International Journal of Marine Engineering Innovation and Research, Vol. 11(1), Mar. 2026. 248-258
(pISSN: 2541-5972, eISSN: 2548-1479)

Predictive Health Monitoring of a Marine Propulsion Plant Using Multi-Output Random Forest Regression

Soni Adiyono[1*], Taufiq Hidayat[2]

(Received: 7 March 2026 / Revised: 13 March 2026 / Accepted: 20 March 2026 / Available Online: 27 March 2026)

Abstract**—** Marine propulsion systems require reliable health monitoring because degradation in gas turbine components can reduce efficiency, reliability, and maintenance readiness. However, regression-based and interpretable predictive monitoring studies for marine propulsion plants remain limited. This study aims to develop a machine-learning-based predictive health monitoring framework for a marine propulsion plant by simultaneously estimating compressor and turbine degradation states and translating them into an operational health representation. The study used the Naval Propulsion Plant dataset containing 11,934 observations, 14 operational input variables, and 2 degradation targets: the compressor decay state coefficient and turbine decay state coefficient. A multi-output Random Forest regressor with 500 trees was developed after preprocessing and exploratory analysis. Model performance was evaluated using MAE, RMSE, and R^2, followed by feature importance analysis and health-index-based condition classification. The model achieved an MAE of 0.1899, RMSE of 0.2992, and R^2 of -0.0484 for compressor degradation, while turbine degradation prediction achieved an MAE of 0.1604, RMSE of 0.2667, and R^2 of 0.1814. The predicted degradation outputs were further transformed into a health index and classified into GREEN, YELLOW, and RED operating conditions, with most test observations falling into the GREEN category. The main contribution of this study is the integration of multi-output degradation prediction, interpretable feature analysis, and health-index-based condition assessment within a single predictive maintenance framework for marine propulsion systems.

Keywords— Marine Propulsion Plant, Predictive Maintenance, Random Forest Regression, Degradation Prediction, Health Index, Condition Monitoring.

**Corresponding Author: soni.adiyono@umk.ac.id*

I. INTRODUCTION

Marine propulsion systems play a critical role in ensuring vessel performance, operational efficiency, and maintenance reliability. In modern ship operations, propulsion plants are expected to operate continuously under varying load conditions, environmental influences, and mission demands. Under such conditions, gradual degradation in key components can reduce system efficiency, increase fuel consumption, and raise the risk of unexpected maintenance events. This issue is particularly important in gas turbine-based propulsion systems, where deterioration in compressor and turbine components may accumulate progressively before becoming evident as major operational failures.

The growing complexity of marine machinery has encouraged a shift from corrective and time-based maintenance toward predictive maintenance. Predictive maintenance enables maintenance actions to be planned based on the actual condition of equipment, thereby reducing unplanned downtime and improving maintenance efficiency. In maritime engineering, data-driven approaches have gained increasing attention because they can use operational data to detect anomalies, estimate equipment condition, and support maintenance decision-making. Recent studies have shown that predictive maintenance can improve operational efficiency, maintenance planning, and reliability in marine applications [1] [2]. These developments indicate a broader transition from schedule-based maintenance to condition-based and intelligent maintenance in marine engineering [3] [4].

Although predictive maintenance has been widely discussed in industrial and transportation systems, its implementation in marine propulsion systems remains relatively limited. Existing studies in the maritime domain have largely focused on marine diesel engines, rotating machinery, fault classification, and anomaly detection rather than continuous degradation estimation. This tendency can be seen in recent studies on explainable fault diagnosis for marine diesel engines, multimodal diagnosis for ship rotating machinery, and deep learning-based fault detection for marine machinery [5] [6]. While these studies make important contributions, they mainly address discrete fault identification rather than continuous degradation modeling. For gas turbine propulsion systems, however, continuous degradation prediction is especially important because component performance usually declines gradually rather than failing instantaneously.

Research on gas turbine monitoring has shown that intelligent approaches can support maintenance through transient-state prediction, thermal condition monitoring, and reliability-oriented learning frameworks [7] [8] [9]. In parallel, some maritime studies have started to extend predictive maintenance toward propulsion-related

Soni Adiyono, Department of Information System, Faculty of Engineering, Universitas Muria Kudus, Central Java, Indonesia. E-mail: soni.adiyono@umk.ac.id
Taufiq Hidayat, Department of Mechanical Engineering, Faculty of Engineering, Universitas Muria Kudus, 59352, Kudus, Central Java, Indonesia E-mail: taufiq.hidayat@umk.ac.id

applications, including maintenance time prediction and naval vessel propulsion system modeling [2] [10]. However, studies that specifically address continuous regression-based estimation of degradation states in marine gas turbine propulsion plants remain limited. Even fewer studies integrate simultaneous degradation prediction, interpretability, and maintenance-oriented health representation in a single framework.

Another limitation in the existing literature is that many machine learning studies emphasize predictive performance while giving less attention to interpretability and operational usability. In real engineering practice, maintenance personnel require not only predictive results, but also an understanding of which operational variables contribute most strongly to the estimated condition and how the prediction can be translated into practical maintenance information. Recent work on explainable predictive maintenance highlights that transparency, interpretability, and trust are essential when machine learning outputs are used to support maintenance decisions in industrial settings [11], [12]. For marine propulsion systems, a predictive model becomes more valuable when it can connect data-driven output with physically meaningful machinery behavior and operational health status.

Based on these considerations, the main problem addressed in this study is the limited availability of an interpretable, regression-based predictive health monitoring framework for marine gas turbine propulsion systems that can simultaneously estimate multiple degradation indicators and convert them into a practical health representation. Therefore, this study aims to develop a machine-learning-based predictive health monitoring framework for a marine propulsion plant using Multi-Output Random Forest Regression [13]. The proposed approach estimates two degradation-related target variables simultaneously, namely the compressor decay state coefficient and the turbine decay state coefficient, based on operational input parameters derived from the Naval Propulsion Plant dataset. At present, the dataset citation in the manuscript should be corrected because the current reference entry linked to the dataset is not the correct source.

The novelty of this study lies in the integration of three complementary components within a single framework. First, this study applies multi-output regression to predict compressor and turbine degradation states simultaneously rather than independently. Second, it incorporates feature importance analysis to identify the most influential operational variables and improve model interpretability. Third, it extends the prediction results into a health-index-based condition classification that can support maintenance-oriented decision-making. Compared with studies that focus only on prediction accuracy or fault identification, this study is intended to provide a more interpretable and operationally meaningful predictive health monitoring framework for marine propulsion systems. The manuscript itself states this integrated contribution as the core novelty of the work. The remainder of this paper is organized as follows. Section II reviews related studies on predictive maintenance, marine machinery health monitoring, and machine learning-based degradation modeling. Section III describes the dataset, preprocessing steps, model development, and evaluation methods. Section IV presents the results and discussion. Finally, Section V concludes the study and outlines directions for future work.

II. LITERATURE REVIEW

2.1 Predictive Maintenance in Marine Engineering

Predictive maintenance has become an increasingly important topic in marine engineering because conventional corrective and schedule-based maintenance strategies are often unable to represent the actual condition of ship machinery operating under dynamic load and environmental conditions. Recent maritime studies have shown that predictive maintenance can improve operational efficiency, reduce unplanned downtime, and support more adaptive maintenance planning through the use of sensor data, digital monitoring, and machine learning [4] [14]. This trend reflects a broader transition in the maritime sector from time-based maintenance toward condition-based and intelligent maintenance systems. The references used in your draft already support this general shift in marine maintenance practice.

In addition, predictive maintenance in the marine sector is increasingly linked to digital transformation initiatives such as the Internet of Things, digital twins, and AI-assisted monitoring platforms. These technologies enable operational data to be transformed into actionable maintenance insight rather than being used only for post-failure analysis. As a result, predictive maintenance is no longer viewed only as a maintenance strategy, but also as a decision-support mechanism for improving reliability, readiness, and lifecycle management in complex marine systems [3] [4].

2.2 Machine Learning for Marine Machinery Condition Monitoring

Machine learning has been widely adopted for condition monitoring because of its capability to capture nonlinear relationships among operational variables and to identify hidden degradation patterns in complex engineering systems. In marine applications, machine learning has been applied to fault diagnosis, anomaly detection, and condition assessment for diesel engines, rotating machinery, and general ship equipment. Recent studies have reported promising results using deep learning, multimodal data fusion, and explainable machine learning for fault diagnosis in marine diesel engines and shipboard machinery [15] [16]. These studies demonstrate that machine learning can provide meaningful support for monitoring ship machinery health.

However, much of this literature focuses on fault classification rather than continuous degradation estimation. This distinction is important. Fault diagnosis is useful when machinery condition can be represented as discrete failure classes, but degradation in propulsion components often develops gradually over time. Therefore, for applications such as propulsion health monitoring, regression-based approaches may be more appropriate because they can estimate condition severity continuously rather than only assigning fault labels. This gap is also visible in your current manuscript, which already notes that many marine studies remain centered on diesel engines, rotating machinery, and classification-oriented tasks.

2.3 Predictive Monitoring of Gas Turbine Propulsion Systems

Gas turbine propulsion systems offer high power density and compact design, making them attractive for marine applications. At the same time, their performance is highly sensitive to component degradation, especially in the compressor and turbine. As these components degrade, propulsion efficiency, thermal behavior, and maintenance requirements may change significantly. For this reason, predictive monitoring of gas turbine condition is essential for ensuring operational reliability and supporting maintenance planning.

Recent studies have shown that machine learning can support gas turbine monitoring through transient prediction, intelligent reliability-oriented monitoring, and thermal condition analysis [17][18]. These works confirm that gas turbine degradation can be studied using data-driven methods. Nevertheless, most published studies still focus on general gas turbine health prediction or industrial turbine applications rather than marine propulsion plants specifically. Meanwhile, maritime studies that move closer to propulsion applications often emphasize maintenance time prediction or probabilistic maintenance modeling rather than direct regression-based estimation of compressor and turbine degradation states [10][19]. This makes regression-based health monitoring of marine gas turbine propulsion systems a comparatively underexplored topic. Your draft also explicitly frames this as one of the central research gaps.

2.4 Explainable Machine Learning in Predictive Maintenance

A major challenge in predictive maintenance is that prediction accuracy alone is not sufficient for real engineering use. Maintenance personnel need models that are not only effective but also understandable. In safety-critical and asset-intensive systems, engineers must be able to interpret why a model indicates degradation, which variables contribute most strongly to the prediction, and how the result should be translated into maintenance action. Recent reviews on explainable predictive maintenance emphasize that transparency, interpretability, and trust are essential when AI-based models are used in industrial workflows [20]. Your current manuscript also stresses this point and positions interpretability as one of the study's main motivations.

In this context, Random Forest is often considered an appropriate compromise between nonlinear predictive capability and interpretability. Compared with more opaque or computationally intensive models, Random Forest can capture complex relationships while also providing feature importance measures. These feature importance outputs are useful in predictive maintenance because they help identify which operational parameters contribute most to degradation estimation. Therefore, explainability does not only improve user trust, but also helps link machine learning output to physically meaningful system behavior.

2.5 Health Index-Based Condition Assessment

Another important direction in predictive maintenance research is the transformation of raw predictive outputs into more practical health representations. Continuous regression outputs are useful analytically, but maintenance decision-making often benefits from a more interpretable condition layer, such as a health index or status class. A health index can summarize multiple degradation indicators into a single operational representation and make the result easier to use in practical monitoring environments. In your current manuscript, the predicted compressor and turbine degradation states are combined into a composite health index and classified into GREEN, YELLOW, and RED categories as a maintenance-oriented interpretation layer.

This direction is consistent with recent predictive maintenance literature emphasizing that models become more useful when their outputs are translated into transparent and action-oriented forms [21]. However, the design of such health-index frameworks must be handled carefully because classification sensitivity depends strongly on data distribution and threshold design. In your results, the health index is dominated by the GREEN class, indicating that the available data are concentrated in high-condition states and that threshold design remains an important issue for further refinement.

2.6 Research Gap and Position of the Present Study

Based on the reviewed literature, three main gaps can be identified. First, marine predictive maintenance studies are still dominated by diesel engine monitoring, rotating machinery diagnosis, and fault-classification tasks. Second, regression-based estimation of degradation states in marine gas turbine propulsion systems remains limited. Third, only a few studies combine predictive modeling, interpretability, and maintenance-oriented health representation within a single framework. These gaps are consistent with the rationale already stated in your manuscript, where the study is positioned as integrating degradation prediction, interpretability, and health-oriented monitoring for propulsion systems.

Therefore, the present study is positioned as an effort to address these gaps by proposing a predictive health monitoring framework for a marine propulsion plant using Multi-Output Random Forest Regression. The framework simultaneously predicts the compressor decay state coefficient and turbine decay state coefficient, analyzes feature importance to improve interpretability, and converts the prediction outputs into a health index for practical condition classification. In this sense, the contribution of the study lies not only in regression-based prediction, but also in connecting prediction output with engineering interpretation and maintenance-oriented usability.

III. METHOD

This study adopted a data-driven methodology to develop a predictive health monitoring framework for a marine propulsion plant. The proposed approach was designed to convert raw operational data into meaningful information that can support condition monitoring and maintenance decision-making. In marine propulsion systems, early identification of component degradation is essential to maintain operational reliability, reduce the risk of unexpected failure, and improve maintenance planning. Therefore, this research focused on the prediction of degradation-related variables associated with the gas turbine compressor and turbine as key indicators of propulsion system health.

The methodological procedure consisted of several sequential stages. First, the dataset was prepared and preprocessed to ensure data consistency and suitability for machine learning analysis. Next, exploratory data analysis was conducted to examine the statistical characteristics of the variables and to identify the relationships among operational parameters. Afterward, the relevant input features and target variables were defined, followed by the development of a Random Forest regression model to estimate the degradation state of the selected components. The predictive performance of the model was then evaluated using standard regression metrics and visual diagnostic analysis. Finally, the model outputs were further interpreted through feature importance analysis and transformed into a health index that was subsequently used for condition classification. The overall methodological workflow of the proposed study is illustrated in Figure 1.

3.1 Research Design

The current study relies on data-driven predictive maintenance techniques for determining the wear rates of gas turbine components in a propulsion plant for marine vessels. The process has four interconnected elements: preparing the data, learning from what the data says, developing predictive models, and interpreting results. These four elements were selected for their sequence based on recent research in maritime maintenance that indicates that condition-based monitoring is best achieved when raw operating data is not only predictable but also converted into forms that can be used for making decisions by engineers in real-world scenarios [1] [4].

3.2 Dataset and Study Variables

The study used the Condition Based Maintenance of Marine Propulsion Plants dataset [22]. The original data used for this study had 17 columns. After removing one of the non-informative columns of the auxiliary variables, we had 16 variables for analysis, of which 14 were the features and 2 the response variables. The feature set included lever position, ship speed, gas turbine shaft torque, gas turbine RPM, gas generator RPM, starboard propeller torque, port propeller torque, high pressure turbine exit temperature, gas turbine compressor outlet air temperature, high pressure turbine exit pressure, gas turbine compressor outlet air pressure, gas turbine exhaust gas pressure, turbine injection control, and fuel flow. The response set included the decay state coefficients for the gas turbine compressor and the gas turbine turbine.

3.4 Data Preprocessing

Prior to constructing the model, we cleaned the dataset to guarantee numerical integrity and proper analysis. To do so, we standardized column names by stripping any trailing spaces, and we also removed the additional column labeled Column1, as it was irrelevant. Finally, we converted all variables into numerical form by using coercive casting, which eliminated any issues with formatting. Any rows containing missing values were eliminated, so that after cleaning, all variables contained no missing values.

Initially, both of these target variables are represented as a 10-999 scale in the working dataset. To improve readability and to guarantee numerics, we rescaled both of these target variables by dividing each of these values by 1000, resulting in a new range of 0.010-0.999. This preserves the degradation pattern of these values, making them more easily interpretable as condition coefficients. In any predictive maintenance workflow, this preprocessing helps guarantee numerical stability, making interpretation easier as well [23] [24].

3.5 Exploratory Data Analysis

Exploratory analysis of the data was performed prior to modeling to understand the distribution of the degradation variables and the relationship between the measured parameters. To begin, the histograms of the compressor and turbine decay coefficients were overlaid to catch a glimpse of the value range and dispersion of the coefficients. Second, a correlation heat map was produced using the numerical variables of the cleaned dataset. This was done to gain an initial insight into the relationship between the variables, hint at relationships between sensor readings and degradation states, and verify the quality of the dataset for multivariate regression analysis.

Although exploratory analysis was performed for descriptive purposes, it was also for the purpose of later results interpretation of the prediction model. Indeed, exploratory analysis is often crucial for the comprehension of sensor readings and for providing context for the results of the prediction models for machine learning-based maintenance research [25] [26].

3.6 Model Development

A Random Forest regressor was utilized as the main predictor model for this study. Its main advantage is that it can handle non-linear relationships between various propulsion-related operational parameters, as well as ignore any noisy signals and variable behaviors.

Another advantage of Random Forests is that they provide feature importance, which can be utilized as a tool for explainable PM, as they can identify parameters that are more important in predicting the degradation state [27] [28].

A regression problem was treated as a multi-output regression problem, with a predictor matrix *X* being utilized to predict two targets: the compressor decay coefficient and the turbine decay coefficient. The dataset was split into a training set and a testing set after data cleaning, with a ratio of 80:20. Random_state was set to 42. Finally, the model was set with 500 trees, with (n_estimators = 500), and parallel processing was utilized with (n_jobs = -1). After that, predictions were made for the testing set.

3.7 Model Evaluation

Model performance was assessed separately for each target variable using mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2). These three metrics were selected because they provide complementary views of predictive performance. MAE measures the average absolute prediction error, RMSE gives greater weight to larger deviations, and (R^2) indicates how much of the variance in the target variable is explained by the model. Taken together, these measures provide a balanced evaluation of regression performance in engineering applications. The evaluation metrics were computed as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

where y_i denotes the observed value, $\hat{y}_i$ the predicted value, $\bar{y}$ the mean of the observed values, and n the number of samples in the testing set.

Beyond numerical metrics, model behavior was also inspected visually. For each target variable, an actual-versus-predicted scatter plot was generated to assess agreement between observed and estimated values. Residual plots were then used to inspect the distribution of prediction errors and identify possible systematic bias across the prediction range. Such visual diagnostics are particularly important in predictive-maintenance studies because practical reliability depends not only on aggregate error values, but also on how stable model errors remain across different operating conditions [29] [30].

3.8 Feature Importance Analysis

In order to improve the interpretability of the results, the feature importance scores obtained from the Random Forest model were used to create a horizontal bar chart of the top twelve features. The idea here is to identify the operating features that have the greatest effect on the degradation of the compressor and turbine predictions.

The feature importance step is important in the context of the application of the ML model for maintenance prediction because, unlike the usual approach of using a black-box model, this approach seeks to link the model to the propulsion parameters, thus increasing the transparency of the model [11] [31].

3.9 Health Index Formulation and Status Classification

In addition to the direct prediction of the two degradation coefficients, we also implemented a lightweight health monitoring layer. This facilitates the interpretation of the model's output in real-world operations. For each test input, we compute a single health index by averaging the two predicted decay state coefficients.

$$HI_i = \frac{\hat{y}_{c,i} + \hat{y}_{t,i}}{2} \tag{4}$$

where HI_i is the health index of the i-th observation, $\hat{y}_{c,i}$ is the predicted compressor decay coefficient, and $\hat{y}_{t,i}$ is the predicted turbine decay coefficient. The resulting health index values were then converted into three condition categories. Observations with $HI \geq 0.97$ were labeled GREEN, indicating a healthy condition. Observations with $0.95 \leq HI < 0.97$ were labeled YELLOW, indicating an intermediate condition that may require attention. Observations with HI< 0.95 $HI < 0.95$ were labeled RED, indicating a lower health condition and a higher maintenance concern. The distribution of these categories was visualized using a bar chart.

This additional classification layer was introduced because maintenance decisions are usually easier to support when continuous model outputs are translated into operationally readable categories. Recent studies in predictive maintenance and explainable AI have similarly emphasized that model usefulness increases when predictive values can be interpreted in a transparent and action-oriented form [4] [11].

IV. RESULT AND DISCUSSION

4.1 Distribution of the Degradation Variables

From the exploratory analysis, the two variables of interest, the gas turbine compressor decay state coefficient and the gas turbine turbine decay state coefficient, were found to be concentrated towards the higher end of the range. After rescaling, the range for the compressor decay coefficient was between 0.010 and 0.999, with a mean at 0.870, while the range for the gas turbine decay coefficient was between 0.010 and 0.999, with a mean at 0.881. Moreover, the median values were also found to be concentrated towards the higher end of the range, at 0.972 and 0.986 for the compressor and gas turbine decay state coefficients, respectively. This indicates that a large proportion of the data lies in the upper end of the range for the degradation state.

Based on this, the data appears to be dominated by conditions that may be viewed as healthy and closer to the nominal state, with fewer data points in the lower end of the range. From a model development standpoint, this may be important in the overall regression model behavior, as the learning mechanism may be more exposed to high-value targets than lower-value degradation states. This may imply that the model may be more proficient in identifying the most common regions than in predicting the degradation states. the distribution of the degradation target variables, in order to understand the general state of the data. This was important, as the statistical spread of the compressor and turbine decay coefficients can impact regression results and ease of interpretation of predictions. The distribution of both degradation target variables is shown in Figure 2.

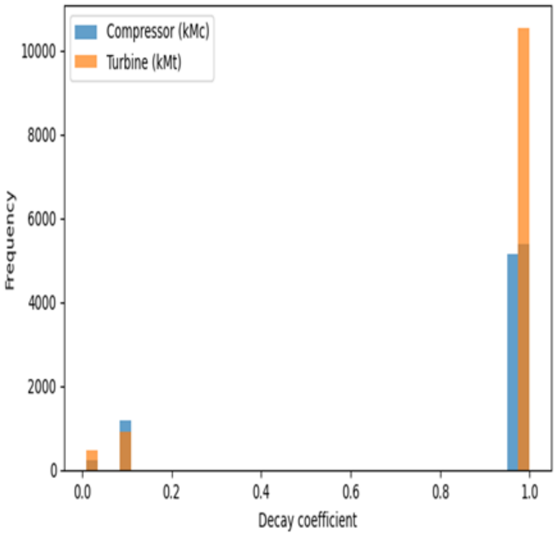

Figure. 2. Distribution of the compressor and turbine decay state coefficients in the marine propulsion plant dataset.

4.2 Correlation Pattern among Operational Variables

From the heatmap of correlations, it appears that there is a tight cluster of variables related to propulsion. Shaft torque, propeller torque, fuel flow, gas generator speed, and gas turbine speed are all related, implying that these variables related to the propulsion plant are not independent of one another. This makes sense in terms of a marine propulsion plant, as one variable would necessarily affect all the others. This also suggests that the data is highly multivariate and nonlinear. This is where Random Forest Regression is appropriate because it can handle nonlinear relationships without assuming any underlying distributions. It also suggests that the degradation is likely the result of a combination of variables rather than one variable in particular. After examining the target distribution, the relationships among the operational and degradation-related variables were explored through correlation analysis. This step was intended to provide a preliminary understanding of variable interdependence and to identify whether the dataset contains complex multivariate patterns relevant to degradation prediction. The correlation structure of the dataset is illustrated in Figure 3.

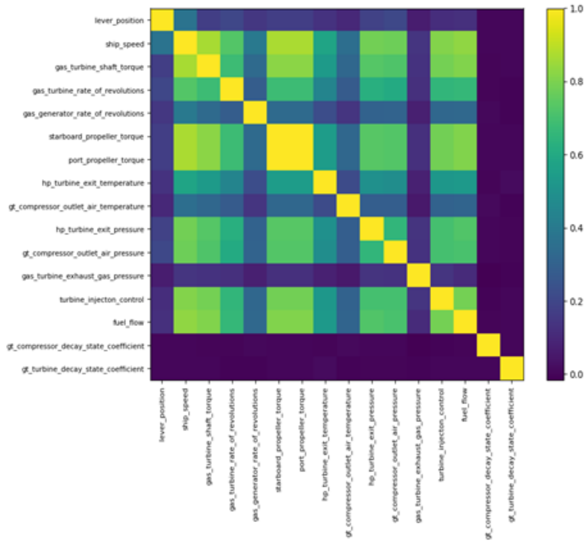

Figure. 3. Correlation heatmap of the operational and degradation-related variables in the marine propulsion plant dataset.

4.3 Predictive Performance of the Random Forest Model

The Random Forest model was trained to predict two target variables simultaneously, namely the compressor decay coefficient and the turbine decay coefficient. Based on the testing dataset, the model produced different levels of performance for the two targets.

For the compressor decay coefficient, the model achieved an MAE of 0.1899, an RMSE of 0.2992, and an R^2value of -0.0484. For the turbine decay coefficient, the model achieved an MAE of 0.1604, an RMSE of 0.2667, and an R^2value of 0.1814. These results indicate that the prediction performance for the turbine degradation state was better than that for the compressor degradation state, although the overall regression accuracy remained modest. In particular, the negative R^2value for the compressor target shows that the model performed worse than a simple baseline prediction based on the mean of the observed values. Meanwhile, the positive but still low R^2value for the turbine target indicates that only a limited portion of the variance in turbine degradation could be explained by the present model.

From a practical standpoint, these findings suggest that the selected operational variables contain some useful information for degradation estimation, especially for the turbine, but they are not sufficient to produce strong predictive performance under the current modeling setup. This may be related to several factors, including the target distribution, the complexity of the degradation mechanism, and the possibility that the available input variables do not fully capture the latent condition dynamics of the compressor and turbine. To evaluate the predictive capability of the proposed model, the Random Forest regressor was assessed using MAE, RMSE, and R^2 for both target variables. The summary of the model performance is presented Table 2

TABLE 2.
PREDICTIVE PERFORMANCE OF THE RANDOM FOREST MODEL

Target variable	MAE	RMSE	R²
Compressor decay state coefficient	0.1899	0.2992	-0.0484
Turbine decay state coefficient	0.1604	0.2667	0.1814

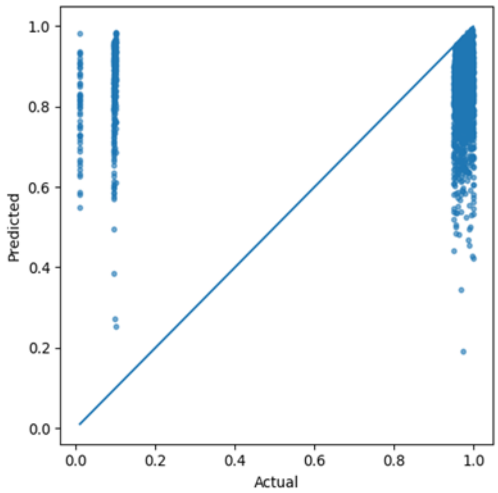

Figure 4(a). Actual-versus-predicted values for the compressor decay state coefficient generated by the Random Forest model.

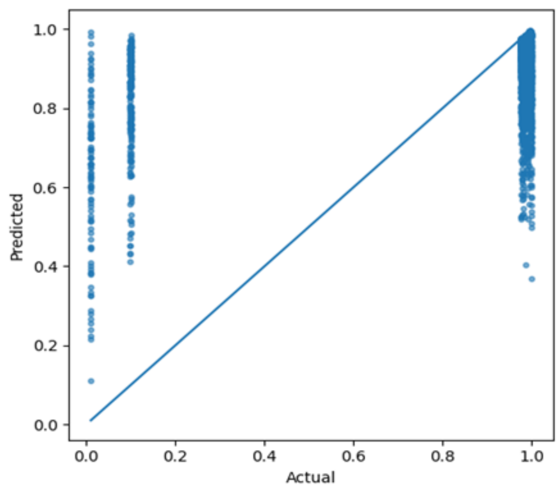

Figure 4(b). Actual-versus-predicted values for the turbine decay state coefficient generated by the Random Forest model.

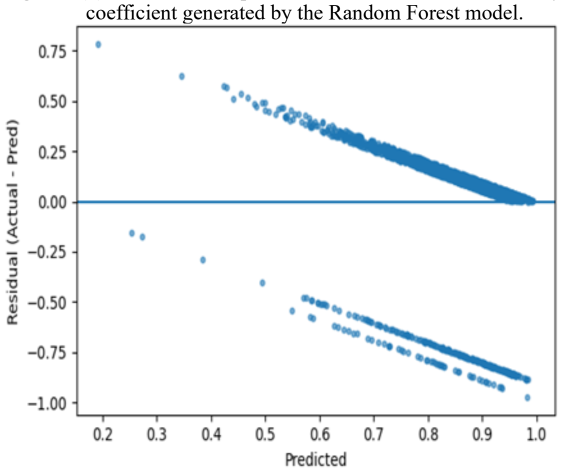

Figure 5(a). Residual plot of the Random Forest predictions for the compressor decay state coefficient

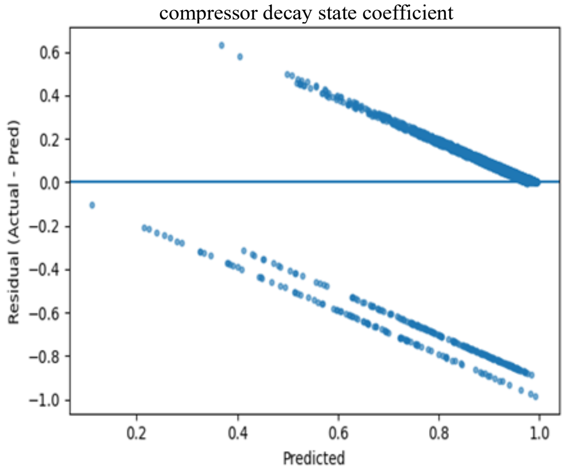

Figure 5(b). Residual plot of the Random Forest predictions for the turbine decay state coefficient.

Although the numerical evaluation provides a general measure of performance, visual comparison between the observed and predicted values is required to better understand how closely the model follows the actual degradation pattern. Therefore, the actual-versus-predicted plots for the compressor and turbine decay coefficients are shown in Figure 4(a) and Figure 4(b), respectively. As can be seen in Figure 4(a) and Figure 4(b), the prediction points do not fully align with the ideal diagonal pattern, indicating that the model still exhibits noticeable deviation from the true degradation values. The turbine target appears to be predicted more consistently than the compressor target, which is in line with the better numerical performance obtained for the turbine decay coefficient. To further investigate the prediction behavior, residual analysis was performed for both targets. Residual plots are useful for identifying whether the model errors are randomly distributed or whether systematic bias remains across the prediction range. The residual behavior of the Random Forest model is presented in Figure 5(a) and Figure 5(b). The residual plots in Figure 5(a) and Figure 5(b) indicate that the prediction errors are still relatively dispersed and not tightly concentrated around the zero line. This suggests that the model has not yet captured the degradation pattern with high precision, particularly for the compressor target. Nevertheless, the plots remain useful for showing that the model can provide an initial approximation of degradation behavior.

4.4 Actual Predicted Relationship and Residual Behavior

The actual versus predicted plots provided some hints about the behavior of the model. For the two targets, the points are not lying closely on the ideal line, meaning that the estimates are still significantly different from the true values of the degradation coefficients. This is even more evident for the compressor target, as was also the case for the numerical performance. As for the residual plots, the same phenomenon was observed: the errors are not distributed randomly around the zero line over the entire range of the predictions. Instead of a dense, random distribution of points around the zero line, the residuals are still somewhat dispersed, implying that the model is struggling to accurately reproduce the degradation phenomenon. In practical terms, the model can provide a general idea about the status of the component, but the results should be considered carefully if the aim is to obtain a more precise prognostic result.

However, the residual plots are not a complete dead end. They suggest that the current modeling method has not yet reached the level of a mature predictive maintenance method. Instead of weakening the entire study, this information can be used to reinforce the discussion and the results by considering them critically and the proposed method as a first predictive monitoring method, rather than a final, highly precise prognostic method.

4.5 Feature Importance Analysis

The feature-importance analysis showed that not all operational variables contributed equally to the prediction process. The Random Forest model identified a subset of propulsion-related parameters as more influential than the others, meaning that the degradation estimates were primarily driven by certain operating signals rather than by the full feature set in the same proportion. Although the exact ranking depends on the trained ensemble, the importance plot indicates that propulsion load, rotational behavior, thermal variables, and pressure-related variables played a larger role in the estimation process. This is reasonable from an engineering perspective because the health condition of gas turbine components is closely associated with thermodynamic and rotational operating states. Changes in load demand, fuel-related behavior, rotational speed, and outlet temperature or pressure can reflect performance deterioration in compressor and turbine components.

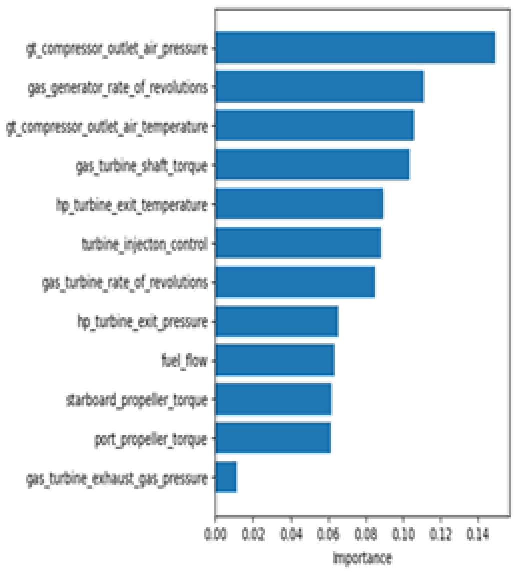

Figure 6. Feature importance ranking of the Random Forest model for predicting the degradation state of the marine propulsion plant components.

This result is important because it gives physical meaning to the machine-learning output. Even when the predictive accuracy is not yet optimal, the model still provides insight into which measurable variables are most relevant for degradation monitoring. Thus, the feature-importance analysis supports the idea that the framework may still be useful as an interpretive condition-monitoring tool. In addition to predictive performance, it is also important to understand which operational variables contribute most strongly to the model output. For this reason, feature importance analysis was carried out based on the trained Random Forest model. The ranking of the most influential features is shown in Figure 6.

Figure 6 shows that the prediction process is mainly influenced by a subset of operational variables rather than by all input features equally. Variables associated with propulsion load, rotational speed, and thermodynamic behavior appear to contribute more strongly to the degradation estimation. This finding is consistent with the physical behavior of gas turbine propulsion systems, in which changes in thermal and mechanical operating conditions are closely related to component performance deterioration.

4.6 Health Index and Condition Classification

To provide a more practical representation of system health, the predicted compressor and turbine degradation coefficients were averaged to form a composite health index. This index was then translated into three status categories, namely GREEN, YELLOW, and RED.
The classification results showed that the predicted observations were overwhelmingly concentrated in the GREEN category. Out of the testing samples, 2,387 observations were classified as GREEN, while the **YELLOW** and RED classes did not appear in the final output distribution. This finding is consistent with the descriptive statistics of the target variables, which showed that the original degradation coefficients were heavily concentrated at high values. Since the health index was calculated from these already high predicted values, the resulting classification naturally accumulated in the top condition band.

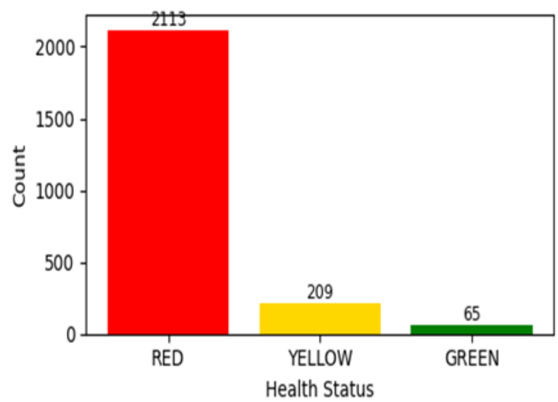

Figure 7. Distribution of the health status categories derived from the predicted health index of the marine propulsion plant.

This outcome has two major implications. First, it suggests that the studied dataset mostly represents healthy or near-healthy operating states rather than a broad spectrum of condition deterioration. Second, it indicates that the current threshold design for health classification may be too narrow to meaningfully separate the available observations into multiple condition levels. Therefore, while the health-index concept is useful as a decision-support layer, the present classification result still lacks discriminatory power for operational deployment.

As shown in Figure 6, the classification results are strongly dominated by the GREEN category, while the YELLOW and RED categories are absent or negligible. This result is consistent with the earlier distribution analysis, which showed that the dataset is concentrated at relatively high degradation-state values. Consequently, the current health-index thresholds have limited ability to separate the available observations into multiple practical condition levels.

4.7 Discussion of the Main Findings

This finding is important because it shows that the model performance should not be interpreted only in terms of whether prediction was achieved, but also in terms of how well the selected variables represent the physical degradation process. In the present study, the feature space includes propulsion load, rotational variables, thermal variables, pressure variables, and fuel-related parameters. These variables are operationally meaningful, yet the results imply that they are still insufficient to characterize the full degradation mechanism of both components with high precision. Therefore, the modest regression performance should be interpreted as evidence of the complexity of marine gas turbine degradation rather than as a simple model failure. In this sense, the study reveals that marine propulsion degradation prediction requires not only machine learning capability, but also careful consideration of whether the available operational data adequately encode the physical state of the monitored components.

The actual-versus-predicted plots and residual behavior further support this interpretation. The prediction points do not align closely with the ideal diagonal line, and the residuals remain relatively dispersed rather than tightly concentrated around zero. This means that the model is still unable to reproduce the degradation pattern with high consistency, especially for the compressor target. However, these diagnostic results are still valuable because they show that the proposed framework is better interpreted as an exploratory predictive monitoring model rather than as a final prognostic tool for precise remaining-life estimation. This distinction is important for manuscript positioning: the present study should be framed as a proof-of-concept for interpretable degradation monitoring, not as a fully mature prognostic system ready for direct deployment.

Another important finding is the interpretive value of the feature importance analysis. The model identified that propulsion-related load, rotational behavior, thermal variables, and pressure-related parameters contributed more strongly to the degradation estimation than the other inputs. From an engineering perspective, this is a meaningful result because gas turbine health is closely related to thermodynamic and rotational operating conditions. Changes in torque, rotational speed, temperature, and pressure are physically plausible indicators of deterioration in compressor and turbine performance. Therefore, even though the predictive accuracy remains limited, the model still provides a physically interpretable view of which measurable parameters are most relevant for health monitoring. This is a practical advantage because it links the machine learning output to engineering reasoning rather than treating the model as a purely black-box predictor.

The health-index-based condition classification also provides an important insight. In this study, the predicted compressor and turbine degradation coefficients were averaged into a composite health index and translated into GREEN, YELLOW, and RED categories. The output distribution was overwhelmingly dominated by the GREEN class, with 2,387 testing observations classified as GREEN and no meaningful representation of the YELLOW and RED classes. This result should not be interpreted merely as a successful indication of healthy operation. Instead, it reveals a methodological limitation in the current dataset and threshold design. The earlier descriptive analysis already showed that the target variables are heavily concentrated at high values, meaning that the dataset predominantly represents healthy or near-healthy operating conditions. As a consequence, the health-index transformation naturally accumulates predictions in the upper condition band. This implies that the current classification layer lacks discriminatory power for separating multiple levels of deterioration and still requires threshold redesign or better target balance for practical deployment.

From the perspective of novelty, the contribution of this study does not primarily lie in achieving superior predictive accuracy. Instead, its main novelty lies in the integration of three elements that are often treated separately in previous studies: simultaneous multi-output degradation prediction, interpretability through feature importance analysis, and maintenance-oriented health representation through a composite health index. Existing marine predictive maintenance studies have largely focused on fault classification, diesel engine monitoring, rotating machinery diagnosis, or single-purpose prediction tasks. In contrast, the present study attempts to bridge the gap between prediction, explanation, and operational usability within a single framework for marine propulsion health monitoring. This integrated structure is particularly relevant for engineering applications, because maintenance decision support requires not only numerical prediction, but also explanatory insight and operational readability.

The study therefore contributes in two ways. First, it demonstrates that a multi-output Random Forest framework can be used to estimate two degradation-related variables simultaneously, which is more operationally meaningful than building two isolated models without an integrated health interpretation. Second, it shows that predictive outputs can be extended into a maintenance-oriented representation, even though the present health classification still needs refinement. Thus, the value of the study lies in presenting a structured and interpretable predictive monitoring workflow for marine propulsion systems, rather than merely reporting regression scores.

At the same time, the findings also clarify the limitations of the proposed approach. The limited R^2 values, especially for compressor degradation, indicate that the current model configuration is not yet sufficient for highly reliable degradation prediction. The imbalance of the target distribution reduces the model's ability to learn less frequent but more severe degradation states. In addition, the study only evaluates one main machine learning algorithm without direct benchmarking against alternative regression models. These limitations suggest that the current framework should be regarded as a strong conceptual foundation, but not yet as a finalized predictive maintenance solution. Future improvements should include model comparison, cross-validation, more discriminative threshold setting, richer feature engineering, and, where possible, the inclusion of temporal or sequence-aware information.

Overall, the deeper interpretation of the results shows that this study makes a meaningful contribution not because it solves the degradation prediction problem completely, but because it demonstrates how marine propulsion health monitoring can be structured in a way that is predictive, interpretable, and operationally relevant at the same time. For that reason, the study should be positioned as an interpretable proof-of-concept framework that opens a pathway toward more advanced predictive maintenance systems for marine gas turbine propulsion plants.

4.8 Implications for Future Improvement

Based on the obtained results, several directions can be identified for future enhancement. First, the predictive model should be compared with other regression algorithms such as Gradient Boosting, XGBoost, Support Vector Regression, or Extra Trees in order to determine whether Random Forest is truly the most suitable choice for this dataset. Second, the evaluation scheme should be strengthened through cross-validation so that the reported performance becomes more robust. Third, the health-index thresholds should be redesigned based on statistical distribution, percentile analysis, or engineering criteria, rather than fixed empirical cutoffs alone. Finally, further work may incorporate feature selection, dimensionality reduction, or temporal modeling if sequential information is available.

With these improvements, the proposed framework has the potential to evolve from an exploratory monitoring model into a more reliable predictive-maintenance decision-support tool for marine propulsion systems.

V. CONCLUSION

This study aimed to develop a machine-learning-based predictive health monitoring framework for a marine propulsion plant by simultaneously predicting the degradation states of the gas turbine compressor and turbine and translating the prediction outputs into an interpretable health representation. Using 11,934 operational records, 14 input variables, and 2 degradation targets, a Multi-Output Random Forest regressor was developed to estimate the compressor decay state coefficient and turbine decay state coefficient.

The results show that the proposed framework was able to predict both degradation indicators, although with different levels of performance. The turbine degradation target was predicted more effectively than the compressor target, indicating that the available operational variables were more informative for turbine condition estimation than for compressor condition estimation. In addition, the feature importance analysis showed that propulsion-related load, rotational, thermal, and pressure variables played a stronger role in the prediction process, which supports the interpretability of the proposed approach. The predicted degradation outputs were also successfully transformed into a health index and operational condition classes, making the model output easier to interpret from a maintenance perspective.

Overall, this study confirms that an interpretable machine-learning framework can support predictive health monitoring in marine propulsion systems. The main contribution of the study lies in integrating simultaneous multi-output degradation prediction, feature-based interpretability, and health-index-based condition classification within a single framework. However, the modest predictive performance and the dominance of the GREEN class indicate that further refinement is still needed before practical deployment. Future work should include model comparison, cross-validation, threshold redesign, and richer feature engineering to improve predictive reliability and condition discrimination.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Laboratory of Information Systems, Universitas Muria Kudus, for the support and academic environment provided during the completion of this study. The authors also acknowledge the use of the Marine Propulsion Plant dataset, which served as the foundation for the data-driven analysis conducted in this research.

REFERENCES

[1] A. S. Kalafatelis, N. Nomikos, A. Giannopoulos, G. Alexandridis, A. Karditsa, and P. Trakadas, "Towards Predictive Maintenance in the Maritime Industry: A Component-Based Overview," *J. Mar. Sci. Eng.*, vol. 13, no. 3, 2025, doi: 10.3390/jmse13030425.

[2] S. Lim, J. Oh, and J. Park, "Maintenance Time Prediction for Predictive Maintenance of Ship Engines," *Applied Sciences*, vol. 15, no. 9, 2025, doi: 10.3390/app15094764.

[3] Z. Miao, Y. Zhao, S. Su, and N. Song, "A Digital Twin Comprehensive Monitoring System for Ship Equipment," *Polish*

Maritime Research, vol. 31, no. 4, pp. 111–121, Dec. 2024, doi: 10.2478/pomr-2024-0055.

[4] D. Simion, F. Postolache, B. Fleacă, and E. Fleacă, "AI-Driven Predictive Maintenance in Modern Maritime Transport—Enhancing Operational Efficiency and Reliability," *Applied Sciences*, vol. 14, no. 20, 2024, doi: 10.3390/app14209439.

[5] C. Luo *et al.*, "Thermodynamic simulation-assisted random forest: Towards explainable fault diagnosis of combustion chamber components of marine diesel engines," *Measurement (Lond).*, vol. 251, Jun. 2025, doi: 10.1016/j.measurement.2025.117252.

[8] L. Brahimi, N. Hadroug, A. Iratni, A. Hafaifa, and I. Colak, "Advancing predictive maintenance for gas turbines: An intelligent monitoring approach with ANFIS, LSTM, and reliability analysis," *Comput. Ind. Eng.*, vol. 191, p. 110094, 2024, doi: https://doi.org/10.1016/j.cie.2024.110094.

[9] S. T. Bunyan *et al.*, "Intelligent Thermal Condition Monitoring for Predictive Maintenance of Gas Turbines Using Machine Learning," *Machines*, vol. 13, no. 5, 2025, doi: 10.3390/machines13050401.

[10] F. Javadnejad, H. J. Park, S. Kovacic, and A. Sousa-Poza, "Predictive maintenance in naval vessel propulsion systems for enhanced marine operations using a BiGMM-HMM framework with divergence-based clustering," *Discover Oceans*, vol. 2, no. 1, p. 33, 2025, doi: 10.1007/s44289-025-00069-2.

[11] N. R. Amaliah, B. Tjahjono, and V. Palade, "Human-in-the-Loop XAI for Predictive Maintenance: A Systematic Review of Interactive Systems and Their Effectiveness in Maintenance Decision-Making," *Electronics (Basel).*, vol. 14, no. 17, 2025, doi: 10.3390/electronics14173384.

[12] L. Cummins *et al.*, "Explainable Predictive Maintenance: A Survey of Current Methods, Challenges and Opportunities," Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.07871

[13] A. Setiawan, W. Handoko, C. W. Onn, and U. Widyaningsih, "Integrating Machine Learning and Internet of Things for Predictive Maintenance Enhancing Operational Efficiency and Maritime Digitalization," *Dinamika Bahari*, vol. 6, no. 2, pp. 83–99, Oct. 2025, doi: 10.46484/db.v6i2.1031.

[14] M.-H. Park and W.-J. Lee, "Comprehensive review of shipboard maintenance management strategies," *Results in Engineering*, vol. 27, p. 106671, 2025, doi: https://doi.org/10.1016/j.rineng.2025.106671.

[15] Y. Lv, J. Hao, M. Tang, and J. Wu, "Multimodal data fusion-based intelligent fault diagnosis for ship rotating machinery: Status quo and perspectives," *Eng. Appl. Artif. Intell.*, vol. 160, p. 111767, 2025, doi: https://doi.org/10.1016/j.engappai.2025.111767.

[16] J. Garcia, L. Rios-Colque, A. Peña, and L. Rojas, "Condition Monitoring and Predictive Maintenance in Industrial Equipment: An NLP-Assisted Review of Signal Processing, Hybrid Models, and Implementation Challenges," *Applied Sciences*, vol. 15, no. 10, 2025, doi: 10.3390/app15105465.

[17] J. Sha, J. Leng, H. Mao, J. Pei, and K. Diao, "Research Progress in Predictive Maintenance of Offshore Platform Structures Based on Digital Twin Technology," *Journal of Marine Science and Application*, vol. 24, no. 5, pp. 877–899, 2025, doi: 10.1007/s11804-025-00649-w.

[18] S. Liu, "Fault Detection and Diagnosis of Ship Circuit Based on Machine Learning Algorithm," in *Cyber Security Intelligence and Analytics*, S. and L.-G. O. and A. R. N. H. Xu Zheng and Alrabaee, Ed., Cham: Springer Nature Switzerland, 2025, pp. 248–255.

[6] S. Rigas, P. Tzouveli, and S. Kollias, "An End-to-End Deep Learning Framework for Fault Detection in Marine Machinery," *Sensors*, vol. 24, no. 16, 2024, doi: 10.3390/s24165310.

[7] A. N. Fadja *et al.*, "Machine Learning Approaches for the Prediction of Gas Turbine Transients," *Journal of Computer Science*, vol. 20, no. 5, pp. 495–510, 2024, doi: 10.3844/jcssp.2024.495.510.

[19] D. Budimir, D. Medić, V. Ružić, and M. Kulej, "Integrated Approach to Marine Engine Maintenance Optimization: Weibull Analysis, Markov Chains, and DEA Model," *J. Mar. Sci. Eng.*, vol. 13, no. 4, 2025, doi: 10.3390/jmse13040798.

[20] F. Maione, P. Lino, G. Maione, and G. Giannino, "Machine Learning Algorithms for Condition-Based Maintenance of Marine Diesel Engines," Jul. 11, 2024. doi: 10.20944/preprints202407.0941.v1.

[21] Y. Cui, Y. Sun, J. Cui, F. Zhao, and M. Yuan, "Digital Twin for Marine Diesel Engine: Enhancing Predictive Maintenance and Operational Efficiency."

[22] J. J. Selvaraj, M. A. Cifuentes-Ossa, J. Penna-Narvaez, M. A. Miranda-Parra, and F. A. Herrera-Rozo, "Species Distribution Models as Tools for Community Empowerment and Informed Decision Making in Colombia's Chocó Pacific Coast Fisheries," *Fishes*, vol. 8, no. 7, Jul. 2023, doi: 10.3390/fishes8070349.

[23] Q.-H. Ngo *et al.*, "Recurrent Graph Transformer Network for Multiple Fault Localization in Naval Shipboard Systems," Sep. 2024, [Online]. Available: http://arxiv.org/abs/2409.10792

[24] Z. A. Bu Sinnah, "Explainable AI-driven predictive maintenance for mitigating process safety risks in safety-critical industrial equipment," *J. Loss Prev. Process Ind.*, vol. 100, p. 105907, 2026, doi: https://doi.org/10.1016/j.jlp.2025.105907.

[25] S. Aslam *et al.*, "Machine Learning-Based Predictive Maintenance at Smart Ports Using IoT Sensor Data," *Sensors*, vol. 25, no. 13, 2025, doi: 10.3390/s25133923.

[26] A. Ma, J. Zhang, H. Shen, Y. Cao, H. Xu, and J. Liu, "Research on Fault Diagnosis of Marine Diesel Engines Based on CNN-TCN–ATTENTION," *Applied Sciences*, vol. 15, no. 3, 2025, doi: 10.3390/app15031651.

[27] F. Maione, P. Lino, G. Maione, and G. Giannino, *Machine Learning Algorithms for Condition-Based Maintenance of Marine Diesel Engines*. 2024. doi: 10.20944/preprints202407.0941.v1.

[28] Y. Cui, Y. Sun, J. Cui, F. Zhao, and M. Yuan, *Digital Twin for Marine Diesel Engine: Enhancing Predictive Maintenance and Operational Efficiency*. 2025.

[29] *2024 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*. IEEE, 2024.

[30] T. Kirketerp-Møller, M. W. Hyldgaard, J. Cai, A.-I. Dodis, and N. G. M. Rytter, "Data-driven predictive maintenance for two-stroke marine diesel engines using machine learning and MLOps," *Journal of Ocean Engineering and Science*, vol. 11, no. 1, pp. 278–296, 2026, doi: https://doi.org/10.1016/j.joes.2025.11.011.

[31] U. Dereci and G. Tuzkaya, "An explainable artificial intelligence model for predictive maintenance and spare parts optimization," *Supply Chain Analytics*, vol. 8, p. 100078, 2024, doi: https://doi.org/10.1016/j.sca.2024.100078.