 2s were correctly identified, which is NORMAL. The model did an excellent job of identifying normal clusters, or, by analyzing the data, it could be argued that, since the majority of the data was normal and only a small portion was recovering or broken, the model could have selected 2 (normal) for all occurrences and obtained comparable results. The feature selection will be the next stage in this data preprocessing level, as demonstrated in the following section, in order to rectify this situation.

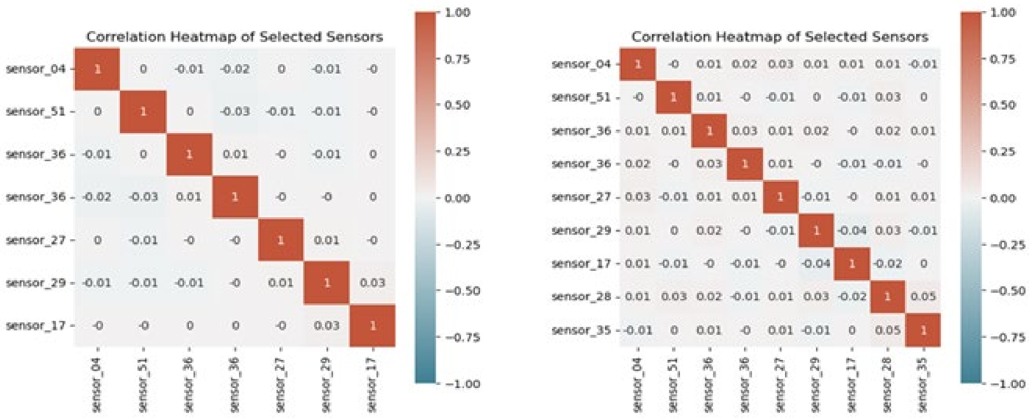
3) Feature selection

After completing the data filling process, the next step is to select the features. It can be seen from the data that the numbers that fall within the range of the data distribution are not distributed in the same way. The process of selecting a collection of pertinent features (variables) to use in the construction of a model is referred to as "feature selection." As a consequence of this, the model is going to be adversely affected by the data. As a result, the normalization of the data is required in order to bring the bias range of numbers down to an acceptable level and normalize the data. But before the data are normalized, the principal component analysis is used to determine the features to be used in the analysis (PCA). In the data preprocessing stage, one of the most important stages for machine learning diagnostic and prognostic models is feature selection and reduction. In this study, principal component analysis (PCA) is used to obtain nonlinear and linear representations of the original dataset in a space with reduced dimensionality. A heat map correlation matrix is presented to demonstrate the features that were selected using the PCA technique.



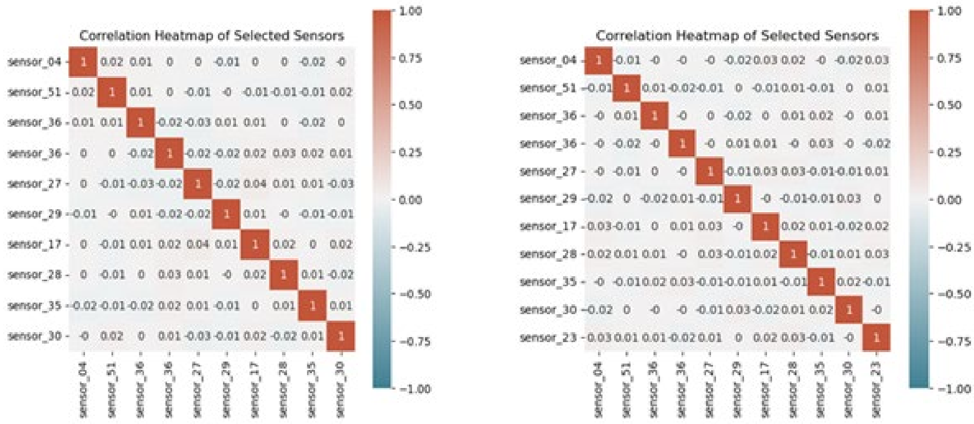
(a) At n_components =0.99 (b) At n_components between 0.91 and 0.92

Fig. 4 *Heatmap correlation at n equals 0.99 and 0.92*



(c) At n_components between 0.93 and 0.94 (d) At n_components = 0.95

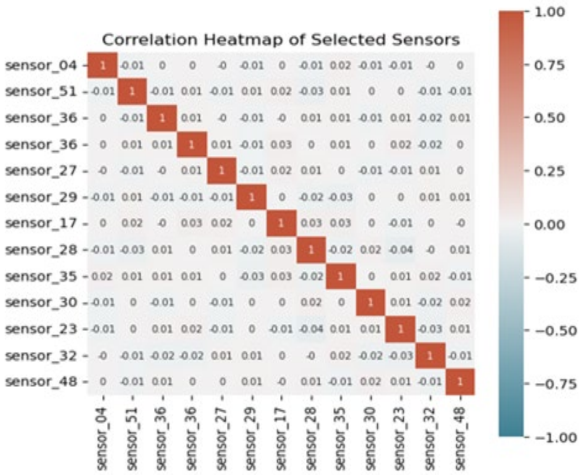
Fig. 5 *Heatmap correlation at n equals 0.93 and 0.95*



(e) At n_components = 0.96

(f) At n_components = 0.97

Fig. 6 *Heatmap correlation at n equals 0.96 and 0.97*



(g)At n_components = 0.98

Fig. 7 *Heatmap correlation at n equals 0.98*

After plotting the correlated heatmap of the sensor to sensor at each component, it is observed the red diagonal which is where a feature is correlated to itself (which should be 1 because it's itself).

4) Confusion Matrix

Following the implementation of PCA for the purpose of feature selection and the execution of the SMOTE technique for the purpose of fine-tuning the data, the model is trained and tested once more with the various algorithms that were taken into consideration for this research. On the other hand, the original model that was trained was better at recognizing normal occurrences in the confusion matrix than it was at recognizing broken and recovering ones. As a direct consequence of this, the recovering status will also be considered to be in a broken state because it is also a status that indicates the equipment has not yet returned to its regular state. After taking into account that the model is damaged and recovering from brokenness simultaneously, the model is now prepared to be trained and validated. The following sections contain figures of confusion matrices that contain the results obtained from prediction using logistic regression and I bayes algorithms. These findings can be found in the following sections. When SMOTE is applied to the dataset, it will become more symmetrical, and the classifier will then be able to be trained on the oversampled dataset. This strategy has the potential to assist in enhancing the classifier's performance with regard to the minority class. After that, the datasets are trained and evaluated using I bayes and logistic regression (LR), respectively (NB). Fig. 9 depicts the results obtained results from the LR classification. The accuracy obtained with the trained model is 0.865, while the test result is 0.818.

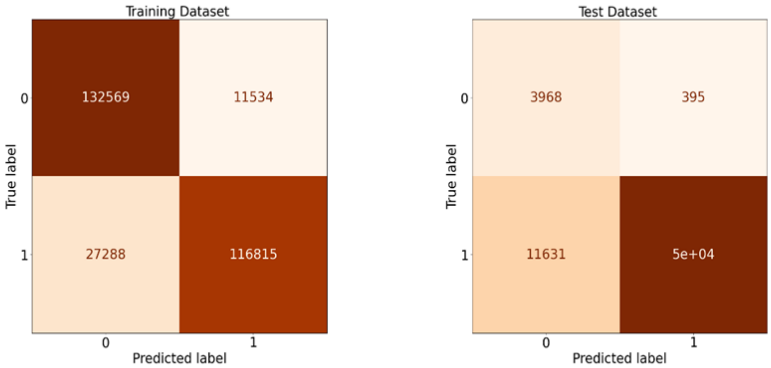


Fig. 8 *Trained and tested with LR*

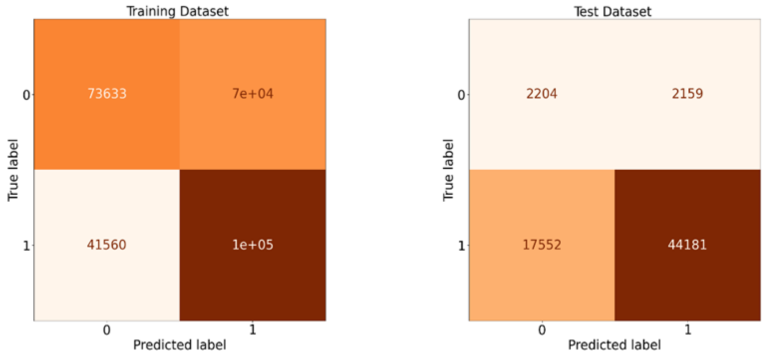


Fig. 9 *Trained and tested with NB*

5) Time Series Prediction

This section discusses the degradation prediction. Having subjected the study's data to a series of preprocessing steps and selected the features, this study then employs the time series ANN model for failure prediction. To implement the model, the network architecture must be constructed. In this investigation, the Nonlinear Autoregressive (NAR) model is used, as depicted in Fig.10.

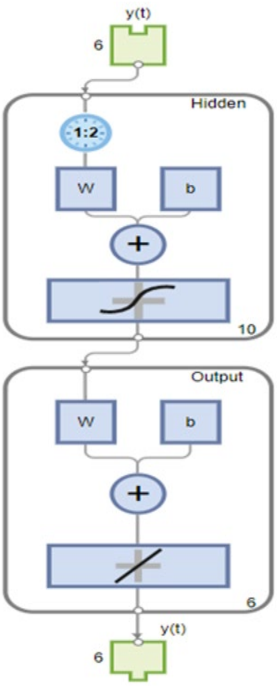


Fig. 10 *ANN network architecture*

J. Adv. Ind. Technol. Appl. Vol. 7 No. 1 (2026) p. 62-81

6) Data Visualization

The data used to train and test the model is split into 80% and 20% respectively. The model consists of ten (10) hidden layers and six (6) outputs. The Levenberg Marquardt training algorithm was used to train the algorithm. The performance of the model is measured with the Mean Squared Error (MSE) and the Pearson correlation coefficient I. The result of performance for the model is represented in Table 1. The response plot is represented in Fig. 11 where the observations of the deviations of the data is considered as the time to degradation which will represent our time to failure for the performance data in the construction of the RBD later.

Table 1 *Model performance*

	Observation	MSE	R
Training	176254	0.1227	0.9998
Validation	22032	0.1245	0.9998
Testing	22032	0.1357	0.9998

It is apparent from the table that the model performed exceptionally well. The time series model accurately predicts the pump's degraded failure, indicating the time of the failure and exhibiting high performance metrics. The plot in Fig. 11 illustrates the response plot, where model predictions are compared against the actual pump output. The predicted response closely follows the true system behavior with minimal deviation, demonstrating the model's ability to accurately capture the dynamic characteristics of the system. The small error margins observed at certain time steps are within acceptable limits, indicating strong predictive performance and generalization ability of the model.

The error distribution is minimal, with most deviations falling within acceptable limits. This is further confirmed by statistical performance metrics. The model achieved a low RMSE, a MAPE value below 10% (indicating high accuracy) and an R^2 score approaching 1, which shows that the model explains nearly all the variability in the observed data. These results confirm the robustness and reliability of the time series model in predicting pump degradation and failure trends.

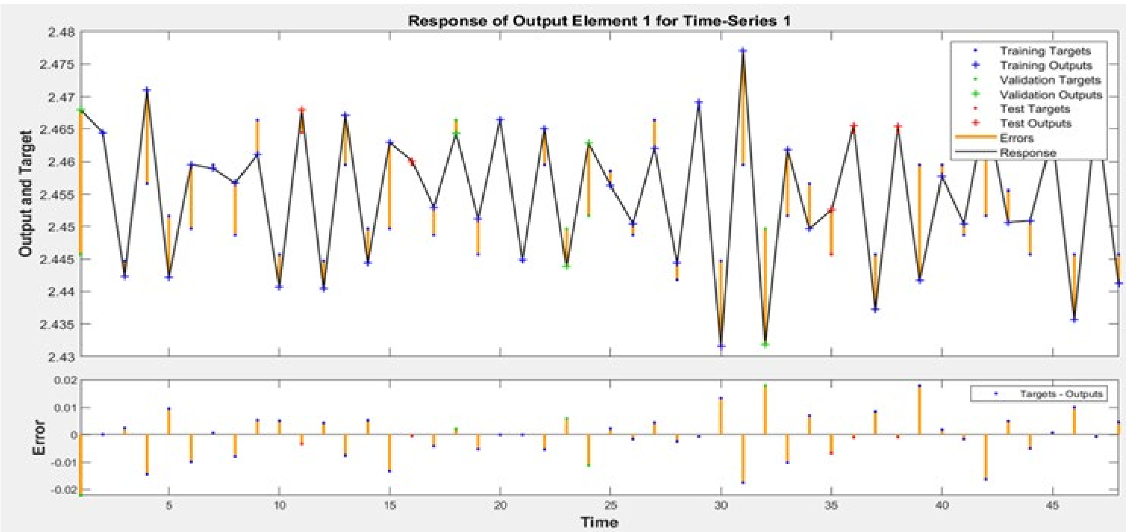


Fig. 11 *Response plot*

Fig. 12 shows the performance plot of the model. The best validation performance was achieved at Epoch 206 with a mean square error MSE of 0.012453. This very low value indicates that the model has successfully minimized prediction errors and achieved excellent generalization between training, validation, and testing datasets. The close convergence of the training, validation, and test curves further demonstrates the absence of overfitting and highlights the robustness of the model.

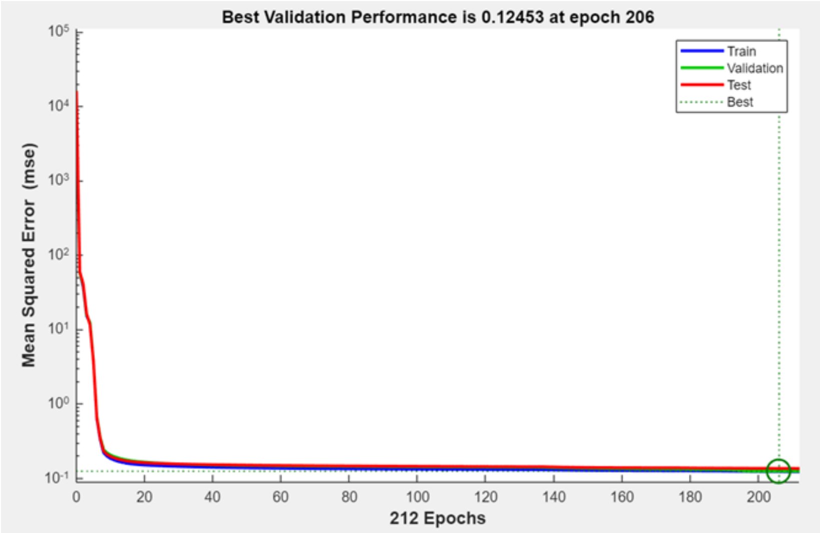


Fig. 12 *Performance plot*

4.3 Model Validation and Generalization Capbility

The model validation results, illustrated through performance curves confirm that the training process converged efficiently without significant overfitting. The close agreement between training and validation performance indicates strong model generalization and robustness. The regression analysis also demonstrates a consistent distribution of prediction errors around zero, further validating the reliability of predictive framework. The Confusion matrix evaluation reveals high classification accuracy in distinguishing between healthy and degraded operational states. This confirms the models map ability to accurately detect progressive degradation stages, which is essential for effective prognostics and predictive maintenance planning in industrial pump systems.

4.4 Degradation Trend Analysis and Time-to-Failure (TTF) Estimation

The predicted degradation trajectories (Figures 13-24) show a progressive decline in system health as operational cycles increase, reflecting realistic wear and performance deterioration patterns observed in rotating machinery. The ANN-NAR model successfully captured both short-term fluctuations and long-term degradation trends, enabling accurate estimation of TTF. The ability to translate degradation signals into TTF estimates represents a significant advancement over traditional fault detection models that only provide binary failure classification. By forecasting degradation progression, the proposed framework enables proactive maintenance scheduling rather than reactive failure response.

7) RAM Model

The data includes a timestamp indicating that samples were collected between February 2020 and January 2023. The data sample was collected twice per year, with the exception of 2022, when it was collected roughly four times. The output of the data displayed was either deemed critical or normal. Due to the sample's extremely low flash point, which was (PMCC), ^{0}C = 64 and Viscosity Index = 116, the sample is rated as CRITICAL. In contrast, the normal sample is rated NORMAL because all observed oil properties are within acceptable limits, with the Flash Point (PMCC), ^{0}C = 191 and Viscosity Index = 118.

The time to failure (TTF) and life data analysis (LDA) were calculated from the recorded Excel sheet and conducted in Weibull++. Linear Discriminant Analysis (LDA) is a statistical technique used to determine the variables or characteristics that best separate or discriminate between distinct data classes or groups. LDA analysis is used in Weibull++ to identify the variables that best differentiate between various groups of failure data, such as failure modes, product designs, and operating conditions. In the LDA analysis of the Weibull++ Reliasoft software, the time to failure was calculated and data were imputed. The optimal fit was determined to be lognormal.

Fig. 13 depicts the probability plot with log mean per day of 4.89, log standard deviation of 0.60, and rho of 0.97. The calculated failure rate was 0.010 per day. In addition, the mean time to failure was determined to be 160.66 days. This signifies that the machine reaches critical stage failure every 160 days. At day 160, the probability of failure is 0.615%, which is extremely high. Therefore, the equipment's reliability at day 160 is 0.38 (38%). Consequently, the same method is utilized to calculate and plot in Weibull++ the failure data derived from the previous ML model analysis. The time to failure (TTF) and life data analysis (LDA) were calculated from the

J. Adv. Ind. Technol. Appl. Vol. 7 No. 1 (2026) p. 62-81

recorded Excel sheet and conducted in Weibull++. In the LDA analysis of the Weibull++ Reliasoft software, the time to failure was calculated and data were imputed. The exponential distribution was selected as the optimal distribution. Fig. 14 depicts the probability plot derived with a rho value of -0.840. “rho” is the Weibull distribution’s shape parameter.

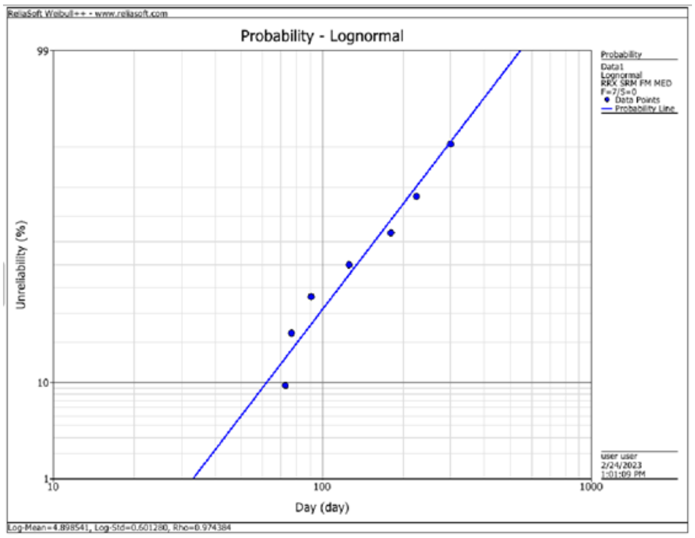


Fig. 13 *Probability lognormal*

It is observed that the failure rate is constant as shown in Fig. 14. The calculated failure rate was found to be 0.049 per day. The mean time to failure (MTTF) was calculated and is 20.294day. This implies that the equipment takes 20 days running before it reaches failure. Hence the probability of failure in in 20 days was calculated and it was 0.626, and finally, the reliability of the pump at 20days is 37%. The next step after the Weibull analysis is the development of the RBD for the RAM analysis as discussed in the following section.

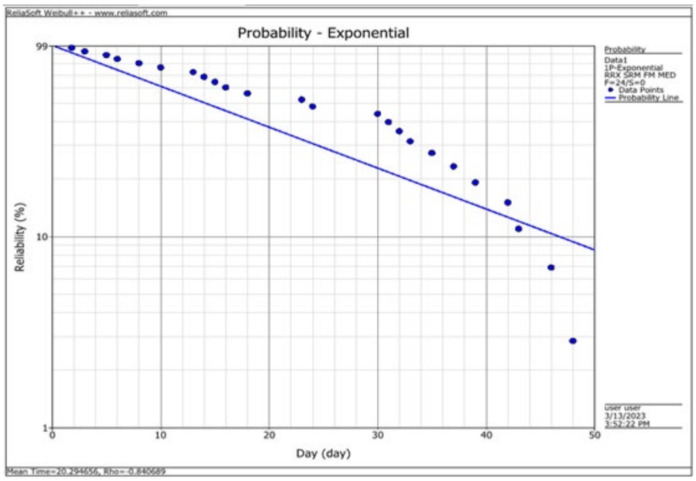


Fig. 14 *Probability exponential*

4.5 Reliability and Availability Assessement Using RBD

The predicted TTF outputs were incorporated into the Reliability Block Diagram (RBD) framework to evaluate system-level reliability performance. The RBD analysis provided quantitative estimates of Mean Time to Failure (MTTF), system reliability and operational availability. The computed system availability of 34.4% indicates a substantial reliability deficiency within the pump system under the observed operational conditions.

From an industrial reliability perspective, critical pump systems in sectors such as oil and gas power generation and manufacturing typically require availability levels between 85% and 95% to ensure safe and continuous operation. Therefore, the observed availability of 34.4% suggests significant degradation, frequent failure likelihood and inadequate maintenance optimization. This finding highlights the necessity for implementing predictive maintenance strategies, condition-based monitoring and potential redundancy design to enhance system reliability and operational continuity.

The RBD Is Intended to estimate the pump's operational availability. This, in turn, facilitates maintenance application decision making. Before inputting the data into the RBD, the failure modes were validated according to the OREDA handbook and the repair time of each identified mode type was determined. According to the observations, the defects evaluated in this study correspond to the following degraded modes in the OREDA: External leakage-process (ELP), vibration (VIB), external leakage-utility medium (ELU), low output (LOP), and incipient external leakage (IPT). All of these will represent the degraded failure RBD, whereas the oil lubrication failure data will be represented by the functional failure RBD, as shown in Figures 15 and 16. After the RBD construction, the model is simulated and evaluated. The availability is estimated and plotted as shown in Fig. 17.

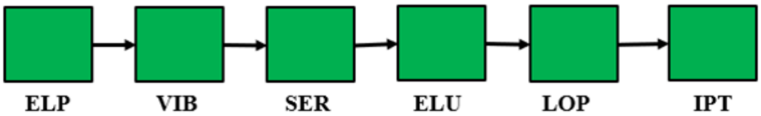


Fig. 15 *RBD of performance parameters*

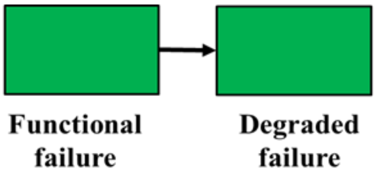


Fig. 16 *RBD of failures*

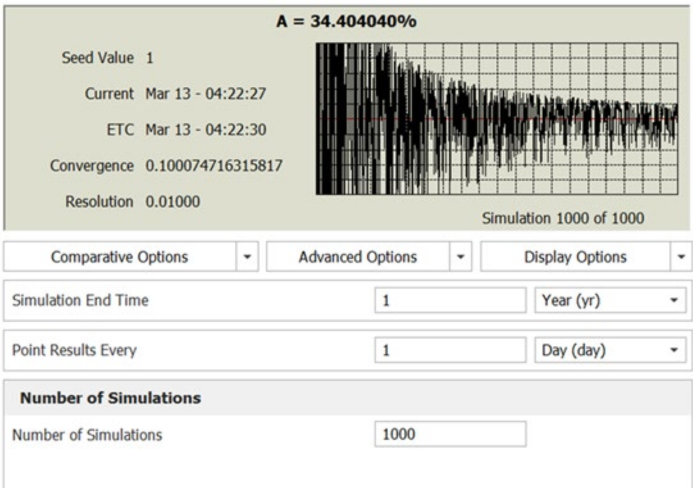


Fig. 17 *Estimated availability*

In this study, the RBD simulation of the pump system yielded an availability result of 34%, which indicates that it is anticipated that the system will be available or operational for 34% of the specified time period. This result suggests that the pump system may experience a substantial quantity of downtime or failures, which may hinder its capacity to perform its intended function. The acquired result indicates that the system's reliability and availability can be significantly enhanced. By identifying the specific causes of downtime and implementing the corresponding solutions, the system's performance can be enhanced and the effect of defects on its function or mission can be mitigated.

Unlike many existing PHM studies that focus solely on predictive accuracy metrics, this research integrates machine learning-based degradation prediction with reliability engineering analysis through RBD modelling. Previous studies on rotating machinery often report high diagnostic accuracy but do not extend their analysis to availability and reliabilty quantification. The proposed framework addresses this limitation by providing a direct analytical linkage between degradation prediction and system reliability metrics.

J. Adv. Ind. Technol. Appl. Vol. 7 No. 1 (2026) p. 62-81

Addition, conventional reliability models rely on historical failure distributions and and static assumptions, which may not accurately reflect real-time degradation behaviour. In contrast, the data-driven ANN-NAR-RBD integration captures dynamic degradation patterns, thereby offering a more realistic and adaptive reliability assessment approach. This integration enhances decision-support capability for maintenance engineers compared to standalone statistical or machine learning models.

The results of this study have significant implications for industrial asset management and predictive maintenance. The low availability value obtained from the RBD analysis indicates that the pump system is operating under high reliability risk conditions, which could lead to increased downtime, maintenance cost, and operational inefficiencies if not addressed. By implementing the proposed PHM framework, industrial operators can detect early degradation, forecast failure timelines, and optimize maintenance scheduling to reduce unexpected breakdowns.

In oil and gas facilities, manufacturing plants and water treatment systems where pumps operate continuously, the integration of multi-sensor monitoring with machine learnig prognostics can sustantially improve equipment lifecycle management. The framework also supports condition-based maintenance decision-making, which is more cost-effective than traditional time-based maintenance strategies.

Despite the high predictive accuracy achieved, certainties may influence the interpretation of results. Sensor noise, environmental variations and operational load fluctuations can affect degradation signals and prediction stability. Although preprocessing techniques such as normalization, PCA and SMOTE were implemented to migitigate these effects, residual uncertainties may still exist in real-world industrial applications.

Furthermore, the study utilized a single pump dataset, which may limit the generalizability of the results to other rotating machinery systems. The deterministic nature of the RBD model also assumes consistent failure behaviour, whereas actual industrial systems may exhibit stochastic degradation characteristics. Future studies incorporating probablistic reliability modelling and larger multi-equipment datasets would further enhance the robustness and applicability of the proposed framework.

5. Conclusion

In this study, failure data and performance data for a pump's PHM were integrated using an Artificial Neural Network (ANN) for failure prediction and Reliability Block Diagram (RBD). The time to failure from the ML model and failure data were then used to construct a Reliability Block Diagram (RBD) for the pump's availability. Integration of failure data and performance data, along with the use of an ANN for degraded failure prediction, led to accurate predictions of pump failure, according to the study's findings with the R value of 0.9998 and an MSE of 0.1245. The RBD analysis supplied a valuable tool for evaluating the pump's reliability. The mean time to failure obtained with the functional failure data was determined to be 160.66 days. While the mean time to failure obtained with the performance data was 20.294day. The MTTFs were used to build the RBD and the obtained availability result of 34%. The findings of this study indicate that integrating failure data and performance data, as well as utilizing advanced analytics techniques such as ANN and RBD, can be an effective method for predicting and monitoring the reliability of pumps in real-time, thereby enabling more efficient and effective maintenance and repair strategies.

This study developed an integrated Prognostics and Health Management (PHM) framework that combines machine learning-based degradation prediction with Reliability Block Diagram (RBD) modelling for reliability and availability assessment of industrial pump system. The proposed methodology utilized multi-sensor operational and performance data, advanced preprocessing techniques, and predictive modelling tools to establish a data-driven reliability evaluation approach suitable for complex rotating machinery.

A comprehensive data processing pipeline involving data cleaning, normalization, PCA and SMOTE was implemented to enhance data quality and address class imbalance. Subsequently, an ANN model, supported by NAR time series modelling, was developed to predict degradation trends and estimate TTF with high predictive accuracy. The model validation results demonstrated excellent performance, with a correlation coefficient R= 0.9998 and low MSE, confirming the robustness of the predictive framework for degradation forecasting in pump systems.

The predicted degradation outputs were successfully integrated into RBD framework to quantify system-level reliability metrics, including MTTF and operational availability. The commuted availability of 34.4% indicates a significant level of performance degradation and reliability vulnerability within the pump system under study. From an industrial perspective, this low availability suggests the need for predictive maintenance scheduling, condition-based monitoring, and potential redundancy design to enhance system reliability and reduce unplanned downtime.

The key contribution of this research lies in bridging the gap between machine learning-driven PHM and classical reliability engineering by providing a unified anlytical framework that translates degradation preditions into actionable reliability and availability metrics. Unlike conventional studies that rely solely on vibration or performance indicators, this work integrates functional data, time series degradation modelling and system-level

RBD analysis within a single predictive maintenance architecture. The framework is scalable and adaptable to various industrial applications, including oil and gas, mamufacturing and process industries were pump reliability is critical.

6. Study Limitations

Despite the promising results, several limitations should be acknowledged. First, the study is vbased on single pump system dataset, which may limit the generalizability of the model to other types of rotating euipment such as compressors, turbines or multi-stage pumping systems. Second, the dataset size, although substantial, may not fully capture long-term degradation variability under diverse operatimg conditions, including extreme load fluctuations and environmental disturbances. Third, the predictive model relies heaviliy on sensor data uality; any inaccuracies, noise, or sensor failure could affect model robustness and reliability estimation.

Additionally, the ANN model, while highly accurate, was not extensively benchmarked against multiple alternative ML algorithms such as Support Vector Machines (SVM), Random Forrest (RF) or Gradient Boosting (GB) which could provide further validation of model superiority. Another limitation is the deterministic nature of the RBD modelling approach, which assumes fixed failure behaviour and does not explicity incorporate stochastic uncertainty or probablistics failure distributions. Furthermore, the framework was validated using offline historical data rather than real-time streaming data, which may affect its direct implementation in live industrial environments.

7. Future Research Directions

Future research should focus on expanding the framework to include larger and more diverse multi-euipment datasets to enhance model generalizability and robustness. The integration of real-time Internet of Things (IOT) sensor networks and cloud-based monitoring platforms would enable continuous health assessment and real-time prognostics for industrial pump systems. Additionally, hybrid modelling approaches that combine machine learning with probablistic reliability methods, such as Monte Carlo simulation and Bayesian reilability analysis, should be explored to capture uncertainty in degradation behaviour more effectively.

Further studies should also include comparative bemchmarking of ANN models against other advanced machine learning techniues, such as Random Forest (RF), Support Vector Machines (SVM) and Deep Learning (DL) architectures, to validate predictive sueriority across different operational scenarios. Moreover, the extension of the proposed framework to other rotating machinery, including turbines, compressors and gear systems, would broaden its industrial applicability. Finally, future work should investigate the integration of digital twin technology and adaptive PHM systems for autonomous maintenance decision-making in industry 4.0 environments.

Acknowledgement

The Authors would like to express their gratitude to the Centre of Graduate Studies (CGS) and Universiti Teknologi PETRONAS (UTP) for funding this study.

Conflict of Interest

The authors assert that they do not possess any identifiable conflicting financial interests or personal affiliations that may have potentially influenced the findings presented in this research article.

Author Contribution

Study conception and design writing, reviewing, and editing the manuscript, generating the original draft, visualizing the data, validating the results, designing the methodology, conducting the investigation, doing formal analysis, and contributing to the conceptualization of the study*: Ruwaida Aliyu;* ***reviewing and editing, and supervision****: Ainul Akmar Mokhtar;* ***Reviewing and editorial aspect:*** *Sirisha Nalakukkala;* ***Critical evaluation and software Analysis****: Temitope Ibrahim;* ***Provision of data and reviewing:*** *Abdulrahman Aminu Ghali;* ***Reviewing and proofreading:*** *Elhuseini Garba. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] A. Alrabghi and A. Tiwari, "State of the art in simulation-based optimisation for maintenance systems," *Comput. Ind. Eng.*, vol. 82, pp. 167–182, 2015, doi: 10.1016/j.cie.2014.12.022.

[2] Kristoffer K. McKee, G. Forbes, I. Mazhar, R. Entwistle, and I. Howard, "A review of major centrifugal pump failure modes with application to the water supply and sewerage industries," *ICOMS Asset Manag. Conf. , Gold Coast, QLD, Aust.*, pp. 1–12, 2011.

[3] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra, "Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0," *Inf. Fusion*, vol. 50, no. July 2018, pp. 92–111, 2019, doi: 10.1016/j.inffus.2018.10.005.

[4] S. Khalid, S. H. Jo, S. Y. Shah, J. H. Jung, and H. S. Kim, "Artificial Intelligence-Driven Prognostics and Health Management for Centrifugal Pumps: A Comprehensive Review," *Actuators*, vol. 13, no. 12, 2024, doi: 10.3390/act13120514.

[5] R. Daniel, "Maintenance Issues 15.1," 2019, pp. 883–916.

[6] M. J. Roemer and G. J. Kacprzynski, "Advanced diagnostic and prognostic technologies for gas turbine engine risk assessment," *Proc. ASME Turbo Expo*, vol. 4, 2000, doi: 10.1115/2000-GT-0030.

[7] Q. Wang and J. Gao, "Research and application of risk and condition based maintenance task optimization technology in an oil transfer station," *J. Loss Prev. Process Ind.*, vol. 25, no. 6, pp. 1018–1027, 2012, doi: 10.1016/j.jlp.2012.06.002.

[8] D. J. Langford *et al.*, "Coding of facial expressions of pain in the laboratory mouse," *Nat. Methods*, vol. 7, no. 6, pp. 447–449, 2010.

[9] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems — Reviews , methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, no. 1–2, pp. 314–334, 2014, doi: 10.1016/j.ymssp.2013.06.004.

[10] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Pet. Res.*, vol. 6, no. 4, pp. 379–391, 2021, doi: 10.1016/j.ptlrs.2021.05.009.

[11] L. Dong, Q. Xiao, Y. Jia, and T. Fang, "Review of research on intelligent diagnosis of oil transfer pump malfunction," *Petroleum*, no. xxxx, 2022, doi: 10.1016/j.petlm.2022.01.002.

[12] R. Mukherjee, R. Reddy Asani, N. Boppana, and M. M. El-Halwagi, "Performance evaluation of shale gas processing and NGL recovery plant under uncertainty of the feed composition," *J. Nat. Gas Sci. Eng.*, vol. 83, no. May, p. 103517, 2020, doi: 10.1016/j.jngse.2020.103517.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[14] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Reliab. Eng. Syst. Saf.*, vol. 182, pp. 208–218, 2019.

[15] M. M. Misren, "Reliability, Availability and Maintainability Analysis for Main Oil Line Pump by Dominant Failure Modes," *Univ. Teknol. Petronas*, vol. 1, no. 1, pp. 1–50, 2014.

[16] N. Hashim, A. Hassan, and M. F. A. Hamid, "Predictive maintenance model for centrifugal pumps under improper maintenance conditions," *AIP Conf. Proc.*, vol. 2217, no. April, 2020, doi: 10.1063/5.0000876.

[17] S. S. Patil, A. K. Bewoor, and R. B. Patil, "Availability Analysis of a Steam Boiler in Textile Process Industries Using Failure and Repair Data: A Case Study," *ASCE-ASME J Risk Uncert Engrg Sys Part B Mech Engrg*, vol. 7, no. 2, 2021, doi: 10.1115/1.4049007.

[18] S. Pirbhulal, V. Gkioulos, and S. Katsikas, "A Systematic Literature Review on RAMS analysis for critical infrastructures protection," *Int. J. Crit. Infrastruct. Prot.*, vol. 33, 2021, doi: 10.1016/j.ijcip.2021.100427.

[19] S. S. Patil and A. K. Bewoor, "Reliability analysis of a steam boiler system by expert judgment method and best-fit failure model method: a new approach," *Int. J. Qual. Reliab. Manag.*, vol. 38, no. 1, pp. 389–409, 2021, doi: 10.1108/IJQRM-01-2020-0023.

[20] E. Calixto and Y. Bot, "RAM analysis applied to decommissioning phase: Comparison and assessment of different methods to predict future failures," *Saf. Reliab. Methodol. Appl. - Proc. Eur. Saf. Reliab. Conf. ESREL 2014*, pp. 225–234, 2015, doi: 10.1201/b17399-35.

[21] R. K. Sharma and S. Kumar, "Performance modeling in critical engineering systems using RAM analysis," *Reliab. Eng. Syst. Saf.*, vol. 93, no. 6, pp. 913–919, 2008, doi: 10.1016/j.ress.2007.03.039.

[22] R. Guo, Z. Zhao, S. Huo, Z. Jin, J. Zhao, and D. Gao, "Research on state recognition and failure prediction of axial piston pump based on performance degradation data," *Processes*, vol. 8, no. 5, pp. 1–19, 2020, doi: 10.3390/PR8050609.

[23] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5695–5705, 2018, doi: 10.1109/TVT.2018.2805189.

[24] Y. L. Tse, M. E. Cholette, and P. W. Tse, "A multi-sensor approach to remaining useful life estimation for a slurry pump," *Meas. J. Int. Meas. Confed.*, vol. 139, pp. 140–151, 2019, doi: 10.1016/j.measurement.2019.02.079.

[25] C. Duan and C. Deng, "Prognostics of Health Measures for Machines with Aging and Dynamic Cumulative Damage," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 5, pp. 2264–2275, 2020, doi: 10.1109/TMECH.2020.2995757.

[26] A. Zhang, H. Srivastav, A. Barros, and Y. Liu, "Study of testing and maintenance strategies for redundant final elements in SIS with imperfect detection of degraded state," *Reliab. Eng. Syst. Saf.*, vol. 209, p. 107393, 2021, doi: 10.1016/j.ress.2020.107393.

[27] H. Srivastav, M. A. Lundteigen, and A. Barros, "Introduction of degradation modeling in qualification of the novel subsea technology," *Reliab. Eng. Syst. Saf.*, vol. 216, no. August, p. 107956, 2021, doi: 10.1016/j.ress.2021.107956.

[28] H. Srivastav, A. Barros, and M. A. Lundteigen, "Modelling framework for performance analysis of SIS subject to degradation due to proof tests," *Reliab. Eng. Syst. Saf.*, vol. 195, no. January 2019, p. 106702, 2020, doi: 10.1016/j.ress.2019.106702.

[29] P. C. Davis, M. A. Thornton, and T. W. Manikas, "Reliability block diagram extensions for non-parametric probabilistic analysis," *10th Annu. Int. Syst. Conf. SysCon 2016 - Proc.*, no. Mvl, 2016, doi: 10.1109/SYSCON.2016.7490656.

[30] P. P. Harihara, "Sensorless fault diagnosis of centrifugal pumps." 2009.

[31] L. Bachus and A. Custodio, "news product and services Book review Know and understand centrifugal pumps," no. November, p. 2003, 2003.

[32] R. K. Mobley, *An introduction to predictive maintenance*. Elsevier, 2002.

[33] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006, doi: 10.1016/j.ymssp.2005.09.012.

[34] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, 2014.

[35] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.

[36] S. Bracke, *Reliability Engineering*. Springer, 2024.

[37] R. Billinton and R. N. Allan, *Reliability evaluation of engineering systems*, vol. 198, no. 3. Springer, 1985.

[38] E. Zio and E. Ferrario, "A framework for the system-of-systems analysis of the risk for a safety-critical plant exposed to external events," *Reliab. Eng. Syst. Saf.*, vol. 114, pp. 114–125, 2013.

[39] M. Compare, P. Baraldi, and E. Zio, "Challenges to IoT-enabled predictive maintenance for industry 4.0," *IEEE Internet things J.*, vol. 7, no. 5, pp. 4585–4597, 2019.

[40] S. S. Patil, A. K. Bewoor, and R. B. Patil, "Availability analysis of a steam boiler in textile process industries using failure and repair data: A case study," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.*, vol. 7, no. 2, 2021, doi: 10.1115/1.4049007.